

# Genetic Polymorphisms Associated with Adverse Pregnancy Outcomes in Nulliparas

Rafael F. Guerrero <sup>\*1,2</sup>, Raiyan R. Khan <sup>\*3</sup>, Ronald J. Wapner <sup>4</sup>, Matthew W. Hahn <sup>1,5</sup>, Anita Raja <sup>6</sup>, Ansaf Salleb-Aouissi <sup>3</sup>, William A. Grobman <sup>7</sup>, Hyagriv Simhan <sup>8</sup>, Robert Silver <sup>9</sup>, Judith H. Chung <sup>10</sup>, Uma M. Reddy <sup>11</sup>, Predrag Radivojac <sup>11</sup>, Itsik Pe'er <sup>3</sup>, David M. Haas <sup>13</sup>

\* These authors contributed equally to this work.

<sup>1</sup>Department of Computer Science, Indiana University, Bloomington, IN

<sup>2</sup>Department of Biological Sciences, North Carolina State University, Raleigh, NC

<sup>3</sup>Department of Computer Science, Columbia University, New York, NY

<sup>4</sup>Columbia University Irving Medical Center, New York, NY

<sup>5</sup>Department of Biology, Indiana University, Bloomington, IN

<sup>6</sup>Department of Computer Science, CUNY Hunter College, NY

<sup>7</sup>Department of Obstetrics and Gynecology, Northwestern University Feinberg School of Medicine, Chicago, IL

<sup>8</sup>University of Pittsburgh Medical Center, Pittsburgh, PA

<sup>9</sup>Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT

<sup>10</sup>Department of Obstetrics and Gynecology, University of California, Irvine, Orange, CA

<sup>11</sup>Department of Obstetrics and Gynecology, Columbia University, New York, NY

<sup>12</sup>Khoury College of Computer Sciences, Northeastern University, Boston, MA

<sup>13</sup>Indiana University School of Medicine, Indianapolis, IN

**Corresponding Author:** David M. Haas, dahaas@iu.edu, Indiana University School of Medicine Indianapolis, IN 46202

## ABSTRACT

**Background:** Adverse pregnancy outcomes (APOs) affect a large proportion of pregnancies and represent an important cause of morbidity and mortality worldwide. Yet, the pathophysiology of APOs is poorly understood, limiting our ability to prevent and treat these conditions.

**Objective:** To search for genetic risk markers for four APOs, we performed genome-wide association studies (GWAS) for preterm birth, preeclampsia, gestational diabetes, and pregnancy loss.

**Study Design:** A total of 9,757 nulliparas from the nuMoM2b study were genotyped. We clustered participants by their genetic ancestry and focused our analyses on the three sub-cohorts with the largest sample sizes: European (EUR, n=6,082), African (AFR, n=1,425), and American (AMR, n=846). Association tests were carried out separately for each sub-cohort and brought together via meta-analysis. Four APOs were tested by GWAS: preeclampsia (n=7,909), gestational length (n=4,781), gestational diabetes (n=7,617), and pregnancy loss (n=7,809). Using the results of the genome-wide associations for each APO, SNP-based heritability of these traits was inferred using LDscore. Putative regulatory effects were inferred by transcriptome-wide association analysis.

**Results:** Two variants were significantly associated with pregnancy loss (rs62021480: OR = 3.29,  $P = 7.83 \times 10^{-11}$ , and rs142795512: OR = 4.72,  $P = 9.64 \times 10^{-9}$ ), implicating genes *TRMU* and *RGMA* in this APO. An intronic variant was significantly associated with gestational length (rs73842644: beta = -0.667,  $P = 4.9 \times 10^{-8}$ ). Three loci were significantly associated with gestational diabetes (rs72956265: OR = 3.09,  $P = 2.98 \times 10^{-8}$ , rs10890563: OR = 1.88,  $P = 3.53 \times 10^{-8}$ , rs117689036: OR = 3.15,  $P = 1.46 \times 10^{-8}$ ), located on or near *ZBTB20*, *GUCY1A2*, and *MDGA2*, respectively. Several loci previously correlated with preterm birth (in genes *WNT4*, *EBF1*, *PER3*, *IL10*, and *ADCY5*), gestational diabetes (in *TCF7L2*), and preeclampsia (in *MTHFR*) were found to be associated with these outcomes in our cohort as well.

**Conclusion:** Our study identified genetic associations with gestational diabetes, pregnancy loss, and gestational length. We also confirm correlations of several previously identified variants with these APOs.

**Keywords:** Genetic association, preeclampsia, preterm birth, gestational diabetes, fetal death, stillbirth, pregnancy loss, miscarriage

**Disclosure Statement:** The authors declare no conflict of interest

**Source of financial support:** Precision Health Initiative of Indiana University, National Institutes of Health award R01HD101246 to DMH and PR. Cooperative agreement funding from the National Heart, Lung, and Blood Institute and the Eunice Kennedy Shriver National Institute of Child Health and Human Development: grant U10-HL119991 to RTI International; grant U10-HL119989 to Case Western Reserve University; grants U10-HL120034 and R01LM013327 to Columbia University; grant U10-HL119990 to Indiana University; grant U10-HL120006 to the University of Pittsburgh; grant U10- HL119992 to Northwestern University; grant U10- HL120019 to the University of California, Irvine; grant U10- HL119993 to University of Pennsylvania; and grant U10-HL120018 to the University of Utah. National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health to Clinical and Translational Science Institutes at Indiana University (grant UL1TR001108) and University of California, Irvine (grant UL1TR000153).

