

## Searching and visualizing genetic associations of pregnancy traits by using GnuMoM2b

Qi Yan<sup>1,\*</sup>, Rafael F. Guerrero<sup>2,3</sup>, Raiyan R. Khan<sup>4</sup>, Andy A. Surujnarine<sup>1</sup>, Ronald J. Wapner<sup>1</sup>, Matthew W. Hahn<sup>2,5</sup>, Anita Raja<sup>6</sup>, Ansaf SallebAouissi<sup>4</sup>, William A. Grobman<sup>7</sup>, Hyagriv Simhan<sup>8</sup>, Nathan R. Blue<sup>9</sup>, Robert Silver<sup>9</sup>, Judith H. Chung<sup>10</sup>, Uma M. Reddy<sup>1</sup>, Predrag Radivojac<sup>11</sup>, Itsik Pe'er<sup>4</sup>, David M. Haas<sup>12</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, New York, NY 10032,

<sup>2</sup>Department of Computer Science, Indiana University, Bloomington, IN 47405,

<sup>3</sup>Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695,

<sup>4</sup>Department of Computer Science, Columbia University, New York, NY 10032,

<sup>5</sup>Department of Biology, Indiana University, Bloomington, IN 47405,

<sup>6</sup>Department of Computer Science, CUNY Hunter College, New York, NY 10065,

<sup>7</sup>Department of Obstetrics and Gynecology, College of Medicine, The Ohio State University, Columbus, OH 43210,

<sup>8</sup>Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh, Pittsburgh, PA 15213,

<sup>9</sup>Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT 84132,

<sup>10</sup>Department of Obstetrics and Gynecology, University of California, Irvine, Orange, CA 92697,

<sup>11</sup>Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115,

<sup>12</sup>Department of Obstetrics and Gynecology, Indiana University, Indianapolis, IN 46202, USA

\*Corresponding authors:

Qi Yan, Ph.D.

Assistant Professor

Department of Obstetrics and Gynecology

Columbia University Irving Medical Center

212-305-8738

[qy2253@cumc.columbia.edu](mailto:qy2253@cumc.columbia.edu)

## **Abstract**

Adverse pregnancy outcomes (APOs) are major risk factors for women's health during pregnancy and even in the years after pregnancy. Due to the heterogeneity of APOs, only few genetic associations have been identified. In this report, we conducted genome-wide association studies (GWAS) of 479 traits that are possibly related to APOs using a large and racially diverse study, Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b). To display the extensive results, we developed a web-based tool GnuMoM2b (<https://gnumom2b.cumcobbgyn.org/>) for searching, visualizing, and sharing results from GWAS of 479 pregnancy traits as well as phenome-wide association studies (PheWAS) of more than 17 million single nucleotide polymorphisms (SNPs). The genetic results from three ancestries (Europeans, Africans, and Admixed Americans) and meta-analyses are populated in GnuMoM2b. In conclusion, GnuMoM2b is a valuable resource for extraction of pregnancy-related genetic results and shows the potential to facilitate meaningful discoveries.

## Introduction

Pregnancy is a unique window into women's future health. Adverse pregnancy outcomes (APOs) are frequent and are related to adverse health, such as hypertension, even years after pregnancy (JOWELL *et al.* 2022). The burden of APOs varies across racial groups in the United States. For example, Black and Hispanic women have a greater prevalence of APOs than White women (CHO *et al.* 2020). To identify genetic determinants of APOs, we recently conducted genome-wide association studies (GWAS) of four APOs (GUERRERO *et al.* 2022) (preterm birth, preeclampsia, gestational diabetes, and pregnancy loss) using samples from a large and racially diverse study, Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b) (HAAS *et al.* 2015), gathered from eight clinical centers in the United States. However, few single nucleotide polymorphisms (SNPs) were identified, probably due to the heterogeneity of APOs. For example, preeclampsia is a heterogeneous APO characterized by high blood pressure and proteinuria. It can be classified by severity (mild, severe, superimposed, and eclampsia) and onset time (early or late), and dichotomizing it into cases and controls may affect statistical power, as the genetic influence on mild preeclampsia could be closer to controls than other cases. Thus, to provide additional insights into genetic influence on APOs, we conducted GWAS of 479 traits that are possibly related to APOs using nuMoM2b.

