The Dynamic Genetic Atlas of 122 Gestational Phenotypes

Siyang Liu[1*#], Hao Zheng[1,2,3*], Yuqin Gu[1,3*], Zijing Yang[1,2*], Yanhong Liu[1], Yuandan Wei[1,3], Xinxin Guo[1], Yanchao Chen[1], Liang Hu[2], Xiaohang Chen[2], Fuquan Zhang[3], Guo-Bo Chen[4], Xiu Qiu[5], Shujia Huang[5#], Jianxin Zhen[3#], Fengxiang Wei[2,6#]

1. School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, Guangdong 518107, China
2. The Genetics Laboratory, Longgang District Maternity and Child Healthcare Hospital of Shenzhen City, Shenzhen, Guangdong, 518172, China
3. Central Laboratory, Shenzhen Baoan Women's and Children's Hospital, Shenzhen, Guangdong 518102, China
4. Center for Productive Medicine, Department of Genetic and Genomic Medicine, Clinical Research Institute, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou 310014, Zhejiang, China
5. Division of Birth Cohort Study, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, 510623, China
6. Longgang Maternity and Child Institute of Shantou University Medical College, Shenzhen, Guangdong, 518172, China

*: Those authors contribute equally as co-first authors

#: Correspondence can be addressed to

22 Siyang Liu liusy99@mail.sysu.edu.cn

23 Shujia Huang shujia.huang@bigcs.org

24 Jianxin Zhen jxzhen@qq.com

25 Fengxiang Wei haowei727499@163.com

26

Abstract

The gestational period, spanning approximately 40 weeks from fertilization to birth, is pivotal in human reproduction. Monitoring the health of pregnant women and newborns during this period involves systematic prenatal and postpartum examinations, guided by indicators established under the national medical insurance system, collectively termed gestational phenotypes. However, our understanding of the genetic basis of these phenotypes and their intricate relationship with maternal long-term health outcomes remain markedly limited. We conducted comprehensive genetic investigations into 122 gestational phenotypes among 121,579 Chinese pregnancies. These phenotypes included anthropometric metrics, comprehensive blood biomarker measurements, and common gestational complications and outcomes. We identified 3,845 genetic loci, 1,385 of which are novel. Our analyses revealed gestation-specific genetic effects, ranging from proportion 0% to 100% for 23 phenotypes, highlighting genes and pathways predominantly enriched in response to hormones, growth and immune function. Longitudinal trajectory genome-wide association study (GWAS) analyses of repeated measures across 24 complete blood cell phenotypes revealed that 17.8% of the genetic variants exhibited significant interactions with gestational timing across five gestational and postpartum periods. Two-sample univariable and multivariable Mendelian Randomization (MR) analyses of 220 mid- and old-age phenotypes suggested causal associations between gestational phenotypes and the risk of chronic diseases in later life. These findings provide initial insights into the genetic foundations of human gestational phenotypes and their relationship with long-term health, laying a basis for advanced population health during gestation.

Introduction

Women's reproductive, maternal, newborn, child, and adolescent health are fundamental aspects of human health development and the key drivers of future population and social progress[1,2]. Amid the global trends of population aging and declining fertility rates[3,4], maternal and child health has increasingly emerged as a critical global public health concern. Particularly noteworthy is the gestational period, starting from fertilization, to birth over approximately 40 weeks, is a crucial phase during which the developing organism undergoes substantial growth and differentiation, while the pregnant individual experiences various physiological changes to support the fetus[5,6]. Given the importance of this period, the health of pregnant women and newborns is monitored through a set of systematic and timed prenatal and postpartum examinations manipulated via the national medical insurance system. These indicators encompass anthropometric metrics, a series of blood and urine measurements and prenatal tests, forming the basis for diagnosing pregnancy disorders and forecasting birth outcomes, collectively making up the gestational phenotypes (**Fig. 1 and Supplementary Table 1**).

Despite the critical nature of the gestation period, the genetic underpinnings underlying the gestational phenotypes and their connections to later age health outcomes remain poorly understood. Of the 122 gestational phenotypes, only 29 have been explored in previous GWAS studies (**Supplementary Table 1**), and 11 were part of our earlier work[7-9]. While extensive genetic studies have been conducted on a variety on various quantitative phenotypes in the general population[10,11], few have addressed the distinct genetic basis of phenotypes during the gestation[8,9,12], and the time-dependent genetic effects during this period remain largely unknown. Additionally, despite observational studies have linked gestational status to maternal older age health status[13-16], evaluations of the causal effects of the spectrum of gestational phenotypes, which usually collected earlier in life, on later age health

status have not yet been achieved, hindering the potential for early screening of diseases for women typically aged 20 to 40.

To address this gap, we collected data on 122 prenatal and postnatal screening indicators and non-invasive prenatal test (NIPT) sequencing data from 121,579 unrelated pregnant women (mean age 30 ± 4). We posed the following questions: (1) What are the genetic determinants for these gestational phenotypes and what is their heritability? (2) Do pregnancy-specific genetic effect exit, and which genes exhibit these effects? (3) Are there gene-by-environment ($G \times E$) interactions for the genetic loci associated with gestational phenotypes? (4) What are the causal relationships between gestational phenotypes and maternal long-term health? To address these questions, we conducted a large-scale genome-wide association study (GWAS) using our developed analytical protocol[7,17], constructing an atlas of quantitative and qualitative trait loci and replicating our findings in external cohorts. We investigated patterns of genetic loci with differential effects during pregnancy and non-pregnancy using a Bayesian clustering algorithm, and examined $G \times E$ effects through repeated sampling of the same phenotypes during pregnancy using regression models. Finally, we used our atlas to estimate potential causal relationships between gestational phenotypes and 220 phenotypes in BioBank Japan (mean age 63 years), mimicking an impact of the gestational phenotypes over the long-term health outcomes[10] (**Fig. 1**).

Results

Study design

We collected NIPT sequencing data and 122 comprehensive pregnancy screening measurements, including biomarkers and electronic medical records, from 121,579 Chinese pregnancies who participated in routine obstetric examinations at two hospitals in Shenzhen City. We categorized the 88 gestational phenotypes into nine molecular test categories: Blood lipid (N=2), Glycemic (N=6), Blood routine (N=32),

Infection (N=14), Kidney (N=3), Liver function (N=8), Thyroid function (N=4), Tang screening (N=16) and NIPT screening (N=3) (**Fig.1, Supplementary Table 2**). Additionally, we included 34 phenotypes in three categories of electronic medical records: Basic information such as maternal height, weight, blood pressure (N=7), Gestational disorders (N=21) and Birth outcomes (N=6) (**Fig. 1, Supplementary Table 3**). Sample sizes for each phenotype ranged from 12,024 (gestational diabetes mellitus with medication in Glycemic) to 115,872 (trisomy 18 risk score in NIPT screening). The distribution of all quantitative gestational phenotypes and the prevalence of common gestational comorbidities were highly consistent between the two hospitals (**Supplementary Table 4-5, Supplementary Fig. 1-2**).

All participants underwent NIPT as part of standard prenatal screening in Shenzhen. We applied our estimated protocol to infer and impute genotypes, estimate family-relatedness, infer population structure, and conduct GWAS. More details about the analyses are provided in the **Methods** section.

3,845 genetic associations with 122 gestational phenotypes

Using the collected phenotype data and NIPT genotypes, we performed GWAS for the 122 gestational phenotypes from Baoan and Longgang hospitals, respectively, followed by a fixed-effect meta-analysis across 12.6 million genetic markers (**Supplementary Table 6**). Linkage disequilibrium (LD) score intercepts were between 0.94 and 1.10 for all 122 phenotypes (**Supplementary Table 7**), which suggested that population structure in our analysis was well controlled. Besides the conventional genome-wide significant threshold ($P < 5\times10^{-8}$), we set a stricter study-wide significant threshold to $P = 4.1\times10^{-10}$ ($P < 5\times10^{-8}/122$) after Bonferroni correction for multiple testing. After GCTA conditional and joint analysis[18], we identified 4,309 independent signals in 3,242 loci associated with at least one phenotypes ($P<5\times10^{-8}$) among nine categories (**Supplementary Table 8**), where

2,942 signals in 1,999 loci reached study-wide significance (P<4.1×10^{-10}) (**Fig. 2, Supplementary Table 9**). The highest number of loci were found in the blood routine category among all the categories, and the phenotype with the highest number of loci was platelet-large cell ratio (P_LCR) (**Fig. 2, Supplementary Table 9**). Within the three categories of basic information, birth outcomes and gestational disorders, we found 367 independent signals in 279 loci ($P < 4.1\times10^{-10}$) and 718 independent signals in 603 loci ($P < 5\times10^{-8}$) (**Supplementary Table 10-11**).

Genetic effect estimations are highly consistent between the two hospitals with squared Pearson's correlation coefficient R^2 greater than 0.76 for nine categories of quantitative traits and R^2 greater than 0.12 for three categories of qualitative traits (**Supplementary Fig. 3**). Only 369 out of 4,309 (8.5%) loci associated with the nine categories of gestational phenotypes and 74 out of 718 (10.3%) associated with the three categories of gestational phenotypes exhibited statistical heterogeneity (P_{het}<0.05) (**Supplementary Table 8-11**). We also performed two replication studies comparing genetic effect estimates with independent cohorts to evaluate the robustness of effect estimates in our study. In replication study I, assuming maternal height is not influenced by a gestational status, we compared genetic effect estimates of maternal height-associated variants with that estimated in the Taiwan Biobank with the GWAS summary statistic for maternal height in this study[19]. The genetic effect estimates are highly consistent between our study and the Taiwan Biobank study with R^2=0.93 (**Supplementary Fig. 4**). There were 121 variants with P values less than the Bonferroni correction threshold ($P < 2.79\times10^{-4}$, 66.9%), and 166 variants with P value less than 0.05 (91.7%) (**Supplementary Fig. 4, Supplementary Table 12**). In replication study II, we compared the genetic effects with an independent cohort (NIPT PLUS pregnancy cohort), which is our additional collection of maternal check-up data from 5,733 pregnant women (**Methods**). Using *two-sample t-test*, 4,059 of the 4,804 variants of the 103 phenotypes with data in NIPT PLUS cohort had consistent

genetic effect size in independent cohort (84.5%) (**Supplementary Fig. 5**, **Supplementary Table 13**). Of the 3,198 variants that were study-wide significance within the study, 2,786 variants had consistent genetic effect (87.1%) (**Supplementary Table 13**). The above results indicate high fidelity of the GWAS findings and genetic effect estimates in this study.

1,385 novel genetic associations

Among the 122 gestational phenotypes, 29 were previously investigated among pregnant women and 70 were investigated in the general populations (**Supplementary Table 1**). After comparison with the GWAS Catalog[20] of previously reported associations with the same or similar phenotypes in either the pregnant or the general populations (**Methods**), 1,189 of the 3,242 loci surpassing the genome-wide threshold (**Supplementary Table 8**), and 651 of the 1,999 loci surpassing the study-wide threshold were novel among nine categories (**Supplementary Table 9**). Among the three categories, 196 of the 603 loci surpassing the genome-wide threshold, as well as 64 of the 279 loci surpassing the study-wide threshold were novel (**Supplementary Table 10-11**). Notably, 5,483 and 5,194 functional variants including missense and nonsense variants ($P < 5\times10^{-8}$) were found for the 3,242 genome-wide significant loci and the 1,999 study-wide significant loci ($P < 4.1\times10^{-10}$), respectively (**Supplementary Table 14-15**). 256 study-wide newly discovered loci with at least one functional variant were illustrated in **Fig. 2**.

Among the 651 study-wide significant novel loci associated with the nine quantitative trait categories, 114 were associated with multiple gestational phenotypes (**Supplementary Table 16**). The genes mapped in these 114 pleiotropic loci were significantly enriched in GO pathway such as response to hormones, compared to an enrichment of more basic celluar processes such as hemostasis, regulation of cell-cell adhension for genes without pleiotropy

(**Supplementary Fig. 6**). Protein-protein interaction network (PPIN) analysis suggests that *IL6* involved in immune responses, *ESR1* related to hormone response and *IGF1* and *EGF* related to cell growth and development are the most essential hub genes (**Supplementary Fig. 7, Supplementary Table 17**). The *ESR1* associated with nine gestational phenotypes, is among the genetic loci with the broadest pleiotropy (**Supplementary Fig. 8, Supplementary Table 16**). The estrogen receptor 1, *ESR1* is a key gene for embryo implantation and affects uterine natural killer cell motility and thus vascular remodeling in early pregnancy and the placenta [21-23]. This is consistent with our findings of variants in *ESR1* were significantly associated with leukocytes (lymphocyte percentage (LYM_P), monocyte absolute value (MON), neutrophils absolute value (NEUT) and white blood cell count (WBC)) during pregnancy (**Supplementary Fig. 8, Supplementary Table 16**). In addition, *ESR1* is first identified to be strongly associated with triglyceride (TG), oral glucose tolerance test 0 hour (OGTT0H), red blood cell count (RBC) and hematocrit (Hct) in our study. Several other genetic loci exhibiting the highest level of pleiotropy in gestational phenotypes include *ELOVL5* (encoding ELOL fatty acid elongase 5) associated with beta human chorionic gonadotropin (FHCG), pappalysin (PAPP) and Tang screening scores of Trisomies 21 and 18, *SLC35A1* (solute carrier family 35 member A1) associated with FHCG, Tang screening score of Trisomy 21 and free thyroxine (FT4) which are important for pregnancy (**Supplementary Table 16**).