## **Introduction**

Adverse pregnancy outcomes (APOs) are a serious threat to the health of pregnant persons and children. APOs affect a significant fraction of pregnancies across the globe and are among the leading causes of morbidity and mortality worldwide<sup>1</sup>. Among the most common APOs are preterm birth (which occurs in over 10% of pregnancies in the United States (US)<sup>2</sup>), preeclampsia (which develops in 5-10% of pregnancies<sup>3</sup>), gestational diabetes (occurring in roughly 6% of pregnancies in the US<sup>4</sup>), and pregnancy loss (estimated to occur in about 20% of pregnancies in the US<sup>5</sup>). APOs are also highly correlated with future disease in birthing parents. For example, gestational diabetes carries a lifetime 50% risk of type 2 diabetes (T2D) in the mother<sup>6</sup>, while preeclampsia might lead to a 2-3 fold increase of cardiovascular disease later in life<sup>7</sup>. Yet, the factors driving these diseases remain poorly understood, hindering efforts in prevention and treatment.

To better understand the mechanisms and improve the prediction of APOs in nulliparous individuals, the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b) consortium recruited and prospectively followed a large cohort of nulliparous people beginning in their first trimester of pregnancy. Participants underwent several assessments over the course of their pregnancies, resulting in a comprehensive profile that included biospecimens, clinical measurements, ultrasounds, behavior (through interviews and questionnaires), physical activity assessment, and dietary content.

By precisely characterizing different aspects of over 10,000 pregnancies, the nuMoM2b cohort has already yielded valuable insights into the factors that contribute to APOs<sup>8,9,10</sup>. Additionally, the availability of biospecimens provides a unique opportunity to study the genetic underpinnings of APOs. The objective of this study was to genotype the nuMoM2b cohort and to search genome-wide for variants associated with four APOs: gestational length (as a proxy for preterm birth), preeclampsia, gestational diabetes mellitus (GDM), and pregnancy loss.

## **Methods**

### *Participants*

The participants of the analysis were enrolled in the nuMoM2b cohort (<https://www.nichd.nih.gov/research/supported/nuMoM2b>), a longitudinal, multiethnic cohort study of nulliparous individuals. All participating centers obtained approval by the local Institutional Review Boards (IRBs) of their corresponding recruitment institutions, with further

details on data collection having been previously described<sup>11,12</sup>. The study enrolled 10,038 nulliparous people from the first trimester of their pregnancy to participate in three study visits during pregnancy. Collection of health status and biomarkers were conducted at regular intervals, and documentation of pregnancy outcomes was performed by chart abstraction using *a priori* definitions. The details of this process were described elsewhere<sup>7</sup>.

### *Phenotype definitions*

**Gestational length:** We opted to use a quantitative phenotype, gestational length, instead of a binary preterm/full term outcome to gain additional information and statistical power by using a more granular phenotype. Gestational length was determined from an estimated due date established by a first-trimester ultrasound crown-rump length measurement and was recorded in weeks<sup>12</sup>. Preterm birth was defined as any live birth that occurred before 37 weeks gestational age. Cases of stillbirth, fetal demise, elective termination, and indicated termination were all excluded from this phenotype group (Table 1).

**Preeclampsia:** Cases include diagnosis of preeclampsia (with and without severe features), eclampsia, and chronic hypertension with superimposed preeclampsia. A detailed description of nuMoM2b study definitions of hypertensive disorders of pregnancy was published in the supplement to the paper by Facco et al<sup>13</sup>. All other individuals were treated as controls.

**Gestational diabetes (GDM):** GDM was diagnosed through clinical evaluation, from fasting blood sugar, sequential 1-hour glucose challenge test followed by a 3-hour glucose tolerance test (GTT), or a single step 2-hour 75-gram GTT<sup>13</sup>. We excluded individuals diagnosed with pregestational diabetes from GDM analyses, and all other individuals were treated as controls.

**Pregnancy loss:** All subjects who had a pregnancy loss, regardless of gestational age, were considered as cases (Table 1). Individuals who underwent termination of pregnancy were excluded from the pregnancy loss analysis. Subjects who had a live birth were treated as controls.

### *Genotyping*

The genotyped cohort comprised 9,757 participants who had adequate samples and agreed to be genotyped. DNA extractions from whole blood were carried out on a QIASymphony instrument (from Qiagen; extraction kit DSP DNA Midi Kit #937355, protocol Blood\_1000\_V7\_DSP) at the Center for Genomics and Bioinformatics (Indiana University, Bloomington). Genotyping was completed at the Van Andel Institute (Grand Rapids, MI, USA)

using the Infinium Multi-Ethnic Global D2 BeadChip (Illumina, Miami, USA). We imposed standard filters for quality control of loci at this stage (cluster separation  $< 0.3$ , AA R Mean  $< 0.2$ , AB R Mean  $< 0.2$ , BB R Mean  $< 0.2$ , 10% GC  $< 0.3$ ) using GenomeStudio v2.4 (Illumina). Genotype calls (in GCT format) for the 1,748,280 loci that passed initial quality control were made with Beeline autoconvert (Illumina).

