OXFORD GENETICS

Searching and visualizing genetic associations of pregnancy traits by using GnuMoM2b

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

[1]Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, New York, NY 10032, USA
[2]Department of Computer Science, Indiana University, Bloomington, IN 47405, USA
[3]Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA
[4]Department of Computer Science, Columbia University, New York, NY 10032, USA
[5]Department of Biology, Indiana University, Bloomington, IN 47405, USA
[6]Department of Computer Science, CUNY Hunter College, New York, NY 10065, USA
[7]Department of Obstetrics and Gynecology, College of Medicine, The Ohio State University, Columbus, OH 43210, USA
[8]Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh, Pittsburgh, PA 15213, USA
[9]Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT 84132, USA
[10]Department of Obstetrics and Gynecology, University of California, Irvine, Orange, CA 92697, USA
[11]Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA
[12]Department of Obstetrics and Gynecology, Indiana University, Indianapolis, IN 46202, USA

*Corresponding author: Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, 622 W 168th St, New York, NY 10032, USA. Email: qy2253@cumc.columbia.edu

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 (GWASs) 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 extensive results, we developed a web-based tool GnuMoM2b (https://gnumom2b.cumcobgyn.org/) for searching, visualizing, and sharing results from a GWAS of 479 pregnancy traits as well as phenome-wide association studies of more than 17 million single nucleotide polymorphisms. The genetic results from 3 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.

Keywords: GWAS; PheWAS; adverse pregnancy outcomes; nuMoM2b

Introduction

Pregnancy is a unique window into a woman'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 the genetic determinants of APOs, we recently conducted genome-wide association studies (GWASs) of 4 APOs (preterm birth, preeclampsia, gestational diabetes, and pregnancy loss; Guerrero *et al.* 2022) using samples from a large and racially diverse study, Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b; Haas *et al.* 2015), gathered from 8 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 heterogenous 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 the genetic influence on APOs, we conducted a 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.cumcobgyn.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)].

Materials and methods

The nuMoM2b study enrolled 10,038 women between 2010 and 2013 from 8 centers in the United States. This study aimed to recruit a large and racially diverse cohort of nulliparous pregnant women (Supplementary Table 1). 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 4 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 (Supplementary Table 2). 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, FL, USA). Quality control measures included assessing sex inconsistencies, autosome missingness, and contamination. Using KING-Robust (Manichaikul *et al.* 2010), kinship was inferred and 1 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 the 1000 Genomes Project Consortium (2015), resulting in 5 ancestry groups: European (EUR; $n = 6,082$), African (AFR; $n = 1,425$), Admixed American (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 GWASs. Furthermore, genotype imputation was performed with the TOPMed Imputation Server (https://imputation.biodatacatalyst.nhlbi.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 (±standard deviation) at visit 1 was 28.06 ± 5.23 for EUR, 23.42 ± 5.38 for AFR, and 24.75 ± 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 (Fig. 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, a GWAS was performed in 3 ancestry groups separately, and only SNPs with minor allele frequency (MAF) > 0.01 and Hardy–Weinberg equilibrium $P > 1 \times 10^{-3}$ were kept. Next, results from 3 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 3 ancestry groups for each SNP.

Manhattan plots are standard tools to visualize the *P*-values of a 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 a GWAS as well as a PheWAS using R shiny (http://shiny.rstudio.com/). This web application is easy to use through a graphical user interface.