We particularly examined a number of phenotypes that had been very poorly studied, such as the screening phenotypes for fetal trisomy. Combined maternal serum Tang and NIPT screening along with fetal ultrasound testing are used to screen for the risk of fetal trisomy[24,25], so these phenotypes are essential for the gestational period. For the NIPT screening scores for T21, T18 and T13, we found 26, 35 and 31 quantitative trait loci ($P < 4.1\times10^{-10}$) (**Supplementary Table 9,**

Supplementary Fig. 9). Notably, intronic variants, eQTLs for *PANX1*, located in chromosome 11, a cell membrane factor highest expressing in human oocytes, eight-cell embryos, and the brain were significantly associated with all three scores in the NIPT screening. Abnormal Pannexin 1 (*PANX1*) has been previously linked to a type of infertility called "oocyte death" in family study[26,27]. In our study, all the variants associated with a decreasing *PANX1* increases the NIPT screening scores, suggesting higher trisomy risk. Meanwhile, variants near and within *PADI4* (peptidyl arginine deiminase 4) were similarly associated with these three phenotypes, and the gene is associated with neutrophil extracellular trap formation (NETosis) and is important for immune regulation during pregnancy[28-30]. In addition, TNF superfamily member 13b (*TNFSF13B*) and ubiquitin specific peptidase 3 (*USP3*) are associated with two of these scores, and they both play a role in the innate immune response[31,32].

Heritability and correlation of gestational phenotypes

To characterize the heritability of the 122 phenotypes, we applied LD Score regression (LDSC)[33]. Among all the phenotypes, maternal serum vitamin B12 levels had the highest SNP heritability of 43.97% (**Supplementary Table 7, Supplementary Fig. 10**). The phenotype with lowest SNP heritability was gestational thrombocytopenia, at 0.0045% (**Supplementary Table 7, Supplementary Fig. 10**). Although the heritability of vitamin B12 in the UK Biobank (UKB) (h^2=0.00953, se=0.0143) is low[34], there is a study of white British twin women that showed a 56% heritability of serum vitamin B12[35]. We hypothesize that the reason for the disparity with UKB is the effect of racial differences and gender of the population.

Phenotypic and genetic correlations were calculated for a total of 88 quantitative phenotypes in nine categories put together (**Supplementary Fig. 11, Supplementary Table 18-19,**). The two heatmaps were basically symmetric along the red diagonal

line suggesting that substantial phenotypic variability derived from the genetic variability (**Supplementary Fig. 11**). Tightly correlated phenotypic domains were clustered within categories, while significant and strong genetic correlations also occur across categories.

Gestation-specific genetic effect

Gestational phenotypes may exhibit pregnancy-specific genetic effects distinct from the same phenotypes outside pregnancy. For 23 out of the 122 phenotypes investigated, including maternal height, we obtained GWAS data in a non-pregnant cohort from the Taiwan Biobank (n = 92,615, age 30-70, southern Chinese ancestry) [19]. Genetic correlations for the same phenotype between the two datasets are all significant ranging from 0.18 to 0.94 (**Supplementary Table 20**). Correlations of genetic effect of NIPT lead SNPs are high between our study and the Taiwan Biobank study with genetic effect R^2 ranging from 0.001 to 0.99 (**Supplementary Fig. 12, Supplementary Table 21**). 900 genetic signals were present with study-wide significance for the 23 phenotypes and 792 signals have genetic effect information in both our and the Taiwan studies (**Supplementary Table 22**).

We then applied a Bayesian clustering algorithm[36] combined with colocalization analysis to explore whether there exists genetic effects that are predominantly influenced by pregnancy compared to non-pregnancy population. Model parameters are detailed in **Supplementary Table 23 and Methods.** Genetic loci associated with each of the gestational trait were categorized into two groups: general and pregnancy-specific, with the remaining loci denoted as unclassified. In total, we have identified 138 genetic signals exhibiting pregnancy specific genetic effects (**Supplementary Fig. 13, Supplementary Table 24**). The general variants indicates that the genetic effects were shared with non-pregnancy phenotypes, and pregnancy-specific cluster indicates the ones dominated by mainly gestational phenotypes. In total, we found 138 out of

the 792 signals exhibiting pregnancy-specific effects (17.4%), while 381 signals demonstrating shared genetic effect (48.1%) and 273 remains uncertain (34.5%). When ranking these 23 phenotypes by the proportion of pregnancy-specific loci among all loci, alpha fetoprotein levels (AFP_S2) or its adjusted form, the alpha-fetoprotein multiples-of-median (AFPMOM_S2) in second trimester screening have the highest proportion of variants exhibiting pregnancy-specific genetic effects, followed by two biomarkers for liver function including alanine transaminase (ALT) and albumin (ALB) (**Fig. 3A, Supplementary Table 24**). Contrastly, the lowest proportion of pregnancy specific variants were observed for maternal height and first systolic blood pressure measurements, consistent with our perception that pregnant women's height is largely unchanged during pregnancy and blood pressure stays stable in early pregnancy.

Pathway enrichment analyses of the pregnancy-specific, general and overall genetic loci were summarized in **Supplementary Fig. 14**. We further conducted a pathway enrichment analyses using genes falling within the pregnancy-specific effect and the general effect clusters to test whether some pathways may significantly differ between the two clusters of genes (two-tailed chisq or fisher exact test, $P < 0.05$) (**Fig. 3B, Supplementary Table 25)**. Among these top 20 significant pathways for the pregnancy-specific genetic loci, there are a variety of pathways that have significantly different gene proportion on both sides, such as the retinoid metabolism and transport, Wnt signaling pathway and pluripotency and regulation of protein catabolic process that only contains pregnancy-specific loci (**Fig. 3B**). In addition, pathways that demonstrated nominally significance include GO pathways of responses to hormone, glucose homeostasis and regulation of lymphocyte proliferation; KEGG pathways of breast cancer, WikiPathways of pathways in cancer and adipogenesis and Reactome gene sets of signaling by nuclear receptors (**Supplementary Table 25**). PPI network hub gene analyses weighting the genes by degree suggests that the most

distinguishing hub genes are similar to the genes that were associated with multiple gestational phenotypes, including the estrogen receptor (*ESR1*), insulin like growth factor 1 (*IGF1*) and interleukin 6 (*IL6*) (**Supplementary Fig. 15, Supplementary Table 26**). These findings suggests that hormone, growth and development mediators and immune regulations are playing an essential role for starting the gestation process.

G (genetics) by T (time) interaction across gestation and postpartum period

Among the 122 gestational phenotypes, the 24 phenotypes belonging to the blood routine category assayed by the complete blood count (CBC) [37,38] are repeatedly assayed for a median number of 5 to 6 measurements in each of the two hospitals (**Supplementary Fig. 16-17, Supplementary Table 27**), offering us an opportunity to understand the patterns and biological meaning of potential longitudinal time-varying genetic effect.

The changing trends of these 24 phenotypes during pregnancy were largely similar between the two hospitals (**Supplementary Fig. 18-19**). Among the leukocyte-related phenotypes, white blood cell (WBC), neutrophil counts (NEUT), and neutrophil percentages (NEU_P) increased significantly during pregnancy. In addition, lymphocyte (LYM) levels decreased, monocyte (MON) levels increased, eosinophils (EOS) were essentially unchanged and basophils (BASO) fluctuated at average levels. Red blood cell counts (RBC), hemoglobin (HGB) levels, hematocrit (Hct) and platelet count (PLT) have a significant downward trend during pregnancy. All CBC phenotypes changed greatly in the delivery and postpartum period, tending to return to pre-pregnancy levels.

We investigated the longitudinal time-varying genetic effect following two steps. First, we applied TrajGWAS[39] to identify genetic variants affecting the phenotypic mean

variability across the 24 CBC gestational phenotypes for each hospital, respectively. By default, TrajGWAS[39] will also compute the effect of genetic variants affecting the within-subject variability. Genetic variants affecting phenotypic mean and within-subject variability (β_g and τ_g significant genetic variants) are shown in the **Supplementary Fig. 20-27**. In total, we identified 645 variants associated with the phenotypic mean reaching a genome-wide significance threshold in both hospitals ($P < 5\times10^{-8}$). Secondly, we included a genotype dosage and gestational day interaction term in the above TrajGWAS regression model for each hospital, testing the statistical significance of the interaction effect size, similar to a locus-based genetic-by-time ($G \times T$) analysis (**Methods**). After integrating the two hospital statistics via inverse-weighted variance fixed effect meta-analysis, 115 out of 645 the variants (17.8%) demonstrated significant interaction effect with gestational days ($P < 5\times10^{-8}$) (**Supplementary Table 28**), while the rest of the 530 variants only affected the phenotypic mean level but the interaction effects were not significant (**Supplementary Table 29**). Among the 23 phenotypes, the percentage of variants with significant $G \times T$ effects ranged from 5.4% to 47.8% (**Supplementary Table 30**).

To explore how genetic variants specifically change during the first, second, third trimesters of pregnancy, delivery, and postpartum period, we performed localized association studies of these 115 variants (**Methods, Supplementary Table 31**) and found 79 out of the 115 variants exhibited non-overlapping 95%CIs for genetic effects in at least two of the five periods (**Supplementary Table 31, Supplementary Fig. 28**). Therefore, we confirm that 79 out of the 645 variants (12.2%) demonstrated substantial and significant changes of genetic effects across the gestation and postpartum period. The 25 variants with the most significant $G \times T$ interaction effects were demonstrated in **Fig. 4A.** 22 out of these 25 variants have strong expression quantitative trait loci (eQTL) information (**Supplementary Table 31**). Taking the

gestational WBC (leukocyte) count as an example, the WBC count steadily increased since the first trimester and reached its highest in the 3rd trimester and delivery (**Fig. 4B**). Among the 27 variants associated with WBC, eight variants (29.6%) exhibiting significant $G \times T$ interaction. Among the eight variants, rs2270401 (17q21.1), a strong eQTL for *GSDMA*, exhibited the most significant $G \times T$ interaction effect, with genetic effects varied substantially across the five periods (first trimester: β=0.270, P=6.44×10^{-491}; second trimester: β=0.396, P=1.62×10^{-1223}; third trimester: β=0.388, P=2.27×10^{-1021}; delivery: β=0.257, P=2.06×10^{-152}; postpartum: β=0.122, P=6.48×10^{-60}) (**Fig. 4C, Supplementary Table 31**). The genetic effect of the rs2270401-A allele was 3.33 times greater in the second trimester than in the postpartum period. Genetic effect of the other seven loci also demonstrated a U-shaped change across the five period (**Supplementary Fig. 28**). The WBC phenotypic and genetic effect changes of the eight loci suggest these loci may play an essential role in the rapid growth of leukocytes during pregnancy and postpartum.

We extracted genetic variants with significant versus non-significant genetic and environmental interactions for enrichment analyses. Gene enrichment analysis for each category of variants were summarized in **Supplementary Fig. 29**. In a chisq and fisher exact test for testing significant difference of pathways with interaction effect and those without, we found pathways enriched in regulation of phagocytosis, engulfment, prefoldin mediated transfer of substrate to CCT/TriC, Heparan sulfate/heparin (HS-GAG) metabolism, Gap junction, nerve development and lipid modification (**Fig. 4D, Supplementary Table 32**). Different from the genes harboring gestation-specific genetic effects, PPI network hub gene analyses weighting the genes by degree suggests that the most distinguishing hub gene is Epidermal Growth Factor (*EGF*) involved in cell growth and development and the GATA binding protein 2 (*GATA2*) which involved in hematopoietic development (**Supplementary Fig. 30, Supplementary Table 33**). This suggests that alterations

in the genetic effects of variants affecting blood cell levels during pregnancy are more related to changes in cell activation and cytokines in pregnant women.

Causal relationship with older-age phenotypes

To explore whether gestational phenome may influence or correlated with the occurrence of chronic diseases in mothers in the distant future, we first estimated causal effects of the 122 gestational phenotypes on 220 phenotypes examined in the BioBank Japan Study (BBJ)[10] using two-sample, inverse variance weighted, bidirectional MR and four additional two-sample approaches as sensitivity analyses. After MR-PRESSO elimination of outliers[40], 73 significant potential causal relationships were found for the 122 gestational phenotypes, using FDR *q*-value less than 0.01 as false discovery rate (FDR) threshold with 1,764 non-collinear phenotypic pairs and excluding Cochran's Q statistic or pleiotropy *P*-value less than 0.05, and these causal relationships were not significant in the reverse MR (Methods). The bidirectional causal effect estimates were presented in a network view in **Fig. 5A,** with the forward MR effect estimates of 88 molecular phenotypes presented in **Fig. 5B and Supplementary Table 34** and the forward MR effect estimates of the other 34 phenotypes presented in **Supplementary Fig. 31 and Supplementary Table 35**.