*Inclusion criteria, QC pipeline, Ancestry determination, and imputation*

We performed preprocessing to determine any sex inconsistencies, autosome missingness ( $>5\%$ ), and contamination (more than 25 related subjects within the same dataset). Given the highly heterogeneous data set, we used the KING-Robust algorithm, a pairwise kinship estimator for GWAS that is robust to the presence of unknown population substructure<sup>14</sup>. We inferred kinship estimates between all pairs of subjects, randomly removing one subject from the pairs of subjects with first- or second-degree relatedness such that we minimized the number of subjects removed. The remaining samples were filtered to remove SNPs with a minor allele frequency (MAF) less than 0.01, genotyping rate  $< 95\%$ , and a Hardy-Weinberg Equilibrium (HWE)  $P < 5 \times 10^{-6}$ .

We determined the ancestry of each subject using SNPweights v.2.1 and leveraging approximately 40,000 ancestry informative markers available from the 1000 Genomes Consortium<sup>15</sup>. Applying a probability cut-off of  $>50\%$ , the samples were clustered into five ancestry groups concordant with the 1000 Genomes Consortium<sup>15</sup> super-populations: African (AFR,  $n=1425$ ), American (AMR;  $n=846$ ), East Asian (EAS;  $n=323$ ), European (EUR,  $n=6082$ ), and South Asian (SAS;  $n=112$ ). We observed a large fraction of highly admixed individuals, thus we also established a sixth group (ADM,  $n=891$ ) of subjects who have none of the five ancestry group percentages reaching  $>50\%$  (Supplemental Figure S1). The heterogeneous genetic ancestry of this cohort would result in high levels of genomic inflation in a single GWAS analysis. To reduce this confounding effect, we clustered the cohort by genetic ancestry for downstream analyses (Supplemental Figure S1, Supplemental Figure S2). Thus, from this point subjects were analyzed in separate groups according to their majority genetic ancestry when doing the filtering preprocessing and imputation.

Not all the ancestry groups contained enough subjects to power a GWAS study. Thus, we only proceeded with imputing the European, African, and American ancestry subjects. These subjects were phased with EAGLE and imputed by Minimac3<sup>16</sup> using the TOPMED Imputation

Server<sup>17,18</sup>. We used version R2 of the TOPMED panel, currently the largest panel of sequenced human genomes, and containing all the ancestries we observed in our cohort. Following imputation, we excluded SNPs with an  $r^2$  quality score  $<0.7$ , MAF  $<0.01$ , genotyping rate  $<95\%$ , and a Hardy-Weinberg Equilibrium (HWE)  $P < 5 \times 10^{-6}$  within each imputed group.

Population structure was determined by PLINK using pruned SNPs from the data (Linkage Disequilibrium pruning  $r^2 < 0.1$ ). The top ten principal components were computed for each of the three sub-cohorts (EUR, AFR, AMR). As maternal age has a non-linear effect on preterm birth, we transformed each subject's age into the distance from the median percentile in all subjects.

### *Genome-wide associations*

Association testing was carried out using regression models implemented in PLINK v1.9<sup>19</sup>. The suggestive association threshold was  $P < 1 \times 10^{-5}$  and the threshold for genome-wide significance was  $P < 5 \times 10^{-8}$ . The model was adjusted for each subject's rank-transformed age and the first ten principal components from our population structure analysis. We used all ten computed principal components due to the high level of population substructure and genetic admixture present in the nuMoM2b data set. Any subjects with missing age or phenotype information were excluded from the analysis. Results from each sub-cohort were fixed-effect meta-analyzed using GWAMA<sup>20</sup> and SNPs that did not appear in at least two of the three individual GWAS were removed. Result plots were displayed using R libraries. The SNPs of interest were further annotated with Variant Effect Predictor<sup>21</sup> to identify their corresponding gene function. The model for gestational length was adjusted for rank-transformed age.

Using the results of the genome-wide associations for each APO, we inferred SNP-based heritability of these traits using LDscore<sup>22</sup>. We searched for phenotypes that had significant genetic correlations with the APOs using the LD-Hub database<sup>23</sup>. Finally, we scanned for putative regulatory effects by carrying out a transcriptome-wide association analysis (TWAS) as implemented in the Fusion software, and using the available expression reference weights for whole blood and adipose tissue<sup>24</sup>, as well as liver, pancreas, vagina, and uterus, and whole blood from the GTEx v7 multi-tissue RNA-seq data set. For TWAS,  $P$ -values were Bonferroni-corrected by the number of genes in each panel. Conditional expression analyses were performed using R scripts from the Fusion software.