To display such extensive results, we have developed a web application, GnuMoM2b (<https://gnumom2b.cumcobbgyn.org/>), for interactively searching, visualizing, and sharing GWAS results. In addition, GnuMoM2b searches for traits associated with any particular SNP to expose pleiotropy across 479 traits (i.e., phenome-wide association study [PheWAS]).

## Results

The GWAS summary statistics of nuMoM2b traits from three individual ancestry group analyses and meta-analyses are populated in GnuMoM2b. The starting page includes usage guidelines. Descriptions of all nuMoM2b traits are provided and searchable by key words. The web interface offers two interactive visualizations: GWAS results queried by trait and PheWAS results queried by SNP. Multiple filters can be used to customize the output, including ancestry selection,  $P$ -value cutoff, minor allele frequency (MAF) cutoff, and number of studies contributing to meta-analysis. The results can be downloaded in text files. GnuMoM2b also links to NCBI dbSNP to provide additional information on a particular SNP. Furthermore, the genomic control (GC) value will appear after loading a GWAS. The density plot (Figure S1) of GC values from all GWAS indicates well-controlled continuous traits across ancestry-specific and meta-analyses and binary traits in the European (EUR) analysis, with slight deflation in some binary traits in the African (AFR), Admixed American (AMR), and meta-analyses, likely due to smaller sample sizes in AFR and AMR. Therefore, users should consistently check GC values.

GnuMoM2b can facilitate meaningful discoveries. In the online example, GWAS shows that rs988551 in *LAMA2* is the most significant SNP ( $P$ -value= $4.7 \times 10^{-9}$ ) associated with gestational hypertension (Trait ID: *acog\_PEGHTN\_7*) from the meta-analysis ( $N_{EUR}=5,636$ ,  $N_{AFR}=1,278$ , and  $N_{AMR}=759$ ) (Figure S2), and PheWAS shows that rs988551 is also associated with other hypertension-related traits ( $P$ -value $<1 \times 10^{-3}$ ). Lately, *LAMA2* has been implicated to be a preeclampsia-dysregulated gene (ZHOU *et al.* 2019). In a recent large GWAS of preeclampsia and gestational hypertension (HONIGBERG *et al.* 2022), rs167479 in *RGL3* was found to be associated with both conditions in the discovery analysis. Our GnuMoM2b replicates this SNP association at a nominal level ( $P$ -value= $1.2 \times 10^{-3}$ ) with gestational hypertension. Note that nuMoM2b

participated as a follow-up cohort in that study, but was not included in the discovery analysis. As another example, a large GWAS reported that *WNT4* was associated with gestational length, and the subsequent functional analysis suggested that *WNT4* was a key regulator of decidualization of the human endometrial stromal cell and subsequent embryo implantation (ZHANG *et al.* 2017). The rs12037376 in *WNT4* was reported in their paper. As their paper focused on European ancestry, we examined the EUR results in GnuMoM2b. In our PheWAS results, we replicated this SNP at a nominal level ( $P$ -value= $6.8 \times 10^{-4}$ ) in association with gestational length (Trait ID: *GawksEND*). Given that cervical length is a crucial determinant of gestational length (BERGHELLA *et al.* 2003), we further examined the association between this SNP and cervical length at 22-29 weeks (Trait ID: *U3BB02*), resulting in a genome-wide significant  $P$ -value= $1.0 \times 10^{-15}$  in EUR. In GnuMoM2b, the meta-analysis of GWAS of cervical length reveals that the most significant SNP is rs12404660 ( $P$ -value= $2.3 \times 10^{-12}$ ) in *WNT4* (Figure S3), and PheWAS shows its association with gestational length is  $P$ -value=0.01. However, the association of both rs12037376 and rs12404660 in *WNT4* with gestational length is not significant at the genome-wide level, probably because stillbirth, fetal demise, elective termination, and indicated termination subjects were not excluded from the gestational length analysis. Thus, GnuMoM2b should be used as an exploration tool to search for preliminary results for further analyses. To reach any solid conclusions, refinement of phenotype definitions and statistical analyses as well as experimental validations are needed. Therefore, caution is needed when interpreting the results from GnuMoM2b. Moreover, the nuMoM2b study measured 9 placental analytes at 2 visits (6-13 weeks and 16-21 weeks) searchable under the “Trait Description” tab. The sample sizes ( $N_{EUR} > 1,000$ ,  $N_{AFR} < 400$ , and  $N_{AMR} < 200$ ) indicate that it could be only reasonable to conduct GWAS in EUR. The GWAS results in EUR show multiple significant loci associated with