We identified numerous previously known and unidentified potential causal associations. Genetically predicated higher gestational BMI contributes to the most of the older-age disorders and phenotypes in our study. It increasing liability of 13 older-age disorders or phenotypes with the largest effects observed on chronic renal failure (*OR* [95%CI]: 1.68 [1.42-2.00]), followed by several cardiovascular disorders such as myocardial infarction, peripheral artery disease, chronic heart failure, unstable angina pectoris, type 1 diabetes and cholelithiasis as well as five medication use of drugs and three physical phenotypes, while gestational BMI slightly decreasing liability of two

older-age disorders including pneumothorax and pulmonary tuberculosis (**Fig. 5A, Supplementary Fig. 31 and Supplementary Tables 35**).

In addition, our MR results suggest contribution of gestational blood lipid and glycemic biomarkers causally associated with older-age disorders. For example, genetically predicted elevated triglyceride levels in pregnant women is associated with an increased risk of taking certain medications, such as the medication use of antithrombotic agents (*OR* [95%CI]: 1.09 [1.06-1.12], $P=1.24\times10^{-11}$) (**Fig. 5C, Supplementary Table 34**). In addition, genetically predicted elevated 1h blood glucose levels on the oral glucose tolerance test (OGTT1H) in pregnant women increase the risk of peripheral arterial disease (PAD) (*OR* [95%CI]: 1.24 [1.13-1.37], $P=1.24\times10^{-5}$) (**Fig. 5D, Supplementary Table 34**), consistent with prior report of glucose abnormalities as a risk factor for PAD[41]. Moreover, elevated levels of genetically predicted OGTT1H also increase the risk of cataract (*OR* [95%CI]: 1.15 [1.09-1.21], $P=3.21\times10^{-8}$) (**Supplementary Table 34**), which is in accordance with the results of the animal model[42].

Furthermore, our results also show that genetically increased infections increase certain disease liability. For instance, genetically increased gestational hepatitis B core antibody positive (HBcAb) increases the risk of gastric cancer (*OR* [95%CI]: 1.12 [1.06-1.19], $P=1.06\times10^{-4}$), chronic hepatitis B infection (*OR* [95%CI]: 1.31 [1.15-1.49], $P=5.27\times10^{-5}$) and hepatic cancer (*OR* [95%CI]: 1.24 [1.13-1.37], $P=5.02\times10^{-6}$) (**Fig. 5A, Supplementary Table 34**), revealing a potential causal relationship between hepatitis B infection observed during pregnancy and extrahepatic cancers. Moreover, genetically predicted quantitative levels of cytomegalovirus IgG (CMV_IgG_quan) were found to be associated with an increased risk of some cardiac diseases (angina pectoris: *OR* [95%CI]: 1.31 [1.16-1.49], $P=1.28\times10^{-5}$; stable angina pectoris: *OR* [95%CI]: 1.27 [1.16-1.40]; medication

use vasodilators used in cardiac diseases: *OR* [95%CI]: 1.26 [1.15-1.38]). This is consistent with previous studies suggesting increased exposure to cytomegalovirus infection is related to coronary artery and vascular atherosclerotic diseases[43,44]. More broadly, we found that genetically predicted increases in BMI during pregnancy were associated with an increased risk of developing a variety of diseases, such as chronic kidney failure (*OR* [95%CI]: 1.68 [1.42-2.00]) and myocardial infarction (*OR* [95%CI]: 1.26 [1.17-1.35]) (**Fig. 5A, Supplementary Table 34, Supplementary Fig. 31**). In general, a less superior physiological state during pregnancy is likely to threaten the future health of the woman.

The causal impact of the gestational phenotypes (collected from pregnancy women with mean age of 30) may directly contribute to the later-age phenotypes (collected from BBJ with ages ranging from 30 to 70) or the causal effect may be mediated through the same phenotypes in a middle age period. To understand this process, we utilized the previously analyzed IVs with pregnancy-specific and shared effects with the 23 phenotypes shared with the Taiwan Biobank, we added GWAS data from the Taiwan Biobank as a second exposure and conducted a multivariable MR (MVMR). Using FDR q value as the threshold of significance (NIPT effect $q < 0.01$), we identified a total of 45 significant causal effects of the gestational phenotypes on the later-age outcomes even after adjusting for the genetic effects of the exposures in the Taiwan Biobank. Among these, 40 effects have a q value > 0.05 in the Taiwan dataset or having an opposite genetic effect compared to ones with the gestational phenotypes suggesting substantial direct causal effects on the outcomes. The remaining 5 effects have a nominally significant P and consistent effect direction in the Taiwan dataset, indicating less strong direct causal effects of the gestational phenotypes on the later age phenotypes (**Supplementary Table 40, Supplementary Fig. 32**). In addition, there are 97 effects having a q value < 0.01 for the Taiwan while having an opposite beta direction or a q value > 0.05 in the NIPT dataset, suggesting that the NIPT

effects on the outcome is primarily dominated by the middle and older age genetic effects. Among the 40 out of the 143 effects (28.0%) with potential direct causal effects, notable relationships include genetically predicted gestational AFP and AFPMOM in the second trimesters is negatively associated with atopic dermatitis and all types of white blood cell, while it is positively associated with thyroid disorders such as Grave disease, Hashimoto thyroiditis as well as medication use thyroid preparations, chronic sinusitis, Type 2 diabetes and medication use drugs in diabetes (**Supplementary Fig. 32 A-B, Supplementary Table 40**). In addition, genetically predicted albumin during pregnancy is negatively associated with chronic hepatitis B infection and cirrhosis. AST and RBC from NIPT datasets are causally associated with the related blood cell information in the BBJ (**Supplementary Fig. 32 C, Supplementary Table 40**). We also compared the results of univariate MR between the 23 gestational phenotypes and the Taiwanese phenotype, and the results have high consistency with an R^2 of 0.79, consistent with the above statistics where the majority, namely 97 versus 45 of the gestational phenotypic effects on outcomes through older-age phenotypes, and through direct effects (**Supplementary Fig. 33, Supplementary Table 41-42**).

Discussion

Our study represents the largest scale GWAS on gestational phenotypes to date and provides the first insight into the dynamic genetic atlas of human gestational phenotypes and their causal associations with older-age health conditions. Using obstetric and NIPT genotype data from 121,579 pregnant women, we identified 3,845 genome-wide significant loci, of which 1,385 loci were novel to this study and the majority of which consists of at least one function variant. More than 84.5% of the genetic associations were replicated in independent cohorts. Genes associated with multiple gestational phenotypes are enriched in response to hormones, compound biosynthetic and multi-organism reproductive processes, and PPI network analysis

reveals that *IL6* involved in immune responses, *ESR1* related to hormone response and *IGF1* and *EGF* related to cell growth and development are the most essential hub genes, providing the first insights into the role of immune, hormone and cell growth responses and the key gene players in gestational process from a human genetic perspective, beyond previous insights achieved via metabolome of primate pregnancy [5].

Notably, the genetic atlas of gestational phenotypes is dynamic, reflecting gestation-specific $G \times E$ and gestation time-varying $G \times T$ interactions. Combining data from the Taiwan Biobank and using Bayesian classification analysis, we identified pervasive gestation-specific genetic effects that present in around 17.4% of the genetic associations (138/792) with 23 gestational phenotypes, ranging from ~0% for early pregnancy systolic blood pressure to 100% for alpha-fetoprotein in second trimester. Genes containing gestation-specific genetic effect exhibit a predominant enrichment in pathways involving cancer, response to hormones, lipid, vitamin and glucose metabolisms, and cell growth and development. PPI network analysis highlights *ESR1*, *IGF1* (but not *EGF*) and *IL6* as the most significant hub genes.

On the other hand, in a longitudinal trajectory GWAS analysis of repeated measurements of 24 complete blood count (CBC) phenotypes, we found significant $G \times T$ interactions present in 12.2% of the genetic variants (79/646) throughout the five gestational, delivery and postpartum periods, with hub genes as *EGF* and *CATA2* in the PPIN, and pathways enriched in regulation of cell growth activity, nerve development and lipid modification.

Conditions during gestational period were deems as unique opportunities for early screening and prediction for late-onset chronic disease risk in women[45]. Our results suggest causal relationships between adverse gestational conditions and

cardiovascular disease, including obesity, dyslipidemia and dysglycemia. We also found that genetically predicted blood glucose abnormalities during pregnancy were associated with an increased risk of cataracts. Previous study has shown that pregnant women with gestational diabetes mellitus, even if they do not have type 2 diabetes, have an increased risk of developing cataracts[46]. In addition, multivariable Mendelian randomization results indicate that high AFP levels during pregnancy are likely to lead to inflammation-related diseases and diabetes, adjusted for mid-life health status. Thus, the state of pregnancy is related to women's health in the long term.

There are some limitations which are connected to future perspectives in our study. First, although the current study represents the largest GWAS for the majority of the phenotype, the current study participants are restricted to East-Asian participants, and the genetic novel discoveries may be different among non-East-Asian population. However, the study methods made available in this study provides a foundation for expanding the participants from non-East-Asian ancestry, which is promising as NIPT tests as well as pregnancy screening programs are undertaken globally around the world. The future comparisons of European and East-Asian gestational phenotypes may lead to new insights into the ethnic difference for gestational phenotypes and its relationship to offspring's health[47]. In addition, the cross-ancestral analysis will facilitate fine-mapping and validation of the functional variants identified in this study. Secondly, while we provide a dynamic view of the genetic atlas of the gestational phenotypes based on a robust statistical framework, unravelling 12 to 17% of gestation-specific genetic effects and time-varying genetic effects, unravelling novel genes involved in immune and hormone response, and cell-growth and development in determining phenotypic distribution specifically during gestation. How these genes may influence and modulate how gestation initializes and progresses, and how they might contribute to abnormal gestations requires further comparisons or investigations in the patient-oriented study and more in-depth mechanistic study conditioning on the

findings in this study. Thirdly, beyond the genetic and causal relationships between gestational phenotypes and broad human complex and older age phenotypes investigated in our study, another important investigation will be the dissection of fetal genetic and intrauterine effects on birth outcomes and offspring health conditioning on the dynamic atlas presented here and expanding birth cohort genome studies[14].

The genetic discoveries in this study provides foundational insights into the dynamics of genetic landscape of the human gestational period. As we continue to expand the genomic data and methods, future efforts will be prioritized to combine the genetic information to dissect the developmental origin of children and adult disorders and for exploring health and risk prediction for women during the gestational time window. These should be conducted in global efforts.

Methods

Study population

From 2017 to 2022, we recruited 121,579 pregnant women who performed maternity checkups in Shenzhen Baoan Women's and Children's Hospital and Longgang District Maternity and Child Healthcare Hospital of Shenzhen City into our study. Pregnant women received prenatal and postnatal serum biomarkers testing and non-invasive prenatal testing (NIPT) in the first or second trimester or both of pregnancy. Maternal genotypes were inferred from low-depth whole-genome sequencing data generated from NIPT using our previously developed advanced methods[7]. In addition, we also collected basic physical measurements of participants and their babies' birth outcomes. The details of the number of participants in each test are presented in Supplementary Table (**Supplementary Table 2-3**). The age distribution of pregnant women and the

prevalence of some gestational comorbidities in the two hospitals are shown in the supplementary material (**Supplementary Table 4-5, Supplementary Fig. 1-2**).

We integrated all the obstetrical data, including genotypic and phenotypic data, to explore the genetic and molecular risk factors underlying the phenotypic spectrum of pregnant biomarkers. This study was approved by the Medical Ethics Committee of the School of Public Health (Shenzhen), Sun Yat-sen University, Longgang District Maternity and Child Healthcare Hospital of Shenzhen City, and Shenzhen Baoan Women's and Children's Hospital. Data collection was approved by the Human Genetic Resources Administration of China (HGRAC).

Phenotype definition

In the genome-wide association study (GWAS) analyzing gestational phenotypes, we collected 122 maternal phenotypes in nine categories concerning Blood lipid (N=2), Glycemic (N=6), Blood routine (N=32), Infection (N=14), Kidney function (N=3), Liver function (N=8), Thyroid function (N=4), Tang screening (N=16) and NIPT screening (N=3), as well as three categories including Basic information (N=7) , Gestational disorders (N=21) and Birth outcomes (N=6) (**Supplementary Table 2-3**). For phenotypes with multiple measurements, we filtered out the record with the earliest measurement date. In longitudinal genetic analyses, we retained all measurements of complete blood count (CBC).