## Results

### *Pregnancy Loss*

This trait showed the highest SNP-based heritability (0.25) of the four APOs studied. We observed two novel associations with pregnancy loss (Table 2): rs62021480 and rs142795512. Located on chromosome 15, rs62021480 encodes a synonymous variant on *RMGA* (Repulsive Guidance Molecule BMP Co-Receptor A), a gene that encodes a glycoprotein that guides developing axons and may act as a tumor suppressor. Rs142795512 is an intronic variant in the gene *TRMU* (TRNA 5-Methylaminomethyl-2-Thiouridylate Methyltransferase), which encodes a mitochondrial tRNA-modifying enzyme. Variants in *TRMU* have been linked to several disease phenotypes, including infantile liver failure<sup>25</sup>.

Through TWAS, we found a correlation between gene expression levels of *TTC38* (Tetratricopeptide Repeat Domain 38) in uterine tissue and pregnancy loss ( $Z=4.83$ ,  $P=1.36 \times 10^{-6}$ ; Supplemental Figure S10).

### *Gestational Length*

The genomic inflations of each sub-cohort GWAS ranged between  $\lambda = 0.99$  and  $\lambda = 1.02$ , while the GWA meta-analysis across sub-cohorts did not show genomic inflation (Supplemental Figure S3). SNP-based heritability for this trait was 23%. A single SNP, rs144380855, reached genome-wide significance in the meta-analysis of gestational length. rs144380855 appears to be associated with a reduced gestational length ( $\beta = -2.07$ ,  $P = 5.3 \times 10^{-10}$ ), and is an intergenic SNP (Table 2). We found evidence of association ( $P < 0.05$ ) in 8 out of 34 SNPs previously reported to be associated with gestational length (Table 3; Supplemental Table S1). These 8 markers were clustered in five genes: *WNT4* (Wnt Family Member 4), *EBF1* (EBF transcription factor 1), *PER3* (Period Circadian Regulator 3), *IL10* (Interleukin 10), and *ADCY5* (Adenylate Cyclase 5).

### *Gestational Diabetes mellitus*

Three significant loci were identified in the GWAS for GDM (Table 2). The top locus included intronic variants rs61167087 and rs72956265, in the gene *ZBTB20*, a Zinc Finger and BTB Domain Containing 20. The second significant association is also an intronic SNP, rs117689036 in *MDGA2*, a MAM Domain-Containing Glycosylphosphatidylinositol Anchor 2. Finally, rs10890563 is a 3'UTR variant on the gene *GUCY1A2*, Guanylate Cyclase 1 Soluble Subunit Alpha 2. Several loci show suggestive associations in the meta-analysis, which we report

in the supplemental text (Figure 4; Supplemental Table S2). We did not find genomic inflation in any of the ancestry groups or in the meta-analysis ( $\lambda = 1.004$ , Supplemental Figure S5), and we estimated SNP-based heritability to be 0.17.

We also found a positive association consistent with an effect on gestational diabetes for one previously reported variant (rs4506565,  $P = 0.001$ ; Table 3). However, we failed to find any confirmatory signal in 17 other variants (Supplemental Table S1).

### *Preeclampsia*

We found no SNPs associated with preeclampsia (Supplemental Figure S6). Again, genomic inflation was low ( $\lambda = 1.005$ , Supplemental Figure S7), and our SNP-based estimate of heritability (0.02) suggests a weak genetic basis for this APO. We confirmed 1 of 22 previously linked variants: rs17367504 (near gene *MTHFR*) showed a consistent effect across our three studied sub-populations (Table 3; Supplemental Table S1).

## **Comment**

### *Principal Findings*

We identified a novel locus associated with gestational length, along with three loci associated with GDM and two additional loci associated with the previously understudied phenotype of pregnancy loss. Our study suggests that variants in *ZBTB20*, a gene that has been previously found to regulate beta cell function and glucose homeostasis in mice<sup>26</sup>, may be involved in the development of GDM. Furthermore, TWAS analysis implicates the gene *TTC38* in pregnancy loss across four tissue types (liver, ovary, pancreas, and uterus). Consistent with this, one of the top SNPs in the pregnancy loss GWAS (rs114058457, Supplementary Table S2), was previously established by Jansen et al. as an eQTL for *TTC38* in blood in a cohort of 4,896 subjects of European ancestry (FDR-corrected  $P < 1.34 \times 10^{-5}$ ).<sup>27</sup>

### *Clinical Implications*

The SNPs that reached genome-wide significance in our study highlight the role of previously unknown genes in our APOs of interest. These genes may further implicate specific biological interactions and pathways that are disrupted in each APO. While not reaching genome-wide significance, we were able to document the directionality of association for several SNPs

previously reported for other APOs. These SNPs, now having been discerned by different investigations, are valuable by identifying genes relevant to pathophysiology—which can lead to better prevention and treatment.

Clinical utility of genetic screening is most effective when there is a preventive measure that can be employed. If strong genetic markers for preeclampsia, for instance, had been discovered in pathway genes for mechanisms that could be influenced by aspirin therapy, these would be prime candidates for screening nulliparas who might not have other risk factors that would lead to recommended prophylaxis<sup>28</sup>. SNPs implicated in shorter gestational length could lead to further screening with cervical length ultrasound or perhaps the administration of preventive progestin therapy. Incorporating genetic screening for APOs into clinical care will necessitate more data and trials, but novel and confirmatory findings such as those presented here are initial steps towards potential clinical utility.