Results

The GWAS summary statistics of numoM2b traits from 3 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 2 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, MAF cutoff, and the number of studies contributing to the meta-analysis. The results can be downloaded as 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 (Supplementary Fig. 1) of GC values from all GWASs indicates well-controlled continuous traits across ancestry-specific and meta-analyses and binary traits in the EUR analysis, with slight deflation in some binary traits in the AFR, 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, the GWAS shows that rs988551 in *LAMA2* is the most significant SNP ($P = 4.7 \times 10^{-9}$) associated with gestational hypertension (Trait ID: *acog_PEgHTN_7*) from the meta-analysis ($N_{EUR} = 5,636$, $N_{AFR} = 1,278$, and $N_{AMR} = 759$; Supplementary Fig. 2), and the PheWAS shows that rs988551 is also associated with other hypertension-related traits ($P < 1 \times 10^{-3}$). Lately, *LAMA2* has been implicated as 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 ($P = 1.2 \times 10^{-3}$) 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 ($P = 6.8 \times 10^{-4}$) 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 $P = 1.0 \times 10^{-15}$ in EUR. In GnuMoM2b, a meta-analysis of the GWAS of cervical length reveals that the most significant SNP is rs12404660 ($P = 2.3 \times 10^{-12}$) in *WNT4* (Supplementary Fig. 3), and the PheWAS shows that its association with gestational length is $P = 0.01$. 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 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 a 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 (Supplementary Fig. 4). With the advent of summarized GWAS data and 2-sample Mendelian randomization (MR)
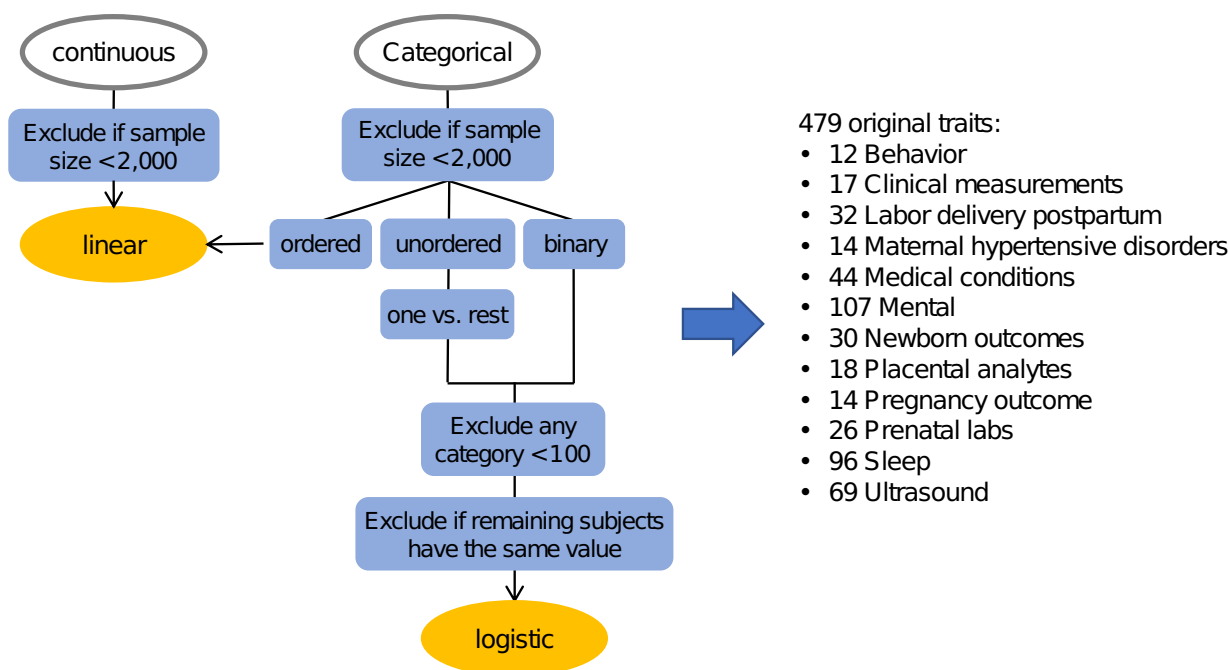
Fig. 1. Preprocessing of GWAS traits. Continuous or ordered categorical traits were analyzed using linear regression models. The unordered categorical traits were recoded into several binary traits following the one vs rest manner, and the binary traits were analyzed using logistic regression models.

methods, conducting MR analyses have become easier for data scientists. Coupled with externally large GWASs 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 3 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 a GWAS of 479 pregnancy-related traits using the nuMoM2b study and hosted the results on GnuMoM2b, which 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 investigate 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 influences 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.