Genotype imputation from Non-Invasive Prenatal Testing data and quality control

NIPT for fetal trisomy by sequencing maternal plasma free DNA (cfDNA) is new genomic sequencing technology[48]. Large-scale low-pass and mediate-coverage sequencing data are available through NIPT. In our previous study, we demonstrated that NIPT data can be used effectively for GWAS analysis[7]. Therefore, it is robust that

we use imputed NIPT data for genetic basis of prenatal and postnatal screening indicators.

Genotype imputation from NIPT data was conducted by genotype likelihoods imputation and phasing method using GLIMPSE software (version 1.1.1)[49] with 10,000 Chinese population high-depth sequencing genotype as reference panel[50]. Loci with minor allele frequency less than 0.001 were filtered in advance for reference panel. To address confounding by population relatives, we excluded the duplicated sequencing records of the same people from different pregnant times and kept the highest sequencing depth.

Family relatedness and principal component analyses

PLINK[51] (version 2.0) was used to select SNPs with an MAF of at least 5%. The kinship was calculated based on the KING-robust kinship estimator. We used a cutoff of approximately 0.354 to identified monozygotic twins and duplicate samples. In addition, we applied PLINK[51] (version 2.0) for principal component analyses (PCA) both before and after genotype imputation. EMU[52] (version 0.9) was used for PCA specifically before genotype imputation.

Variant annotation

We performed variant annotation using Ensembl Variant Effect Predictor (VEP) (version 101)[53], with indexed GRCh38 cache files (version 109). HGVS notations were generated by primary assembled reference FASTA files for *Homo sapiens*. All these data for annotation were pre-downloaded from the Ensembl FTP server (https://ftp.ensembl.org/pub/). Since a variant may overlap with multiple transcripts, we used the --pick option to assign one block of results to each variant based on a set of VEP default criteria. For variants located in intergenic regions, we used the --

nearest option to identify the nearest gene to the protein-coding transcription start site (TSS).

Genome-wide association analysis and conditional analysis

After quality control of the sequencing data and imputation of genotype from the NIPT data, we conducted genome-wide association analysis using PLINK[51] (version 2.0) to examine the association of the 122 phenotypes, assuming additive genotype effects. For quantitative traits, we utilized linear regression model for GWAS, and for qualitative traits, we used logistic regression model. To account for population stratification, we considered some or all the following as covariates in the GWAS for all phenotypes: maternal age, body mass index (BMI), gestational week at the time of indicators measurement, and the top ten principal components. All GWAS covariates were shown in the **Supplementary Table 6**.

Using the GWAS summary statistics from two hospitals, we performed meta-analyses using METAL[54] (version 2011-03-25) with an inverse-variance weighted (IVW) fixed-effect model. To visualize the meta-analyses results, we generated fuji plot using R script and Circos software[55,56]. Then, we used the GWAS meta summary statistics for all the subsequent analyses.

To identify independent genome-wide significant signals ($P < 5 \times 10^{-8}$), we used GCTA[57] multi-SNP-based conditional & joint association analysis (COJO[18]) with a stepwise model selection (--cojo-slct). Variants with MAF < 0.01 were filtered and the colinearity threshold was 0.2. 500kb upstream and downstream block of a genome-wide significant signal was divided into a separate locus. The SNP with the smallest *P* value in each locus was selected as the lead SNP. In addition, if the SNP had *P* value less than 4.1×10^{-10}, it was considered to be a variant that reached the study-wide significance. Similarly, we used the same method to select independent

SNPs from the 220 GWAS summary statistics of BBJ cohort[10] and 23 GWAS summary statistics of Taiwan Biobank[19] for subsequent analysis.

Replication of independent genome-wide significant signals

To examine whether genetic effects of the same phenotype were equivalent between the two cohorts, we performed internal replication and external comparison of independent genome-wide significant SNPs obtained from conditional analysis using two methods. In the internal replication, we compared the direction of beta and *P* value of each SNP in the two hospitals. The criteria for a SNP to be replicated were that it had consistent beta direction and the *P* value was less than the Bonferroni-corrected significant threshold (0.05/number of independent SNPs per phenotype). In the second approach, based on the central limit theorem, we utilized a *two-sample t-test* with the following hypotheses:

$$Null\ hypothesis\ H_0\colon \beta_1 = \beta_2$$

$$Alternative\ hypothesis\ H_1\colon \beta_1 \neq \beta_2$$

The statistic *T* was calculated as follows:

$$T = \frac{\beta_1 - \beta_2}{\sqrt{SE_1^2 + SE_2^2}} \sim t(\upsilon') \quad (1)$$

The degrees of freedom υ' were calculated by the formula:

$$\upsilon' = \frac{(SE_1^2 + SE_2^2)^2}{\frac{SE_1^4}{n_1 - 1} + \frac{SE_2^4}{n_2 - 1}} \quad (2)$$

In the formulas, β represents the estimate of the genetic effect, *SE* represents the standard error of the genetic effect estimate, and *N* represents the sample size of trait int the cohort. The *P* value was calculated using the statistic *T* and degrees of freedom υ' obtained from the formulas, and the SNP was replicated successfully if the *P* value was greater than 0.05. In addition, we used METAL to estimate heterogeneity tests for the genetic effects of SNPs between the two hospitals, and *P* value ≥0.05 were taken

687 to indicate that there was no heterogeneity in the genetic effects of the SNP between
688 the two hospitals.
689
690 We conducted external replication of the identified variants using the NIPT PLUS
691 pregnancy cohort, which is a data collection of 5,733 individuals who sought maternal
692 check-ups at Shenzhen Baoan Women's and Children's Hospital (Shenzhen, China)
693 throughout the entire 40-week gestational period. Between 2020 and 2021, these
694 pregnant women received NIPT in the first or second trimester and provided written
695 informed consent. Besides, these samples were sequenced at a deeper depth compared
696 to traditional NIPT. The genotyping and quality control process of NIPT PLUS cohort
697 participants were the same as Baoan and Longgang. Among the 122 phenotypes in
698 this study, 103 phenotypes could be replicated in the NIPT PLUS pregnancy cohort.
699 We used the same methods as for the internal replication to perform external
700 replication of the variants for these 103 phenotypes.
701
702 **Comparison of genetic effects on maternal height and Taiwanese**
703 **population height**
704 As maternal height is essentially unchanged during pregnancy, we demonstrate the
705 effectiveness of our GWAS results by comparing the genetic effects of height in the
706 Taiwanese population with the genetic effects of maternal height related variants[19].
707 We extracted genome-wide significant independent variants associated with height in
708 Taiwanese population and found corresponding genetic effects in our GWAS
709 summary statistic. Comparisons were made using the Bonferroni correction threshold
710 method and the rate of success in the comparison was calculated.
711
712 **Identification of novel locus and signal**
713 We downloaded all associations (v1.0.2_e109_r2023-02-15) from GWAS Catalog[20]
714 to identify novel associated SNPs. A trait associated locus was considered novel when

the ±500kb range of the lead SNP did not include previously reported associations in the GWAS Catalog for the same trait. In addition, we considered a trait-associated signal as novel when it satisfied that the R^2 between the GWAS Catalog reported SNPs and our SNP was less than 0.2. The R^2 calculations were performed on the LDpair Tool through LDlink[58] (version 5.5.1) based on GRCh38 1000G genome build in the EAS and EUR populations.

Protein-protein interaction network analysis

We selected genes hosting loci in our study for protein-protein interaction (PPI) analysis. Genes with PPI score > 0.4 were chosen to build a network model visualized by Cytoscape (Version:3.10.1)[59]. The degree algorithm implemented in CytoHubba[60] was used to compute the top 10 hub genes. Software parameters used default parameters.

Genetic heritability and correlation estimates

We standardized the phenotypic data and calculated their Spearman correlation coefficients to obtain the phenotypic correlation matrix. Then the meta GWAS summary statistics were used to estimate SNP-based heritability and genetic correlation (r_g) for all phenotypes using LD Score Regression[33,61]. GCTA[62] was used to calculate LD score files using 10,000 Chinese population high-depth sequencing genotype. The genome was chopped into segments with length of 1000kb (--ld-wind 1000), and the cutoff for LD R^2 was 0.01 (--ld-rsq-cutoff 0.01). The phenotypic correlation matrix and genetic association matrix of the two hospitals were extracted to draw heat maps, respectively.

Detection of pregnancy-specific effects

Because of the paucity of currently available GWAS summary statistics for pregnancy and the influence of population structure, we used GWAS summary statistics for

743 similar phenotypes from the Taiwan Biobank[19], which were downloaded for external
744 comparisons. Three conditions were required for an independent SNP to be compared:
745 1) the SNP was present in the externally compared GWAS summary statistic; 2) the
746 effect allele and the direction of the genetic effect were the same as those of the
747 externally data; and 3) the *P* value was less than the Bonferroni-corrected threshold
748 (0.05/number of loci).
749
750 We applied linemodels package[36] (https://github.com/mjpirinen/linemodels) to the
751 GWAS summary statistics from gestational phenotypes GWAS and similar
752 phenotypes GWAS in the general population of East Asia. The analysis included
753 1,318 variants in 23 phenotypes (**Supplementary Table 23**). We classified the
754 variants into two clusters based on their effect sizes, pregnancy specific variants and
755 general variants. If the genetic effect of the variant is more different from that of the
756 general population in East Asia, it is called a pregnancy specific variant (posterior
757 probability of pregnancy > 0.95), and if the genetic effect is more consistent between
758 the two, it is called a general variant (posterior probability of general > 0.95). The
759 clusters were represented by line models where the slope of the pregnancy cluster is
760 fixed to 0, the correlations are fixed to 0.999 in both clusters, and the other parameters
761 are estimated using EM algorithm. The parameters of the various categories are
762 shown in **Supplementary Table 23**. The slope represents the slopes of the two lines,
763 and the scale determines the magnitude of effect sizes, and the correlation determines
764 the allowed deviation from the lines.
765
766 To exclude the effect of LD structure on classification, we performed colocalization
767 analyses of variants in gestational phenotypes and variants in East Asian populations
768 with the same or similar phenotypes and adjusted the linemodes results depending on
769 colocalization results. We set the prior probability of an SNP associated with
770 phenotypes of pregnant women (p_1) and in East Asia (p_2) as 1×10^{-4}, and the prior

probability of an SNP associated with both traits (p_{12}) as 5×10^{-6} [63]. We adjusted the classification results by adopting the following criteria. If a variant classified as a pregnancy cluster in line models had a posterior probability (PPH4) < 0.2, the classification result remained unchanged; otherwise, it was classified as an uncertain cluster; if a variant in a general cluster had a PPH4 > 0.8, the classification result remained unchanged; otherwise, it was classified as an uncertain cluster; if a variant classified as a uncertain cluster in line models had a PPH4 < 0.2, it was classified as an pregnancy cluster; if a variant classified as a uncertain cluster in line models had a PPH4 > 0.8, it was classified as an general cluster. The results of the classification of the other cases are consistent with line models.

Enrichment analysis of variants

Based on the results of line models, we performed enrichment analysis using Metascape (v3.5.20240101)[64]. This method calculates the hit rate with our gene list and the background hit rate based on the number of genes in each section, and the ratio of the two is the enrichment factor. The *P*-value is defined as the probability of obtaining multiple pathways, forming a cumulative hypergeometric distribution, the more negative the value, the less likely it is that the observed enrichment is due to randomness. The analysis was performed with default parameters, namely, the min overlap was set to 3, the *P* value cutoff was 0.01, the min enrichment factor was 1.5. Ontology category included GO Biological Processes, Reactome Gene Sets, KEGG Pathway, WikiPathways, Canonical Pathways and PANTHER Pathway.

We performed separate enrichment analyses for genes with pregnancy-specific variants and genes with general variants, selecting the top 20 pathways with the most significant pregnancy-specific variants and comparing them with the general variants. We also aggregated all genes for enrichment analysis of single gene lists.

Longitudinal trajectories of genetic effects

We conducted analyses using TrajGWAS[39] to identify genetic variants that contribute to change in gestational phenotypes during pregnancy and postpartum. We used gestational day as a time factor to determine whether there was a genetic-time interaction. GWAS for longitudinal biomarkers was performed using the TrajGWAS method, based on a mixed effects model with the following formula:

$$y_{ij} = x_{ij}\beta + g_i\beta_g + G \times T \times \beta_{G\times T} + \varepsilon_{ij} \quad (3)$$

y_{ij} represents the phenotypic value of the jth measurement for the ith pregnant woman. x_{ij} represents fixed-effects covariates, including the first ten principal components, maternal age and gestational days at phenotypic testing, β is its corresponding effect value. g_i denotes the genotype dosage for the ith pregnant woman, corresponding to a genetic effect of β_g. $G \times T$ represents the interaction between maternal genotype dosage and gestational time, corresponding to an effect value of $\beta_{G\times T}$. ε_{ij} is a random-effects term reflecting the variance of the genetic variants affecting the within-subject (WS) variance of the biomarker. The formula for the variance $\sigma^2_{\varepsilon_{ij}}$ of ε_{ij} is given below:

$$\sigma^2_{\varepsilon_{ij}} = \exp\left(\tau w_{ij} + \tau_g g_i + \tau_{G\times T} G \times T + \omega_i\right) \quad (4)$$

w_{ij} is a covariate including the first 10 principal components, maternal age, and gestational days at phenotypic measurement, and τ is the corresponding effect value. τ_g denotes the genetic effect affecting the variance of WS. $\tau_{G\times T}$ indicates the effect value for the interaction of maternal genotype dosage and gestational time. ω_i denotes the random intercept.