placental analytes, such as endoglin and sFlt-1 at visit 1, ADAM-12 at visit 2, and fbHCG at both visits (Figure S4). With the advent of summarized GWAS data and two-sample Mendelian randomization (MR) methods, conducting MR analyses have become easier for data scientists. Coupled with external large GWAS of APOs, GnuMoM2b results offer an opportunity to investigate the causal relations between placental analytes and APOs through MR with SNPs as instrumental variables. However, caution is needed to conduct a credible MR analysis and provide a reasonable interpretation.

In addition to searching and visualizing results online, GnuMoM2b allows users to download the summary statistics of three individual ancestry analyses and meta-analyses, which offers users an opportunity to perform summary data-based analysis, such as fine-mapping, colocalization, and MR.

## **Discussion**

We have conducted GWAS of 479 pregnancy-related traits using the nuMoM2b study and hosted the results on GnuMoM2b that is an easy-to-use web-based application for searching, visualizing, and sharing both genome-wide and phenome-wide association results. The nuMoM2b is a comprehensive cohort with racially diverse participants, offering in-depth clinical and psychosocial phenotyping and longitudinal follow-up during pregnancy. This cohort presents a unique opportunity for a holistic approach to investigating genetic and environmental factors contributing to population morbidity originating in pregnancy. Our GnuMoM2b results complement other large-scale biobank studies, such as the UK Biobank (BYCROFT *et al.* 2018), which lack detailed pregnancy phenotypes. These results greatly expand our understanding of genetic influence on pregnancy traits. GnuMoM2b offers researchers, and obstetricians in

particular, a new resource to readily extract pregnancy-related genetic results. In the future, we will further add nuMoM2b Heart Health Study (nuMoM2b-HHS) genetic results to GnuMoM2b. The nuMoM2b-HHS, a follow-up study of nuMoM2b, is conducted in a subset of nuMoM2b women, 2-7 years after delivery to better understand the impact of pregnancy outcomes on future health (HAAS *et al.* 2016). Lastly, we want to emphasize that GnuMoM2b is a valuable resource for pregnancy data exploration, but understanding the existing biases (e.g., small sample size in particular ancestries, suboptimal trait definitions, and spurious results due to small MAF) is important when interpreting the results, and further analyses are needed to reach any conclusions.