### *Research Implications*

Given the low heritability computed across all four APOs, it appears that environmental or other factors may play an important role in these phenotypes. One possibility is that individual genetic loci each confer small effects to the phenotype as a whole; indeed, our findings are consistent with this hypothesis. The development of statistical methods that leverage such assumptions, such as polygenic risk scoring, may enable better predictive capacities of models. Evolutionarily, given that variants with a strong effect on infant mortality may face strong negative selection, it may be more likely for rare variants of smaller effect to play a larger role in the genetics of preterm birth<sup>29</sup>. Identifying rare causative variants would require sequencing efforts including whole-genome and whole-exome sequencing, as well as larger study populations of diverse genetic ancestry. Additionally, our study has unearthed potential contributors to preterm birth at multiple levels of granularity—SNP, gene, and tissue. Targeted follow-up functional studies of these layers of genetic factors *in vivo* or *in vitro* may better elucidate the underlying biological mechanisms influencing preterm birth phenotypes and ultimately can shape targeted treatments.

### *Strengths and Limitations*

While ours is one of the largest GWAS on traits associated with pregnancy (e.g., it is, to our knowledge, the largest of three GWAS on GDM<sup>30,31</sup>), it highlights the limitations of

prospective studies in a population of highly diverse genetic ancestry. Namely, this is an imbalanced cohort both in terms of cases-controls and genetic ancestries. Our results would benefit from further validation cohorts and deeper sampling of cases, particularly of individuals of non-European genetic ancestry. Larger sample sizes and future meta-analysis would allow us to explore the diversity of types and etiologies of preterm birth and other APOs. As we noted previously, the environment may play a large role in mediating our APOs, and therefore future directions include conducting gene-by-environment interaction studies to further elucidate the interplay between the two.

### *Conclusions*

Our study identified novel genetic associations for three pregnancy phenotypes, reporting new variants correlated with gestational length, gestational diabetes, and pregnancy loss. This study confirms the presence of genetic signals in previously identified pathways for preterm birth, GDM, and preeclampsia. Validation of these new associated SNPs in other cohorts will be important in determining their potential clinical usefulness.

### **Acknowledgements**

We thank Tatiana Foroud and Kymberleigh Pagel for helpful discussion in the planning stage of this work, and Jelena Radivojac for contributing to the design and logistics of DNA extraction and genotyping.

## References

1. Lawn, J. E. & Kinney, M. Preterm birth: now the leading cause of child death worldwide. *Sci. Transl. Med.* **6**, 263ed21 (2014).
2. Martin, J. A., Hamilton, B. E., Osterman, M. J. K. & Driscoll, A. K. Births: Final Data for 2019. *Natl. Vital Stat. Rep.* **70**, 1–51 (2021).
3. Sibai, B., Dekker, G. & Kupferminc, M. Pre-eclampsia. *Lancet* **365**, 785–799 (2005).
4. Deputy, N. P., Kim, S. Y., Conrey, E. J. & Bullard, K. M. Prevalence and Changes in Preexisting Diabetes and Gestational Diabetes Among Women Who Had a Live Birth - United States, 2012–2016. *MMWR Morb. Mortal. Wkly. Rep.* **67**, 1201–1207 (2018).
5. Rossen, L. M., Ahrens, K. A. & Branum, A. M. Trends in Risk of Pregnancy Loss Among US Women, 1990–2011. *Paediatr. Perinat. Epidemiol.* **32**, 19–29 (2018).
6. Kim, C., Newton, K. M. & Knopp, R. H. Gestational Diabetes and the Incidence of Type 2 Diabetes: A systematic review. *Diabetes Care* vol. 25 1862–1868 (2002).
7. Mongraw-Chaffin, M. L., Cirillo, P. M. & Cohn, B. A. Preeclampsia and cardiovascular disease death: prospective evidence from the child health and development studies cohort. *Hypertension* **56**, 166–171 (2010).
8. Catov, J. M. *et al.* Patterns of leisure-time physical activity across pregnancy and adverse pregnancy outcomes. *Int. J. Behav. Nutr. Phys. Act.* **15**, 68 (2018).
9. Catov, J. M. *et al.* Patterns of leisure-time physical activity across pregnancy and adverse pregnancy outcomes. *Int. J. Behav. Nutr. Phys. Act.* **15**, 68 (2018).
10. Goretsky, A. *et al.* Data Preparation of the nuMoM2b Dataset. doi:10.1101/2021.08.24.21262142.
11. Haas, D. M. *et al.* Pregnancy as a Window to Future Cardiovascular Health: Design and Implementation of the nuMoM2b Heart Health Study. *Am. J. Epidemiol.* **183**, 519–530 (2016).
12. Haas, D. M. *et al.* A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). *Am. J. Obstet. Gynecol.* **212**, 539.e1–539.e24 (2015).
13. Facco, F. L. *et al.* Association Between Sleep-Disordered Breathing and Hypertensive Disorders of Pregnancy and Gestational Diabetes Mellitus. *Obstet. Gynecol.* **129**, 31–41 (2017).
14. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
15. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
16. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).
17. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
18. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
19. Rentería, M. E., Cortes, A. & Medland, S. E. Using PLINK for Genome-Wide Association Studies (GWAS) and Data Analysis. *Methods in Molecular Biology* 193–213 (2013) doi:10.1007/978-1-62703-447-0\_8.
20. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
21. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
22. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
23. Zheng, J. *et al.* 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* vol. 33 272–279 (2017).
24. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat.*