There were 24 of these phenotypes, involving phenotypes of the Blood routine categories (N=24). Specific phenotypes are shown in **Supplementary Table 27**. Following the procedure described by Ko et al., we first performed the score test for SNPs, obtaining the direction of the effect (β_g or τ_g) and the *P* value that affected the mean and within-subject variability for each phenotype. Subsequently, the Wald test

was performed on SNPs with a *P* value of β_g or τ_g less than 5×10^{-8} to estimate the effect sizes of the genetic effect of genotype dosage (β_g or τ_g) and the effect of interaction between maternal genotype dosage and gestational time ($\beta_{G\times T}$ or $\tau_{G\times T}$). Subsequently, we performed a meta-analysis of the effect values of these significant SNPs using METAL[54]. TrajGWAS analyses were performed with the Julia package TrajGWAS.jl (https://github.com/OpenMendel/TrajGWAS.jl)[65]. Manhattan and QQ plots in the supplementary figures were drawn with a Julia package MendelPlots.jl (https://github.com/OpenMendel/MendelPlots.jl)[65] with SNP minor allele frequency>0.01. To identify independent significant loci, we used a python script where the variant with the smallest *P* value in the 1Mb range was the lead SNP, and variants within 500kb above and below the lead variant were classified into a locus.

In order to investigate whether the genetic effects of the loci identified by TrajGWAS change during a total of five periods: first, second, and third trimester of pregnancy, day of delivery, and within 42 days postpartum, we performed localized association analyses of these loci. The main steps were consistent with the previous GWAS, except that the PLINK[51] (version 2.0) parameters --chr, --from-bp, and --to-bp were used to point to the loci. We defined a variant with change as a non-overlapping 95% confidence interval (CI) for the genetic effect in at least two of the five periods.

We enriched the variants with significant genotype dosage for significant and non-significant interactions between genetics and environment. Methods were the same as for the enrichment analyses described above.

Mendelian randomization network of pregnancy, Taiwanese and BBJ traits

To identify whether pregnancy status affects the development of future disease progression, bidirectional mendelian randomization analyses were performed on

cohort of this study and Japanese Biobank cohort data with the TwoSampleMR package[66,67]. We used GWAS summary statistics from two cohorts and independent SNPs obtained from GCTA-COJO as genetic instruments ($P < 5\times10^{-8}$) in our MR analyses. We also calculated the proportion of variance in phenotypes explained (PVE) by the instruments according to the formula[68] (3) and calculated the F-statistic according to formula (4) to evaluate the strength of instruments, with F values greater than 10 indicating sufficient strength.

$$PVE = \frac{2\hat{\beta}^2 MAF(1-MAF)}{2\hat{\beta}^2 MAF(1-MAF) + \left(se\left(\hat{\beta}\right)\right)^2 2NMAF(1-MAF)} \quad (5)$$

$$F = \frac{N-nsnp-1}{nsnp} \times \frac{PVE}{1-PVE} \quad (6)$$

The assessments of the instrumental variables are shown in the **Supplementary Table 36-39**. We conducted sensitivity analyses using alternative MR methods that including: 1) MR Egger; 2) Simple mode; 3) Weighted median; 4) Weighted mode. Among the several methods of MR analyses, the inverse variance weight (IVW) method which confers the greatest statistical power for estimating causal associations was used as primary reference[69].

To examine and address horizontal pleiotropy in causal relationships inferred from MR analyses, we performed MR-PRESSO[40] using the results of the first two-sample MR. The outliers were found using MR-PRESSO and tested for differences between the pre-correction and post-correction results. The outliers would be excluded, and the two-sample MR was performed again if the difference was significant.

Robust causal estimates were defined as those that were significant at FDR q value of IVW[70] ($q < 0.01$) and showed no evidence of heterogeneity ($P > 0.05$) and horizontal pleiotropy ($P > 0.05$) after MR-PRESSO.

We also performed two-sample MR of 23 Taiwanese phenotypes with 220 BBJ phenotypes. The methodology is the same as in the previous section. Genetic instrumental variable information is displayed in **Supplementary Table 43-44**.

To adjust for the influence of mid-life health status on causal relationships, we incorporated 23 phenotypes from the Taiwan Biobank[19] as a second exposure in the multivariable mendelian randomization analysis[71,72]. The gestational phenotype and the Taiwan Biobank phenotype were used as exposures, and the 220 phenotypes from the BBJ were used as outcomes. We also used the TwoSampleMR package. The significance threshold is $q < 0.01$.

Reference

1. Stock, S.J. & Aiken, C.E. Barriers to progress in pregnancy research: How can we break through? *Science* **380**, 150-153 (2023).
2. Qiao, J. *et al.* A Lancet Commission on 70 years of women's reproductive, maternal, newborn, child, and adolescent health in China. *Lancet* **397**, 2497-2536 (2021).
3. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950-2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1160-1203 (2020).
4. Chen, X. *et al.* The path to healthy ageing in China: a Peking University-Lancet Commission. *Lancet* **400**, 1967-2006 (2022).
5. Yu, D. *et al.* A multi-tissue metabolome atlas of primate pregnancy. *Cell* **187**, 764-781.e14 (2024).
6. Hunter, P.J. *et al.* Biological and pathological mechanisms leading to the birth of a small vulnerable newborn. *Lancet* **401**, 1720-1732 (2023).
7. Liu, S. *et al.* Genomic Analyses from Non-invasive Prenatal Testing Reveal Genetic Associations, Patterns of Viral Infections, and Chinese Population History. *Cell* **175**, 347-359.e14 (2018).
8. Zhen, J. *et al.* Genome-wide association and Mendelian randomisation analysis among 30,699 Chinese pregnant women identifies novel genetic and molecular risk factors for gestational diabetes and glycaemic traits. *Diabetologia* **67**, 703-713 (2024).
9. Yang, Z. *et al.* Genetic Basis of Altered Platelet Counts and Gestational Thrombocytopenia in Pregnancy. *Blood* (2023).
10. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415-1424 (2021).

11. Graham, S.E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675-679 (2021).
12. Elliott, A. *et al.* Distinct and shared genetic architectures of gestational diabetes mellitus and type 2 diabetes. *Nat Genet* **56**, 377-382 (2024).
13. Juliusdottir, T. *et al.* Distinction between the effects of parental and fetal genomes on fetal growth. *Nat Genet* **53**, 1135-1142 (2021).
14. Huang, S. *et al.* The Born in Guangzhou Cohort Study enables generational genetic discoveries. *Nature* **626**, 565-573 (2024).
15. Mitha, A. *et al.* Neurological development in children born moderately or late preterm: national cohort study. *Bmj* **384**, e075630 (2024).
16. Crump, C. *et al.* Adverse pregnancy outcomes and long term risk of ischemic heart disease in mothers: national cohort and co-sibling study. *Bmj* **380**, e072112 (2023).
17. Liu, S. *et al.* Utilizing Non-Invasive Prenatal Test Sequencing Data Resource for Human Genetic Investigation. *bioRxiv*, 2023.12. 11.570976 (2023).
18. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, s1-3 (2012).
19. Chen, C.Y. *et al.* Analysis across Taiwan Biobank, Biobank Japan, and UK Biobank identifies hundreds of novel loci for 36 quantitative traits. *Cell Genom* **3**, 100436 (2023).
20. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* **51**, D977-d985 (2023).
21. Kelleher, A.M., DeMayo, F.J. & Spencer, T.E. Uterine Glands: Developmental Biology and Functional Roles in Pregnancy. *Endocr Rev* **40**, 1424-1445 (2019).
22. Mills, M.C. *et al.* Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nat Hum Behav* **5**, 1717-1730 (2021).

23. Gibson, D.A. *et al.* Profiling the expression and function of oestrogen receptor isoform ER46 in human endometrial tissues and uterine natural killer cells. *Hum Reprod* **35**, 641-651 (2020).
24. Schlaikjær Hartwig, T. *et al.* Cell-free fetal DNA for genetic evaluation in Copenhagen Pregnancy Loss Study (COPL): a prospective cohort study. *Lancet* **401**, 762-771 (2023).
25. Chitty, L.S. Use of cell-free DNA to screen for Down's syndrome. *N Engl J Med* **372**, 1666-7 (2015).
26. Wang, W. *et al.* Homozygous variants in PANX1 cause human oocyte death and female infertility. *Eur J Hum Genet* **29**, 1396-1404 (2021).
27. Sang, Q. *et al.* A pannexin 1 channelopathy causes human oocyte death. *Sci Transl Med* **11**(2019).
28. Brinkmann, V. *et al.* Neutrophil extracellular traps kill bacteria. *Science* **303**, 1532-5 (2004).
29. Bouvier, S. *et al.* NETosis Markers in Pregnancy: Effects Differ According to Histone Subtypes. *Thromb Haemost* **121**, 877-890 (2021).
30. Guillotin, F. *et al.* Vital NETosis vs. suicidal NETosis during normal pregnancy and preeclampsia. *Front Cell Dev Biol* **10**, 1099038 (2022).
31. Yu, G. *et al.* APRIL and TALL-I and receptors BCMA and TACI: system for regulating humoral immunity. *Nat Immunol* **1**, 252-6 (2000).
32. Duan, T. *et al.* USP3 plays a critical role in the induction of innate immune tolerance. *EMBO Rep* **24**, e57828 (2023).
33. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
34. lab, N. UK Biobank-Neale lab. (2018).
35. Dalmia, A. *et al.* A genetic epidemiological study in British adults and older adults shows a high heritability of the combined indicator of vitamin B(12)

977 status (cB(12)) and connects B(12) status with utilization of mitochondrial
978 substrates and energy metabolism. *J Nutr Biochem* **70**, 156-163 (2019).
979 36. Pirinen, M. linemodels: clustering effects based on linear relationships.
980 *Bioinformatics* **39**(2023).
981 37. Li, A., Yang, S., Zhang, J. & Qiao, R. Establishment of reference intervals for
982 complete blood count parameters during normal pregnancy in Beijing. *J Clin*
983 *Lab Anal* **31**(2017).
984 38. Zhu, J., Li, Z., Deng, Y., Lan, L. & Yang, J. Comprehensive reference
985 intervals for white blood cell counts during pregnancy. *BMC Pregnancy*
986 *Childbirth* **24**, 35 (2024).
987 39. Ko, S. *et al.* GWAS of longitudinal trajectories at biobank scale. *Am J Hum*
988 *Genet* **109**, 433-445 (2022).
989 40. Verbanck, M., Chen, C.Y., Neale, B. & Do, R. Detection of widespread
990 horizontal pleiotropy in causal relationships inferred from Mendelian
991 randomization between complex traits and diseases. *Nat Genet* **50**, 693-698
992 (2018).
993 41. Bonaca, M.P., Hamburg, N.M. & Creager, M.A. Contemporary Medical
994 Management of Peripheral Artery Disease. *Circ Res* **128**, 1868-1884 (2021).
995 42. Ranaei Pirmardan, E. *et al.* Pre-hyperglycemia immune cell trafficking
996 underlies subclinical diabetic cataractogenesis. *J Biomed Sci* **30**, 6 (2023).
997 43. Jonasson, L., Tompa, A. & Wikby, A. Expansion of peripheral CD8+ T cells
998 in patients with coronary artery disease: relation to cytomegalovirus infection.
999 *J Intern Med* **254**, 472-8 (2003).
1000 44. Borgia, M.C. *et al.* Further evidence against the implication of active
1001 cytomegalovirus infection in vascular atherosclerotic diseases. *Atherosclerosis*
1002 **157**, 457-62 (2001).
1003 45. Parikh, N.I. *et al.* Adverse Pregnancy Outcomes and Cardiovascular Disease
1004 Risk: Unique Opportunities for Cardiovascular Disease Prevention in Women:

A Scientific Statement From the American Heart Association. *Circulation* **143**, e902-e916 (2021).

46. Auger, N., Tang, T., Healy-Profitós, J. & Paradis, G. Gestational diabetes and the long-term risk of cataract surgery: A longitudinal cohort study. *J Diabetes Complications* **31**, 1565-1570 (2017).