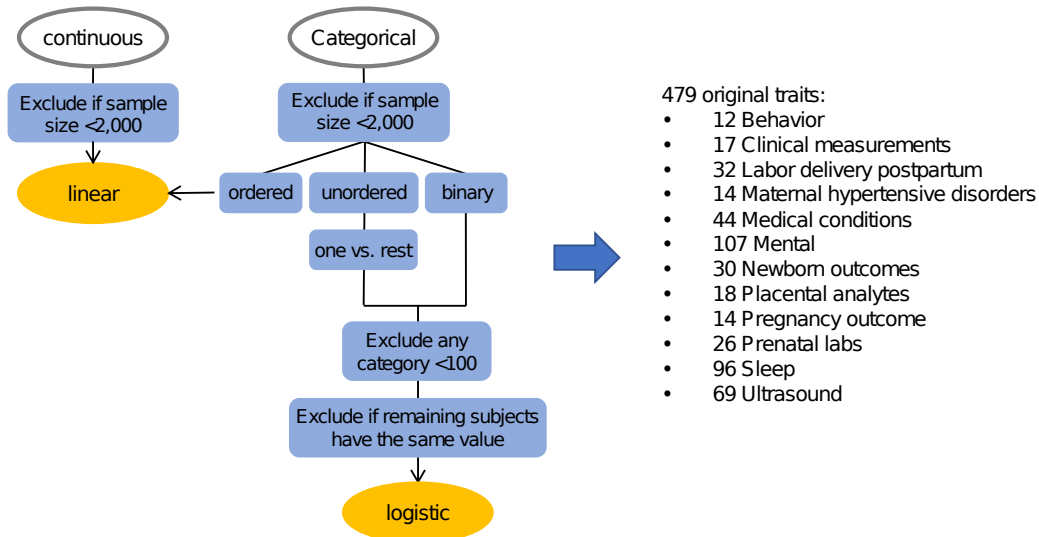
## **Materials and methods**

The nuMoM2b study enrolled 10,038 women between 2010-2013 from eight centers in the United States. The study aimed to recruit a large and racially diverse cohort of nulliparous pregnant women (Table S1). Its main objective was to identify maternal characteristics, such as genetic factors, physiological responses, and environmental factors that can predict APOs (HAAS *et al.* 2015). Participants were followed longitudinally and underwent four study visits from the first trimester to delivery. Throughout pregnancy, various data were collected, such as interviews, questionnaires, research ultrasounds, maternal biometric measurements, and biospecimens (Table S2). The nuMoM2b study methods have been described in detail elsewhere (HAAS *et al.* 2015), and the study was approved by the Institutional Review Boards at all participating centers (GUERRERO *et al.* 2022). Genome-wide genotyping was conducted using the Infinium Multi-Ethnic Global D2 BeadChip (Illumina, Miami, USA). Quality control measures included assessing sex inconsistencies, autosome missingness, and contamination. Using KING-Robust

(MANICHAIKUL *et al.* 2010), kinship was inferred and one participant from each pair with first- or second-degree relatedness was randomly removed. Instead of directly using self-reported race, ancestry was determined using SNPweights, leveraging data from 1000 Genomes Project (GENOMES PROJECT *et al.* 2015), resulting in five ancestry groups: EUR (n=6,082), AFR (n=1,425), AMR (n=846), East Asian (EAS; n=323), and South Asian (SAS; n=112). Due to insufficient sample sizes, EAS and SAS were excluded from downstream GWAS. Furthermore, genotype imputation was performed with the TOPMed Imputation Server (<https://imputation.biodatacatalyst.nih.gov/>). Other details about genome-wide genotyping, genotype imputation, quality control, and ancestry estimation were previously described (GUERRERO *et al.* 2022). In this study, after quality control, 8,076 independent subjects (estimated ancestries: 5,891 EUR, 1,374 AFR, and 811 AMR) and 17,177,813 genotyped and imputed SNPs were included in the analyses. The mean age ( $\pm$ standard deviation) at visit 1 was 28.06 $\pm$ 5.23 for EUR, 23.42 $\pm$ 5.38 for AFR, and 24.75 $\pm$ 5.66 for AMR participants. In selecting traits, our aim was to include as many genetically related traits as possible. We first excluded traits that had no genetic relevance, and further filtering was done based on sample size (Figure 1). In the end, we retained 479 traits. Detailed procedures are provided in the supplementary material. The analyses were conducted with PLINK software, version 2 (CHANG *et al.* 2015), using linear regression for continuous or ordered categorical traits and logistic regression for binary traits under an additive genetic model, adjusting for maternal age, distance to median maternal age, and the first 10 principal components calculated using genotypic data. For each trait, GWAS was performed in three ancestry groups separately, and only SNPs with MAF>0.01 and Hardy-Weinberg equilibrium (HWE)  $P$ -value> $1\times 10^{-3}$  were kept. Next, results from three individual ancestry groups for each trait were pooled using the DerSimonian and Laird method



for meta-analyses with random effects implemented in METAL (WILLER *et al.* 2010; HEMANI 2022), which allows different true effect sizes across three ancestry groups for each SNP.



