



OPEN

# Genetic predictors of blood pressure traits are associated with preeclampsia

Elizabeth A. Jasper<sup>1,2,4,5</sup>, Jacklyn N. Hellwege<sup>3,4,6</sup>, Joseph H. Breeyear<sup>3</sup>, Brenda Xiao<sup>7</sup>, Gail P. Jarvik<sup>8</sup>, Ian B. Stanaway<sup>9</sup>, Kathleen A. Leppig<sup>10</sup>, Geetha Chittoor<sup>11</sup>, M. Geoffrey Hayes<sup>12,13,14</sup>, Ozan Dikilitas<sup>15</sup>, Iftikhar J. Kullo<sup>16</sup>, Ingrid A. Holm<sup>17,18</sup>, Shefali Setia Verma<sup>19</sup>, Todd L. Edwards<sup>4,20</sup> & Digna R. Velez Edwards<sup>1,4,5,21</sup>✉

Preeclampsia, a pregnancy complication characterized by hypertension after 20 gestational weeks, is a major cause of maternal and neonatal morbidity and mortality. Mechanisms leading to preeclampsia are unclear; however, there is evidence of high heritability. We evaluated the association of polygenic scores (PGS) for blood pressure traits and preeclampsia to assess whether there is shared genetic architecture. Non-Hispanic Black and White reproductive age females with pregnancy indications and genotypes were obtained from Vanderbilt University's BioVU, Electronic Medical Records and Genomics network, and Penn Medicine Biobank. Preeclampsia was defined by ICD codes. Summary statistics for diastolic blood pressure (DBP), systolic blood pressure (SBP), and pulse pressure (PP) PGS were acquired from Giri et al. Associations between preeclampsia and each PGS were evaluated separately by race and data source before subsequent meta-analysis. Ten-fold cross validation was used for prediction modeling. In 3504 Black and 5009 White included individuals, the rate of preeclampsia was 15.49%. In cross-ancestry meta-analysis, all PGSs were associated with preeclampsia ( $OR_{DBP} = 1.10$ , 95% CI 1.02–1.17,  $p = 7.68 \times 10^{-3}$ ;  $OR_{SBP} = 1.16$ , 95% CI 1.09–1.23,  $p = 2.23 \times 10^{-6}$ ;  $OR_{PP} = 1.14$ , 95% CI 1.07–1.27,  $p = 9.86 \times 10^{-5}$ ). Addition of PGSs to clinical prediction models did not improve predictive performance. Genetic factors contributing to blood pressure regulation in the general population also predispose to preeclampsia.

Preeclampsia, a pregnancy complication characterized by hypertension after 20 gestational weeks, is a major cause of maternal and neonatal morbidity and mortality. Globally, preeclampsia is estimated to complicate 2.0% to 8.0% of all pregnancies<sup>1</sup>. In the United States (U.S.), the prevalence of preeclampsia increased by 25%

<sup>1</sup>Division of Quantitative and Clinical Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 600, Rm 616, Nashville, TN 37203, USA. <sup>2</sup>Center for Precision Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>3</sup>Present address: Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA. <sup>4</sup>Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, TN, USA. <sup>5</sup>Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>6</sup>Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>7</sup>Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA. <sup>8</sup>Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA, USA. <sup>9</sup>Division of Nephrology and Harborview Medical Center Kidney Research Institute, Department of Medicine, University of Washington Medical Center, Seattle, WA, USA. <sup>10</sup>Genetic Services, Kaiser Permanente of Washington, Seattle, WA, USA. <sup>11</sup>Department of Population Health Sciences, Geisinger, Danville, PA, USA. <sup>12</sup>Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>13</sup>Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>14</sup>Department of Anthropology, Northwestern University, Evanston, IL, USA. <sup>15</sup>Departments of Internal Medicine, Cardiovascular Medicine, Mayo Clinician-Investigator Training Program, Mayo Clinic, Rochester, MN, USA. <sup>16</sup>Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA. <sup>17</sup>Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA. <sup>18</sup>Department of Pediatrics, Harvard Medical School, Boston, MA, USA. <sup>19</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>20</sup>Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>21</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, USA. ✉email: digna.r.velez.edwards@vumc.org

between 1987 and 2004<sup>2</sup>. The risk of preeclampsia has also increased over the years, with women giving birth in 2003 at a 6.7-fold increased risk of preeclampsia compared to women who gave birth in 1980<sup>3</sup>. In addition to the rising rates and risk of preeclampsia, the condition is costly. In the U.S. in 2012, one study estimated the cost of preeclampsia in the year following delivery to be 2.18 billion dollars<sup>4</sup>. In addition to its immediate economic cost, the pregnancy complication has massive effects on the health of the mother and infant. Risk of stillbirth, neonatal death, fetal growth restriction, lower birth weights, preterm birth, and perinatal mortality is higher for women with preeclampsia<sup>5,6</sup>. There is growing evidence that preeclampsia is not just an issue of pregnancy, rather it effects the long-term health of mothers and infants. Preeclamptic women are later at increased risk of developing and dying from cardiovascular disease while children exposed to preeclampsia have higher blood pressures, body mass index (BMI), and risk of hypertension<sup>5,7</sup>.