47. Patel, R.R., Steer, P., Doyle, P., Little, M.P. & Elliott, P. Does gestation vary by ethnic group? A London-based study of over 122,000 pregnancies with spontaneous onset of labour. *Int J Epidemiol* **33**, 107-13 (2004).

48. Cheung, S.W., Patel, A. & Leung, T.Y. Accurate description of DNA-based noninvasive prenatal screening. *N Engl J Med* **372**, 1675-7 (2015).

49. Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics* **53**, 120-126 (2021).

50. Cheng, S. *et al.* The STROMICS genome study: deep whole-genome sequencing and analysis of 10K Chinese patients with ischemic stroke reveal complex genetic and phenotypic interplay. *Cell Discov* **9**, 75 (2023).

51. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

52. Meisner, J., Liu, S., Huang, M. & Albrechtsen, A. Large-scale inference of population structure in presence of missingness using PCA. *Bioinformatics* **37**, 1868-1875 (2021).

53. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).

54. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).

55. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-45 (2009).

1032 56. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese
1033 population links cell types to complex human diseases. *Nat Genet* **50**, 390-400
1034 (2018).
1035 57. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: A Tool for
1036 Genome-wide Complex Trait Analysis. *The American Journal of Human*
1037 *Genetics* **88**, 76-82 (2011).
1038 58. Machiela, M.J. & Chanock, S.J. LDlink: a web-based application for exploring
1039 population-specific haplotype structure and linking correlated alleles of
1040 possible functional variants. *Bioinformatics* **31**, 3555-7 (2015).
1041 59. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of
1042 biomolecular interaction networks. *Genome Res* **13**, 2498-504 (2003).
1043 60. Chin, C.-H. *et al.* cytoHubba: identifying hub objects and sub-networks from
1044 complex interactome. *BMC Systems Biology* **8**, S11 (2014).
1045 61. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases
1046 and traits. *Nat Genet* **47**, 1236-41 (2015).
1047 62. Yang, J. *et al.* Genetic variance estimation with imputed variants finds
1048 negligible missing heritability for human height and body mass index. *Nat*
1049 *Genet* **47**, 1114-20 (2015).
1050 63. Wallace, C. Eliciting priors and relaxing the single causal variant assumption
1051 in colocalisation analyses. *PLoS Genet* **16**, e1008720 (2020).
1052 64. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the
1053 analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).
1054 65. Zhou, H. *et al.* OPENMENDEL: a cooperative programming project for
1055 statistical genetics. *Hum Genet* **139**, 61-71 (2020).
1056 66. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference
1057 across the human phenome. *eLife* **7**(2018).

67. Li, J., Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLOS Genetics* **13**(2017).

68. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-13 (2010).

69. Slob, E.A.W. & Burgess, S. A comparison of robust Mendelian randomization methods using summary data. *Genetic Epidemiology* **44**, 313-329 (2020).

70. John D. Storey, A.J.B., Alan Dabney, David Robinson. qvalue: Q-value estimation for false discovery rate control. R package version 2.36.0 (2024).

71. Sanderson, E. Multivariable Mendelian Randomization and Mediation. *Cold Spring Harb Perspect Med* **11**(2021).

72. Vabistsevits, M. Setting up Multivariable Mendelian Randomization analysis in R. (22 March, 2021).

Figure Legends

Figure 1. Schematic overview of the study.

We curated a comprehensive dataset comprising 122 pregnancy-related phenotypes from 120,579 Chinese pregnant women. We conducted an in-depth analysis of the genetic architecture underlying these phenotypes, identified biological pathways enriched for genetic variants, and explored genetic variants associated with alterations in the longitudinal trajectories of these phenotypes. Additionally, we investigated potential causal relationships between gestational phenotypes and various diseases.

Figure 2. Genetic associations with nine categories of gestational phenotypes.

The top-left inset presents a histogram depicting the number of loci associated with gestational phenotypes for each trait, categorized into novel and known loci. In the main panel, a Fuji plot illustrates the results of the genome-wide association analysis (GWAS) across all phenotypes within nine categories. Independent variants identified through meta-analysis ($P < 4.1 \times 10^{-10}$) are shown. The number of associated phenotypes per pleiotropic locus is represented by filled colored boxes, with functional and newly discovered loci labeled directly on the plot. Detailed information on functional loci can be found in Supplementary Table 15. Pleiotropic and trait-specific associations are distinguished by circles of varying sizes. The order of the 88 gestational phenotypes in both the histogram and Fuji plot corresponds to that in Supplementary Table 2.

Figure 3. Classification of the genetic effects of SNPs on gestational phenotypes and Taiwan Biobank phenotypes, along with the identification of pathways enriched for pregnancy specific and general variants.

Panel **A** presents a plot classifying the genetic effects of SNPs ($P < 4.1 \times 10^{-10}$, Bonferroni correction threshold) on gestational phenotypes, utilizing a Bayesian classification algorithm and colocalization analysis. Panel **B** features a butterfly bar plot illustrating the top 20 most significant pathways for pregnancy-specific variants, identified through Metascape enrichment analysis. On the left side, the plot shows the enrichment *P* values for pregnancy-specific variants and the percentage of genes enriched in each pathway relative to the total number of pregnancy-specific genes. The right side displays the corresponding results for general variants. Pathways showing significant differences in gene enrichment, as determined by a chi-square test, are marked with an asterisk (*).

Figure 4. Results of trajectory changes in blood cell traits during pregnancy.

Panel **A** presents forest plots illustrating the genetic effects across five time periods with significant genetic and environmental interactions for the first 25 variants, as well as for at least two of the five periods where no overlap in genetic effects was observed. Different colors represent each period, indicating beta values and 95% confidence intervals, with annotations highlighting whether the variant has a P-value less than 5×10^{-8}. The eQTL status of each variant is indicated adjacent to the variant label. Detailed information can be found in Supplementary Table 31. Panel **B** depicts changes in white blood cell (WBC) counts across the first, second, and third trimesters, at delivery, and during the postpartum period, based on data from two hospitals. A generalized additive model was used to smooth the curve, with the ribbon around the curve representing the 95% confidence interval. The first trimester spans from conception to 14 weeks (98 days), the second trimester covers weeks 14 through 28 (196 days), and the third trimester extends from week 28 to delivery. Panel **C** displays a forest plot of the genetic effect of the rs2270401-A allele at 17q21.1 (*MED24*) on WBC counts, based on a meta-analysis across the first, second, and third

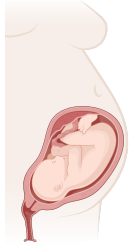
trimesters, at delivery, and postpartum. Points represent the genetic effect (beta), with error bars indicating the 95% confidence interval. Panel **D** presents the results of enrichment analyses for variants showing significant and non-significant genetic and environmental interactions.

BASO: basophil absolute value; BASO_P: basophil percentage; HGB: hemoglobin level; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; RDW_CV: red blood cell volume distribution width variation; PLT: platelet count; MPV: mean platelet volume; PCT: platelet crit; PDW: platelet distribution width; P_LCR: platelet-large cell ratio.

Figure 5. Causal inferences between pregnancy-related phenotypes and BBJ phenotypes.

Panel **A** displays bidirectional Mendelian Randomization (MR) estimates, illustrating causal links between all pregnancy-related phenotypes (blue nodes) and traits from the Biobank Japan (BBJ) cohort (red nodes). Arrows indicate the direction of association, with red denoting an increasing effect and blue representing a decreasing effect. Panel **B** features a heatmap of MR results, showing nine categories of gestational phenotypes as the exposures (right) and BBJ phenotypes as the outcomes (bottom). Relationships with *P* values below the false discovery rate (FDR) threshold are annotated with an asterisk (*, $q < 0.01$). Panel **C** presents an MR scatterplot where maternal triglyceride (TG) levels serve as the exposure and the use of antithrombotic agents in the BBJ cohort as the outcome. Panel **D** shows an MR scatterplot with maternal oral glucose tolerance test 1-hour plasma glucose (OGTT1H) levels as the exposure, and peripheral artery disease in the BBJ cohort as the outcome.

Study Cohort



Genotype

Genome sequencing from NIPT
N=121,579

Phenotypes

Blood lipid & Glycemic (N=8)
Blood routine (N=32)
Infection (N=14)
Liver, Kidney & Thyroid function (N=15)
Tang & NIPT screen (N=19)

Basic information (N=7)
Birth outcome (N=6)
Gestational disorders (N=21)

Genetic association

Baoan (N=50,963)
Longgang (N=70,616)
meta
5,027 independent signals in 3,845 loci

Consistency
Baoan vs Longgang
(N=121,579)

External replication
NIPT PLUS
(N=5,733)