- Genet.* **48**, 245–252 (2016).
25. Zeharia, A. *et al.* Acute infantile liver failure due to mutations in the TRMU gene. *Am. J. Hum. Genet.* **85**, 401–407 (2009).
  26. Zhang, Y. *et al.* The zinc finger protein ZBTB20 regulates transcription of fructose-1,6-bisphosphatase 1 and  $\beta$  cell function in mice. *Gastroenterology* **142**, 1571–1580.e6 (2012).
  27. Jansen, R. *et al.* Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum. Mol. Genet.* **26**, 1444–1451 (2017).
  28. Gestational Hypertension and Preeclampsia: ACOG Practice Bulletin, Number 222. *Obstet. Gynecol.* **135**, e237–e260 (2020).
  29. Momozawa, Y. & Mizukami, K. Unique roles of rare variants in the genetics of complex diseases in humans. *J. Hum. Genet.* **66**, 11–23 (2021).
  30. Kwak, S. H. *et al.* Polymorphisms in KCNQ1 are associated with gestational diabetes in a Korean population. *Horm. Res. Paediatr.* **74**, 333–338 (2010).
  31. Moen, G.-H. *et al.* Genetic determinants of glucose levels in pregnancy: genetic risk scores analysis and GWAS in the Norwegian STORK cohort. *Eur. J. Endocrinol.* **179**, 363–372 (2018).
  32. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
  33. Zhang, G. *et al.* Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *N. Engl. J. Med.* **377**, 1156–1167 (2017).
  34. Kovac, U. *et al.* The Association of Polymorphisms in Circadian Clock and Lipid Metabolism Genes With 2 Trimester Lipid Levels and Preterm Birth. *Front. Genet.* **10**, 540 (2019).
  35. Menon, R., Fortunato, S. J., Edwards, D. R. V. & Williams, S. M. Association of genetic variants, ethnicity and preterm birth with amniotic fluid cytokine concentrations. *Ann. Hum. Genet.* **74**, 165–183 (2010).
  36. Chiefari, E., Arcidiacono, B., Foti, D. & Brunetti, A. Gestational diabetes mellitus: an updated overview. *Journal of Endocrinological Investigation* vol. 40 899–909 (2017).
  37. Powe, C. E. & Kwak, S. H. Genetic Studies of Gestational Diabetes and Glucose Metabolism in Pregnancy. *Curr. Diab. Rep.* **20**, 69 (2020).
  38. Thomsen, L. C. V. *et al.* The antihypertensive MTHFR gene polymorphism rs17367504-G is a possible novel protective locus for preeclampsia. *J. Hypertens.* **35**, 132–139 (2017).

**Table 1.** Number of subjects in the nuMoM2b cohort used for the GWAS. Each row represents the number of subjects remaining after applying the step. Bolded numbers represent the final number of subjects used for each separate GWAS (GDM: gestational diabetes, GL: gestational length, PEC: preeclampsia, PL: pregnancy loss).