Data availability

The summary-level GWAS and PheWAS data are available and can be downloaded at GnuMoM2b (https://gnumom2b.cumcobgyn.org/). The individual-level data are available through the NICHD's Data and Specimen Hub (DASH) at https://doi.org/10.57982/gjxm-yz73.

Supplemental material available at GENETICS online.

Funding

Conflicts of interest

The author(s) declare no conflict of interest.

Literature cited

The 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Abecasis GR, Altshuler DM, Bentley DR, Chakravarti A, Clark AG, et al. A global reference for human

genetic variation. Nature. 2015;526(7571):68–74. doi:10.1038/nature15393.

Berghella V, Bega G, Tolosa JE, Berghella M. Ultrasound assessment of the cervix. Clin Obstet Gynecol. 2003;46(4):947–962. doi:10.1097/00003081-200312000-00026.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–209. doi:10.1038/s41586-018-0579-z.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):7. doi:10.1186/s13742-015-0047-8.

Cho L, Davis M, Elgendy I, Epps K, Lindley KJ, Mehta PK, Michos ED, Minissian M, Pepine C, Vaccarino V, et al. Summary of updated recommendations for primary prevention of cardiovascular disease in women: JACC state-of-the-art review. J Am Coll Cardiol. 2020;75(20):2602–2618. doi:10.1016/j.jacc.2020.03.060.

Guerrero RF, Khan RR, Wapner RJ, Hahn MW, Raja A, Salleb-Aouissi A, Grobman WA, Simhan H, Silver R, Chung JH, et al. Genetic polymorphisms associated with adverse pregnancy outcomes in nulliparas. medRxiv 22271641. https://doi.org/10.1101/2022.02.28.22271641, 1 March 2022, preprint: not peer reviewed..

Haas DM, Ehrenthal DB, Koch MA, Catov JM, Barnes SE, Facco F, Parker CB, Mercer BM, Bairey-Merz CN, Silver RM, et al. Pregnancy as a window to future cardiovascular health: design and implementation of the nuMoM2b Heart Health Study. Am J Epidemiol. 2016;183(6):519–530. doi:10.1093/aje/kwv309.

Haas DM, Parker CB, Wing DA, Parry S, Grobman WA, Mercer BM, Simhan HN, Hoffman MK, Silver RM, Wadhwa P, et al. A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). Am J Obstet Gynecol. 2015;212:539.e1–539.e24.

Hemani G. explodecomputer/random-metal: Adding random effects model (v0.1.0). Zenodo. 2022. doi:10.5281/zenodo.6974696.

Honigberg MC, Truong B, Khan RR, Xiao B, Bhatta L, Vy HMT, Guerrero RF, Schuermans A, Selvaraj MS, Patel AP, et al. Polygenic prediction of preeclampsia and gestational hypertension. Nat Med. 2023;29(6):1540–1549. doi:10.1038/s41591-023-02374-9.

Jowell AR, Sarma AA, Gulati M, Michos ED, Vaught AJ, Natarajan P, Powe CE, Honigberg MC. Interventions to mitigate risk of cardiovascular disease after adverse pregnancy outcomes: a review. JAMA Cardiol. 2022;7(3):346–355. doi:10.1001/jamacardio.2021.4391.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867–2873. doi:10.1093/bioinformatics/btq559.

Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190–2191. doi:10.1093/bioinformatics/btq340.

Zhang G, Feenstra B, Bacelis J, Liu X, Muglia LM, Juodakis J, Miller DE, Litterman N, Jiang P-P, Russell L, et al. Genetic associations with gestational duration and spontaneous preterm birth. N Engl J Med. 2017;377(12):1156–1167. doi:10.1056/NEJMoa1612665.

Zhou C, Yan Q, Zou QY, Zhong XQ, Tyler CT, Magness RR, Bird IM, Zheng J. Sexual dimorphisms of preeclampsia-dysregulated transcriptomic profiles and cell function in fetal endothelial cells. Hypertension. 2019;74(1):154–163. doi:10.1161/HYPERTENSIONAHA.118.12569.

Editor: Y. Li