Genome-wide association study

24 gestational phenotypes with multiple measurements in 5 periods

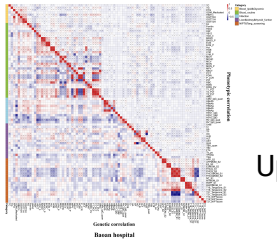
TrajGWAS

localized association

GWAS of longitudinal trajectories

Post-GWAS

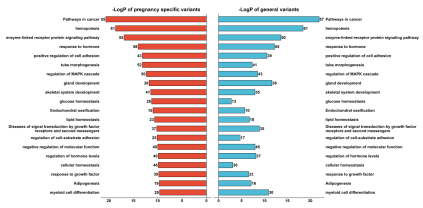
Heritability & Correlation



LDSC

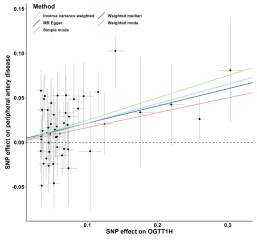
Up to 44% heritability

Different variant analysis



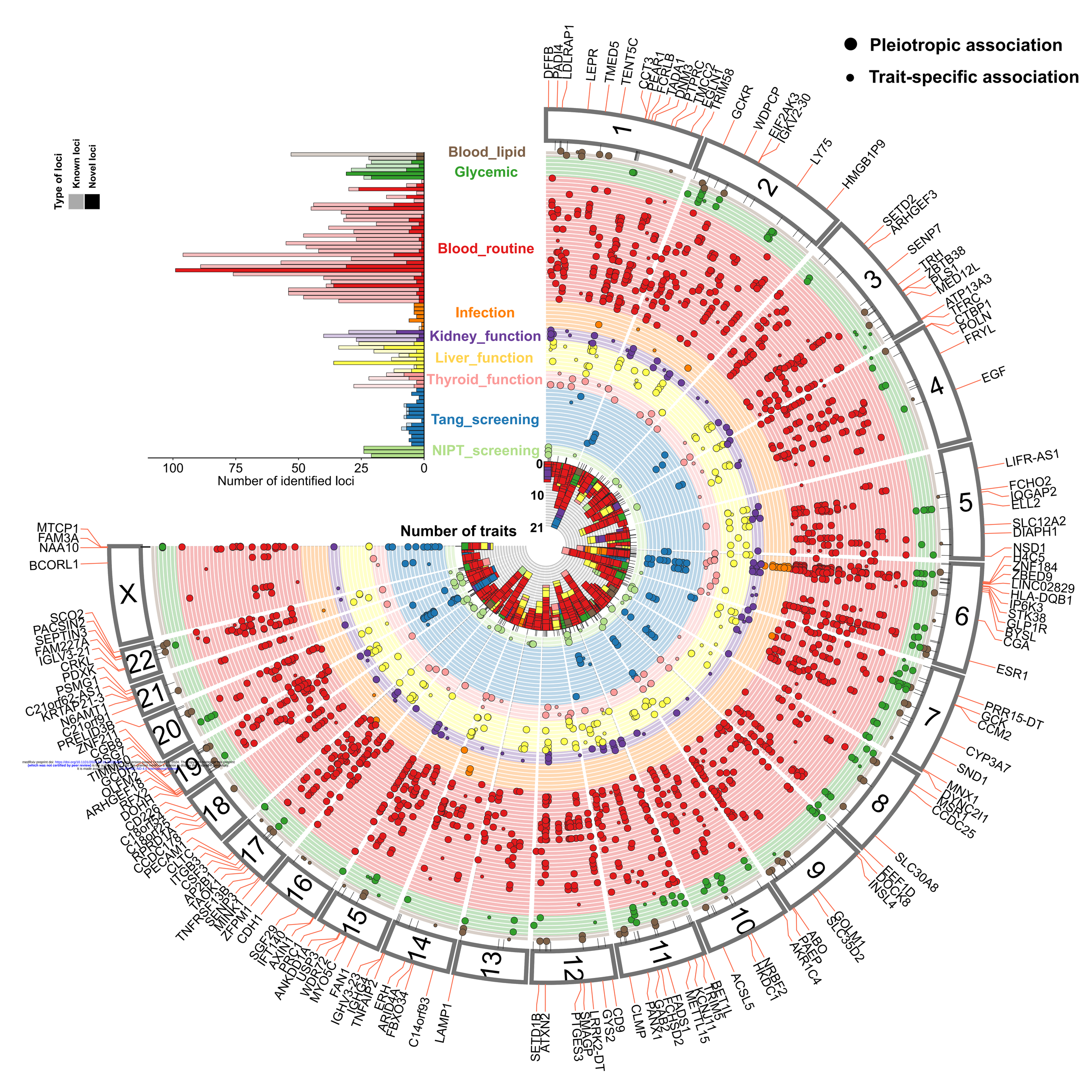
Genetic difference

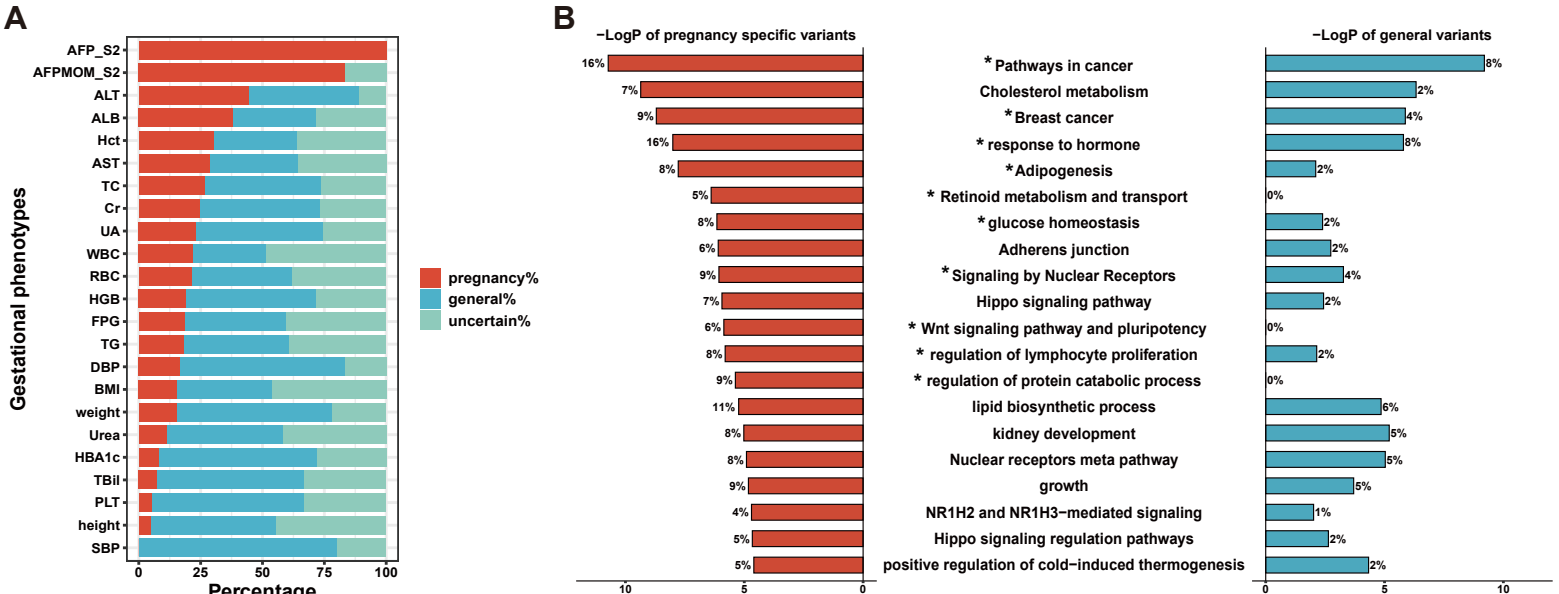
Mendelian randomization

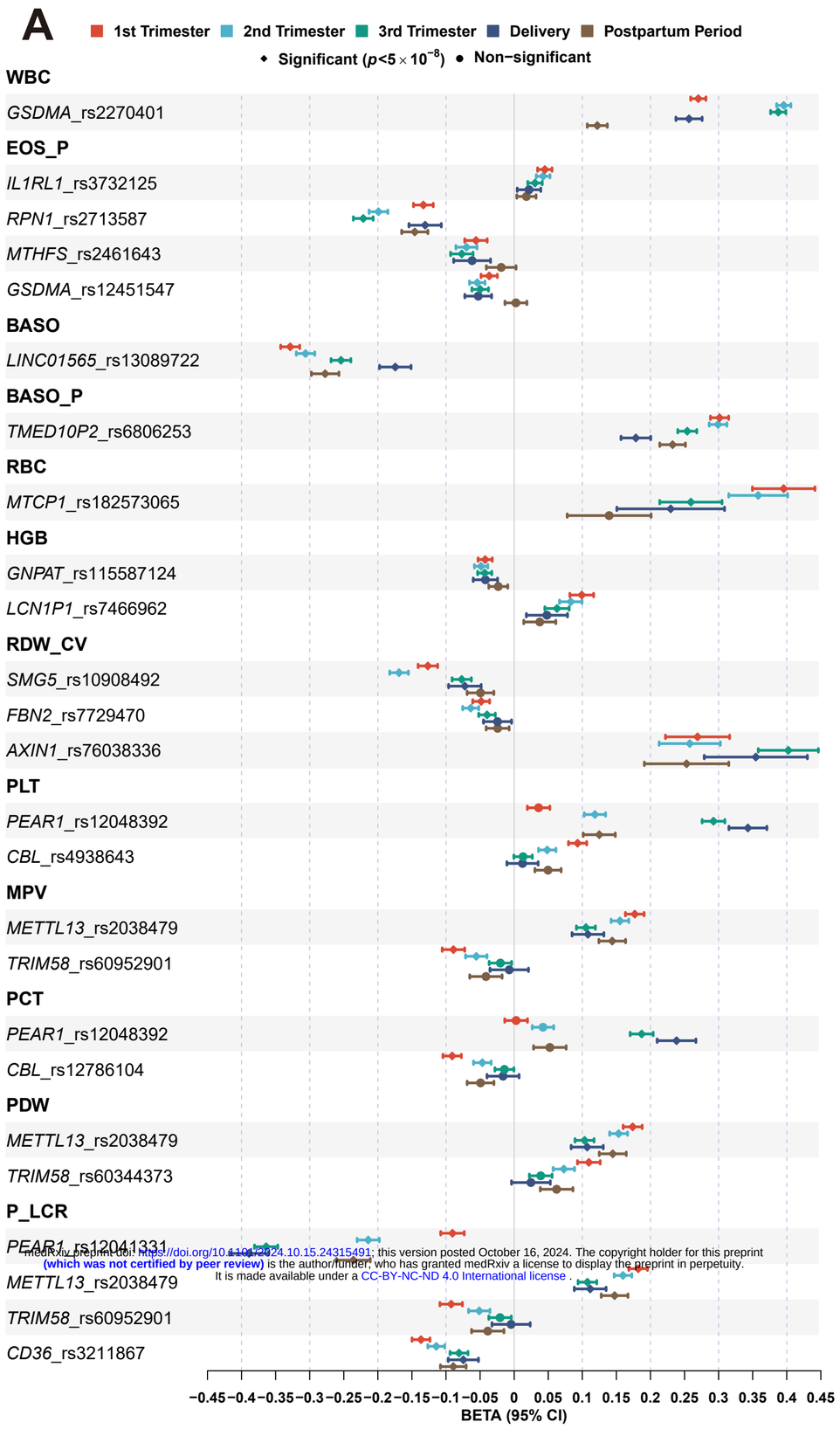


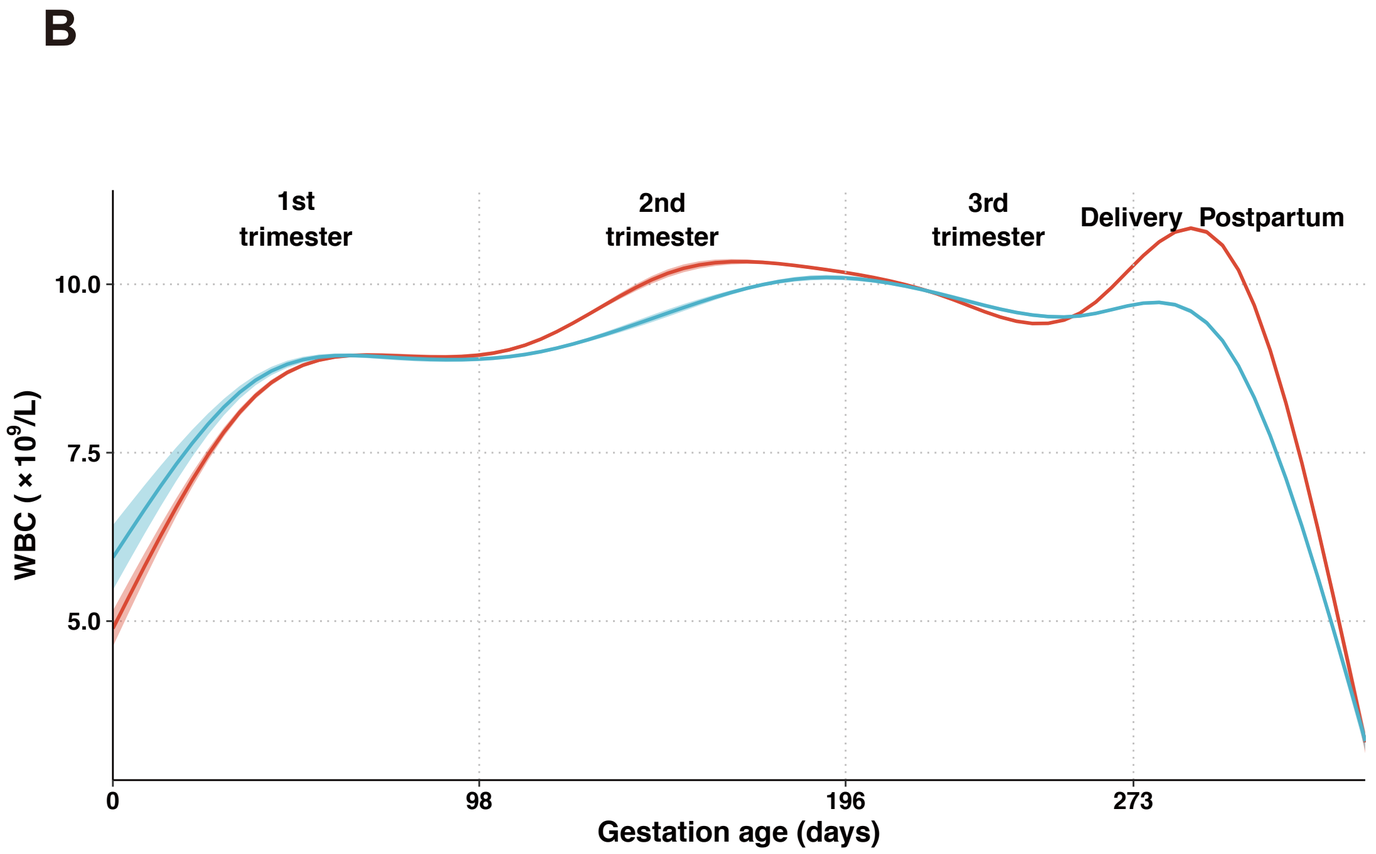
TwoSample MR
&
Multivariable MR

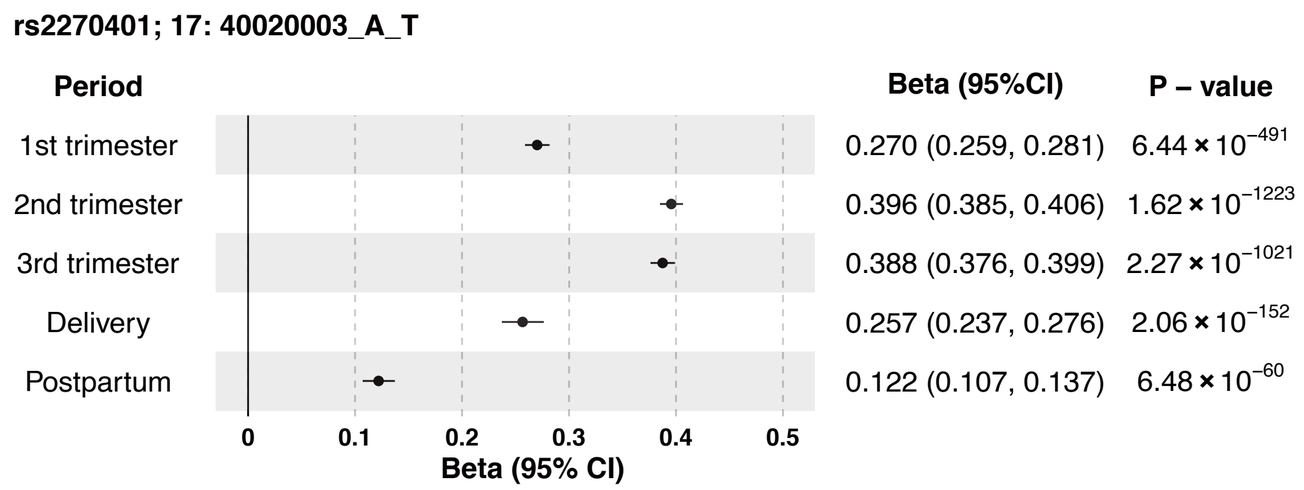
Future health

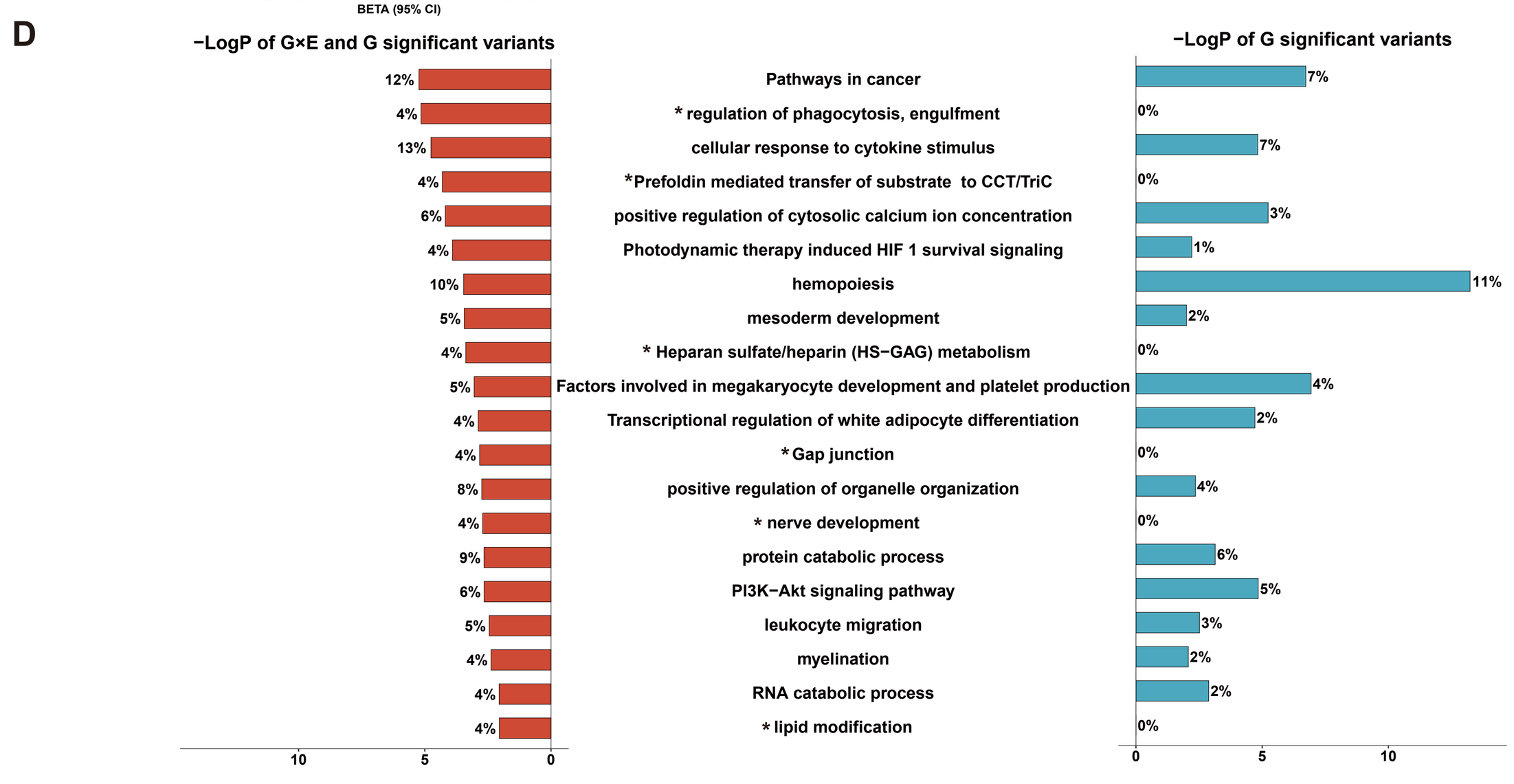


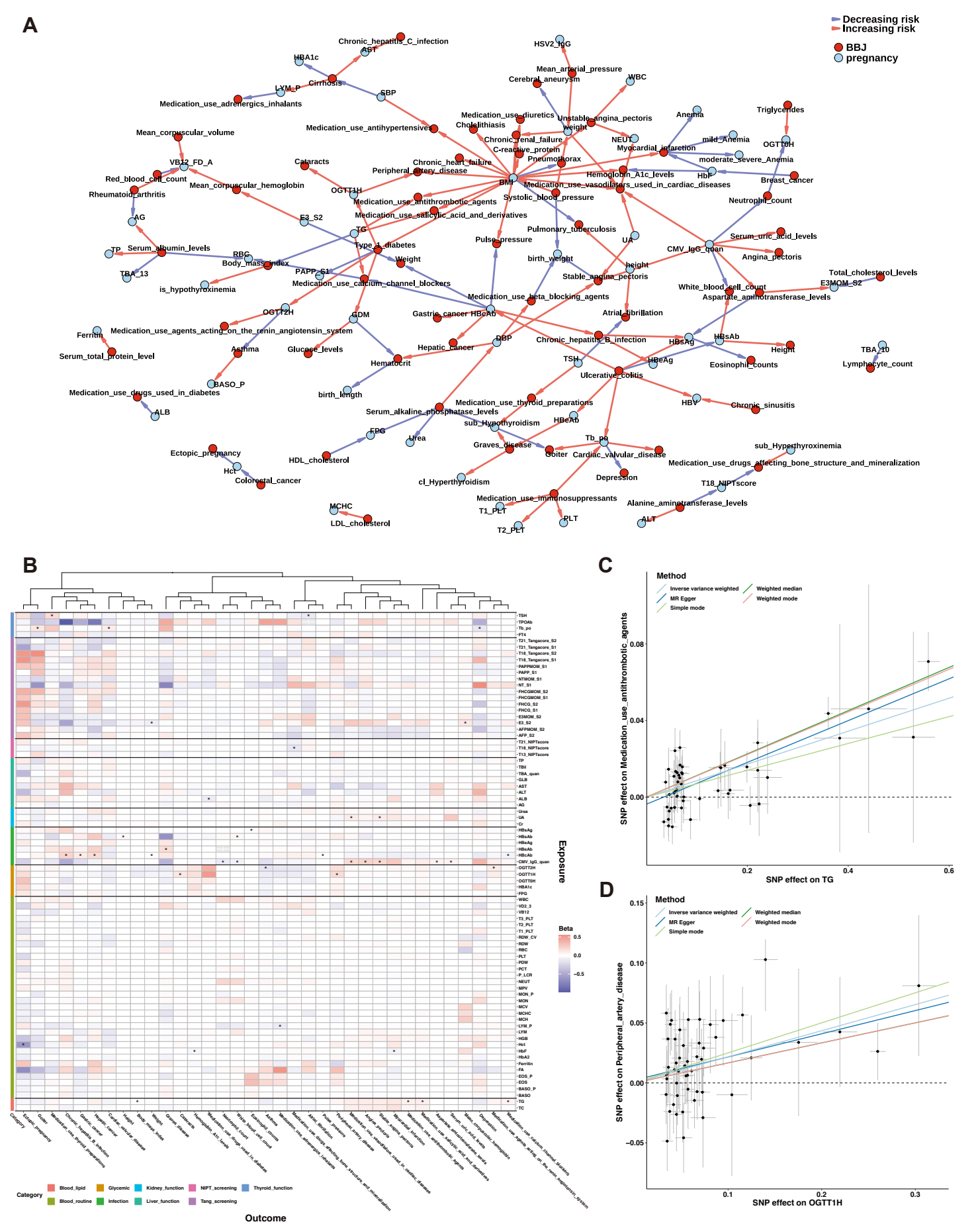

A
AFP_S2
AFPMOM_S2
ALT
ALB
Hct
AST
TC
Cr
UA
WBC
RBC
HGB
FPG
TG
DBP
BMI
weight
Urea
HBA1c
TBil
PLT
height
SBP
Gestational phenotypes
0
25
50
75
100
Percentage
pregnancy%
general%
uncertain%
B
−LogP of pregnancy specific variants
16%
7%
9%
16%
8%
5%
8%
6%
9%
7%
6%
8%
9%
11%
8%
8%
9%
4%
5%
5%
10
5
0
* Pathways in cancer
Cholesterol metabolism
* Breast cancer
* response to hormone
* Adipogenesis
* Retinoid metabolism and transport
* glucose homeostasis
Adherens junction
* Signaling by Nuclear Receptors
Hippo signaling pathway
* Wnt signaling pathway and pluripotency
* regulation of lymphocyte proliferation
* regulation of protein catabolic process
lipid biosynthetic process
kidney development
Nuclear receptors meta pathway
growth
NR1H2 and NR1H3−mediated signaling
Hippo signaling regulation pathways
positive regulation of cold−induced thermogenesis
−LogP of general variants
8%
2%
4%
8%
2%
0%
2%
2%
4%
2%
0%
2%
0%
6%
5%
5%
5%
1%
2%
2%
0
5
10










A
Decreasing risk
Increasing risk
BBJ
pregnancy
Chronic_hepatitis_C_infection
AST
HBA1c
Cirrhosis
LYM_P
SBP
Medication_use_adrenergics_inhalants
HSV2_IgG
Mean_arterial_pressure
Cerebral_aneurysm
WBC
Medication_use_diuretics
Cholelithiasis
Unstable_angina_pectoris
weight
Anemia
Triglycerides
Medication_use_antihypertensives
Chronic_renal_failure
C-reactive_protein
Pneumothorax
NEUT
Myocardial_infarction
mild_Anemia
OGTT0H
moderate_severe_Anemia