**Figure 1. Preprocessing of GWAS traits.** The continuous or ordered categorical traits were analyzed using linear regression models. The unordered categorical traits were recoded into several binary traits following the one vs. rest manner, and the binary traits were analyzed using logistic regression models.

Manhattan plots are standard tools to visualize  $P$ -values of GWAS on a genome-wide level and identify genetic loci associated with the trait. GnuMoM2b is developed to highlight top genetic hits and draw Manhattan plots of GWAS as well as PheWAS using R shiny (<http://shiny.rstudio.com/>). This web application is easy to use through a graphical user interface.

### Data availability

The summary-level GWAS and PheWAS data are available and can be downloaded at GnuMoM2b (<https://gnumom2b.cumcobb.org/>). The individual-level phenotype and genotype data have been reported earlier (HAAS *et al.* 2015; GUERRERO *et al.* 2022) and are available upon request.

## **Funding**

Support for performing DNA extraction and GWAS from the Indiana University Grand Challenges Precision Diabetes project funding. D.M.H. was partially funded by R01HD101246 from the U.S. National Institutes of Health (NIH). Original funding for nuMoM2b sample and data collection are noted in the study methods paper (HAAS *et al.* 2015).

## **Conflict of interest**

None declared.

## REFERENCES

- Berghella, V., G. Bega, J. E. Tolosa and M. Berghella, 2003 Ultrasound assessment of the cervix. *Clin Obstet Gynecol* 46: 947-962.
- Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott *et al.*, 2018 The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562: 203-209.
- Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell *et al.*, 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: 7.
- Cho, L., M. Davis, I. Elgendy, K. Epps, K. J. Lindley *et al.*, 2020 Summary of Updated Recommendations for Primary Prevention of Cardiovascular Disease in Women: JACC State-of-the-Art Review. *J Am Coll Cardiol* 75: 2602-2618.
- Genomes Project, C., A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison *et al.*, 2015 A global reference for human genetic variation. *Nature* 526: 68-74.
- Guerrero, R. F., R. R. Khan, R. J. Wapner, M. W. Hahn, A. Raja *et al.*, 2022 Genetic Polymorphisms Associated with Adverse Pregnancy Outcomes in Nulliparas. medRxiv.
- Haas, D. M., D. B. Ehrenthal, M. A. Koch, J. M. Catov, S. E. Barnes *et al.*, 2016 Pregnancy as a Window to Future Cardiovascular Health: Design and Implementation of the nuMoM2b Heart Health Study. *Am J Epidemiol* 183: 519-530.
- Haas, D. M., C. B. Parker, D. A. Wing, S. Parry, W. A. Grobman *et al.*, 2015 A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). *Am J Obstet Gynecol* 212: 539 e531-539 e524.
- Hemani, G., 2022 explodecomputer/random-metal: Adding random effects model (v0.1.0). Zenodo.
- Honigberg, M. C., B. Truong, R. R. Khan, B. Xiao, L. Bhatta *et al.*, 2022 Genome-wide meta-analysis identifies novel maternal risk variants and enables polygenic prediction of preeclampsia and gestational hypertension. medRxiv.
- Jowell, A. R., A. A. Sarma, M. Gulati, E. D. Michos, A. J. Vaught *et al.*, 2022 Interventions to Mitigate Risk of Cardiovascular Disease After Adverse Pregnancy Outcomes: A Review. *JAMA Cardiol* 7: 346-355.
- Manichaikul, A., J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale *et al.*, 2010 Robust relationship inference in genome-wide association studies. *Bioinformatics* 26: 2867-2873.
- Willer, C. J., Y. Li and G. R. Abecasis, 2010 METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190-2191.
- Zhang, G., B. Feenstra, J. Bacelis, X. Liu, L. M. Muglia *et al.*, 2017 Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *N Engl J Med* 377: 1156-1167.
- Zhou, C., Q. Yan, Q. Y. Zou, X. Q. Zhong, C. T. Tyler *et al.*, 2019 Sexual Dimorphisms of Preeclampsia-Dysregulated Transcriptomic Profiles and Cell Function in Fetal Endothelial Cells. *Hypertension* 74: 154-163.