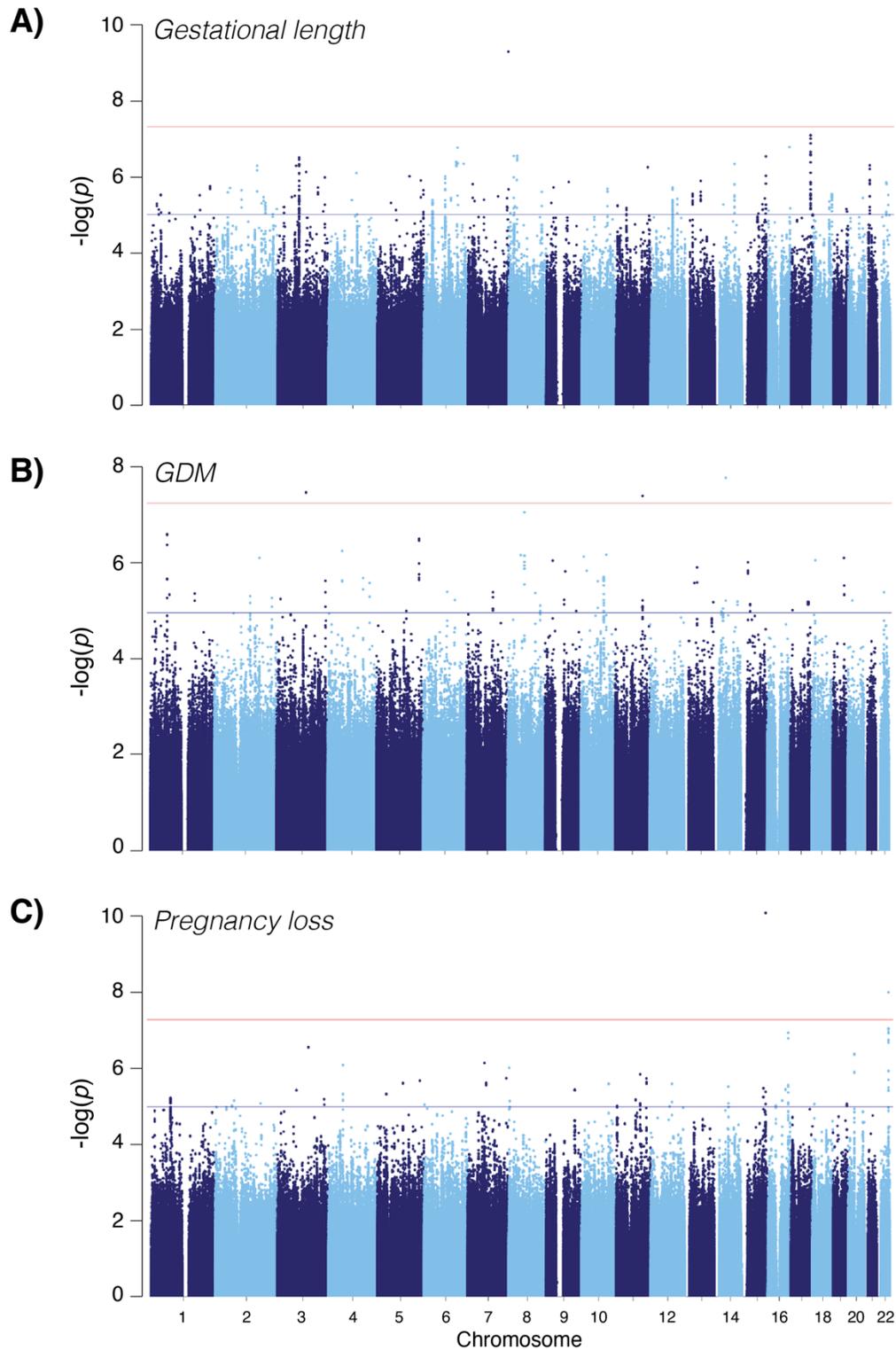
	<b>African Ancestry</b>				<b>Admixed American Ancestry</b>				<b>European Ancestry</b>			
<b>Initial</b>	1425				846				6082			
<b>Preprocessing</b>	1384				811				5896			
<b>After removal of related Subjects</b>	1374				811				5891			
<b>GWAS</b>	GDM	GL	PEC	PL	GDM	GL	PEC	PL	GDM	GL	PEC	PL
<b>Available Phenotypes</b>	<b>1258</b>	1355	<b>1355</b>	<b>1308</b>	<b>754</b>	775	<b>775</b>	<b>775</b>	<b>5605</b>	5726	<b>5779</b>	<b>5726</b>
<b>Live and Spontaneous</b>	-	<b>770</b>	-	-	-	<b>512</b>	-	-	-	<b>3499</b>	-	-

**Table 2.** All genome-wide significant SNPs across the four GWAS analyses performed. The allele frequency at the population level (as reported on Gnomad<sup>32</sup>) is indicated for African ancestry (AFR AF), European ancestry (EUR AF), and Admixed American ancestry (AMR AF). Odds ratio (OR) is computed for the pregnancy loss (logistic regression) GWAS, while beta is computed for the gestational length (GWAS). “GWAS effect” reflects the direction of the odds ratio or beta of the SNP in each sub-cohort (EUR, AFR, AMR; + indicates positive effect, - indicates negative effect, ? indicates the SNP was missing from the population).