Mean_corpuscular_volume
VB12_FD_A
Red_blood_cell_count
Rheumatoid_arthritis
Mean_corpuscular_hemoglobin
Cataracts
Chronic_heart_failure
Peripheral_artery_disease
BMI
HbF
Hemoglobin_A1c_levels
Breast_cancer
Medication_use_vasodilators_used_in_cardiac_diseases
OGTT1H
Medication_use_antithrombotic_agents
Systolic_blood_pressure
Neutrophil_count
AG
E3_S2
Medication_use_salicylic_acid_and_derivatives
TG
Pulmonary_tuberculosis
Pulse_pressure
UA
Serum_uric_acid_levels
TP
Serum_albumin_levels
RBC
Body_mass_index
Type_1_diabetes
Weight
birth_weight
CMV_IgG_quan
Angina_pectoris
TBA_13
PAPP_S1
Medication_use_calcium_channel_blockers
Stable_angina_pectoris
height
Total_cholesterol_levels
E3MOM_S2
is_hypothyroxinemia
White_blood_cell_count
Aspartate_aminotransferase_levels
Medication_use_beta_blocking_agents
OGTT2H
GDM
Gastric_cancer
HBcAb
Atrial_fibrillation
Ferritin
Medication_use_agents_acting_on_the_renin_angiotensin_system
Asthma
Glucose_levels
Hepatic_cancer
DBP
Chronic_hepatitis_B_infection
HBsAg
HBsAb
Height
Serum_total_protein_level
Hematocrit
TSH
HBeAg
Eosinophil_counts
TBA_20
Lymphocyte_count
Ulcerative_colitis
BASO_P
Medication_use_drugs_used_in_diabetes
birth_length
HBV
Chronic_sinusitis
ALB
Serum_alkaline_phosphatase_levels
Medication_use_thyroid_preparations
sub_Hypothyroidism
HBeAb
FPG
Urea
Tb_p6
Graves_disease
Goiter
Cardiac_valvular_disease
sub_Hyperthyroxinemia
Ectopic_pregnancy
HDL_cholesterol
Medication_use_drugs_affecting_bone_structure_and_mineralization
Hct
Colorectal_cancer
cl_Hyperthyroidism
Depression
T18_NIPTscore
Medication_use_immunosuppressants
Alanine_aminotransferase_levels
MCHC
T1_PLT
PLT
ALT
LDL_cholesterol
T2_PLT
B
Exposure
Beta
0.5
0.0
−0.5
Category
Blood_lipid
Glycemic
Kidney_function
NIPT_screening
Thyroid_function
Blood_routine
Infection
Liver_function
Tang_screening
Outcome
C
Method
Inverse variance weighted
Weighted median
MR Egger
Weighted mode
Simple mode
SNP effect on Medication_use_antithrombotic_agents
SNP effect on TG
0.08
0.04
0.00
0.2
0.4
0.6
D
Method
Inverse variance weighted
Weighted median
MR Egger
Weighted mode
Simple mode
SNP effect on Peripheral_artery_disease
SNP effect on OGTT1H
0.15
0.10
0.05
0.00
−0.05
0.1
0.2
0.3