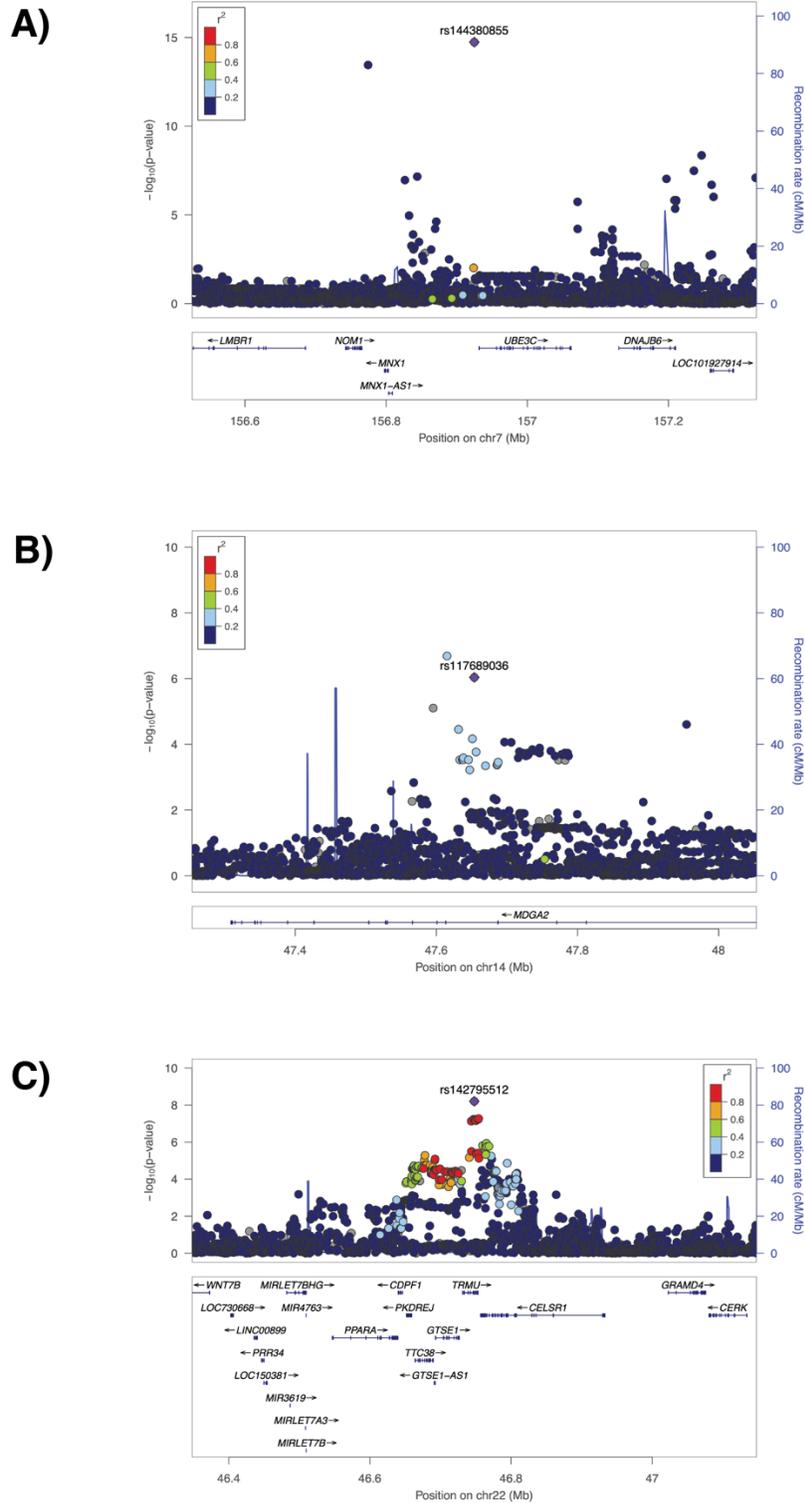
rsID	SNP	AFR AF	EUR AF	AMR AF	Beta/OR	<i>P</i>	GWAS Effect	APO
rs144380855	chr7:157132227:T:C	0.033	0.017	0.017	-2.07	5.3x10 <sup>-10</sup>	?--	Gestational length
rs72956265	chr3:114815630:A:G	0.071	0.010	0.014	3.09	2.98x10 <sup>-8</sup>	+++	GDM
rs10890563	chr11:106675610:C:T	0.24	0.074	0.18	1.88	3.53x10 <sup>-8</sup>	+++	GDM
rs117689036	chr14:47184572:A:C	0.0034	0.023	0.0071	3.15	1.46x10 <sup>-8</sup>	+?+	GDM
rs62021480	chr15:93045079:C:G	0.178	0.057	0.027	3.29	7.83x10 <sup>-11</sup>	++?	Pregnancy loss
rs142795512	chr22:46351905:C:T	0.011	0.025	0.025	4.72	9.64x10 <sup>-9</sup>	++?	Pregnancy loss

**Table 3.** Previously identified variants that were validated in this study (association with relevant APO  $P < 0.05$ ). Sample size (n), value of normalized statistic (Z-score) and significance (P-value) are shown for the meta-analysis of three sub-populations.

APO	rsID	SNP	n	Z-score	P-value	Gene	Reference
Gestational length	rs12037376	chr1:22135618:G:A	4076	3.61	$3.14 \times 10^{-4}$	<i>WNT4</i>	33
	rs56318008	chr1:22143914:C:T	4076	3.63	$2.84 \times 10^{-4}$	<i>WNT4</i>	33
	rs2963463	chr5:158468041:T:C	4076	-2.88	0.004	<i>EBF1</i>	33
	rs2946171	chr5:158494932:T:G	4076	-2.80	0.005	<i>EBF1</i>	33
	rs2946169	chr5:158491951:C:T	4076	-2.58	0.010	<i>EBF1</i>	33
	rs228669	chr1:7809988:T:C	4076	2.29	0.022	<i>PER3</i>	34
	rs1800896	chr1:206773552:T:C	4076	2.21	0.021	<i>IL10</i>	35
	rs9861425	chr3:123354036:A:C	4076	-1.97	0.048	<i>ADCY5</i>	33
GDM	rs4506565	chr10:112996282:A:T	7602	3.17	0.0015	<i>TCF7L2</i>	36,37
Preeclampsia	rs17367504	chr1:11802721:A:G	7894	-2.25	0.024	<i>MTHFR</i>	38



**Figure 1.** Manhattan Plot of GWAS Results: GWA meta-analysis was carried out using the inverse variance-weighted method implemented in GWAMA. The solid red line indicates the threshold for genome-wide significance ( $P < 5 \times 10^{-8}$ ), and the blue line represents the threshold for suggestive hits ( $P < 1 \times 10^{-5}$ ). A) Gestational length; B) GDM; C) Pregnancy loss.



**Figure 2.** Selective regional association plots demonstrating regional linkage disequilibrium for A) rs144380855 associated with gestational length, B) rs117689036 associated with gestational diabetes, and C) pregnancy-loss associated SNP rs142795512