Pre-existing maternal conditions, pregnancy-specific characteristics, environmental factors, and family history have been identified as risk factors for preeclampsia. Increasing age of mothers, higher BMI, pre-existing hypertension, and diabetes all place women at increased risk of preeclampsia during pregnancy<sup>8–11</sup>. Having preeclampsia in a prior pregnancy and/or a family history of preeclampsia are significant risk factors for preeclampsia. The odds of preeclampsia are 2.9 (95% confidence interval (CI) 1.70–4.93) times greater in individuals who have experienced it in a previous pregnancy compared to those without a personal history of the condition. Individuals with family members who had preeclampsia are 7.19 (95% CI 5.85–8.83) times more likely to develop it compared to those without a family history<sup>9–11</sup>. Heritability studies of preeclampsia present strong support for a genetic basis for preeclampsia with twin studies estimating heritability for preeclampsia to be approximately 55%<sup>12</sup>. Several small genome wide association studies (GWAS) have been performed, with less than 50 associated polymorphisms identified to date<sup>13–15</sup>. However, identified risk factors and polymorphisms have yet to fully explain the exact biologic mechanism leading to preeclampsia. There is some evidence supporting that genetic predisposition to hypertension may be associated with preeclampsia in European ancestry and Asian women<sup>14</sup>. Several polymorphisms associated with preeclampsia have also been previously associated with blood pressure traits, though the shared genetic risk has yet to be fully evaluated<sup>14–16</sup>. Recent studies have further sought to investigate the connection between preeclampsia or hypertensive disorders of pregnancy, blood pressure, and hypertension outside of pregnancy<sup>16–22</sup>. Though they uncover more of the genetic background of preeclampsia and its relationship to blood pressure traits, these studies utilized populations that were primarily of European descent, potentially limiting their generalizability to other non-European populations.

As preeclampsia is a blood pressure disorder and puts individuals at greater risk of hypertension later in life, we investigated whether collective blood pressure-raising alleles may contribute to preeclampsia in Black and White females. To increase the understanding of preeclampsia etiology and identify the potential shared genetic architecture with other blood pressure traits, we evaluated the association between polygenic scores (PGS) for three blood pressure traits—diastolic blood pressure (DBP), systolic blood pressure (SBP), and pulse pressure (PP)—and preeclampsia in several cohorts of Black and White reproductive-aged females with documented pregnancies. We also investigated the utility of blood pressure PGSs for the prediction of preeclampsia, determining whether the addition of genetic components improve the performance of prediction models composed of two major clinical risk factors: BMI and age.

## Results

### Population characteristics

In total, 8513 individuals were included in the study; 2979 individuals from BioVU, 3279 from the Electronic Medical Record and Genomics Network (eMERGE) cohort, and 2255 from the Penn Medicine Biobank (PMBB) cohort (Table 1). There were 1319 cases of preeclampsia, with 17%, 12%, 18% in BioVU, eMERGE, and PMBB. As identified through their electronic health records (EHRs), 41.16% were Black (38.47% in BioVU, 34.74% in eMERGE, 54.06% in PMBB cohort) and 58.84% were White (61.53% in BioVU, 65.26% in eMERGE, and 45.94% in PMBB) in the overall cohort. The overall rate of preeclampsia was 15.49%. Higher percentages of preeclampsia were found in Black individuals compared to White individuals (18.78% vs 13.20%). Within cohorts and across the entire population, cases had higher average BMIs and ages compared to controls.

	BioVU		eMERGE		PMBB		Total	
	Cases (n = 509)	Controls (n = 2470)	Cases (n = 398)	Controls (n = 2881)	Cases (n = 412)	Controls (n = 1843)	Cases (n = 1319)	Controls (n = 7194)
Race								
Non-Hispanic Black	212 (41.65%)	934 (37.81%)	178 (44.72%)	961 (33.36%)	268 (65.05%)	951 (51.60%)	658 (49.89%)	2846 (39.56%)
Non-Hispanic White	297 (58.35%)	1536 (62.19%)	220 (55.28%)	1920 (66.64%)	144 (34.95%)	892 (48.40%)	661 (50.11%)	4348 (60.44%)
BMI, kg/m <sup>2</sup> , mean (sd)	33.3 (8.81)	29.2 (7.49)	31.4 (8.19)	28.7 (8.21)	33.7 (10.08)	29.8 (8.47)	32.9 (9.05)	29.2 (8.04)
Age, years, mean (sd)	30.7 (6.53)	29.7 (6.73)	36.6 (5.53)	36.0 (6.22)	36.4 (6.87)	34.9 (6.37)	34.3 (6.36)	33.6 (6.44)

**Table 1.** Population characteristics. *BMI* body mass index, *sd* standard deviation.

### Blood pressure trait PGSs-preeclampsia associations

In Black individuals, PGSs for DBP and SBP were not significantly associated with preeclampsia in the meta-analyses (odds ratio (OR)<sub>DBP</sub> = 0.99, 95% confidence interval (CI) 0.88–1.12, *p*-value = 0.90; OR<sub>SBP</sub> = 1.10, 95% CI 0.99–1.23, *p*-value = 0.09) (Table 2). The PGS for PP was significantly associated with preeclampsia in Black individuals, with every one standard deviation increase in the PP PGS, the odds of preeclampsia is associated with a 1.16-fold increase (OR<sub>PP</sub> = 1.16, 95% CI 1.03–1.31, *p*-value = 0.01). In White individuals, DBP, SBP, and PP PGSs were all significantly associated with preeclampsia (OR<sub>DBP</sub> = 1.14, 95% CI 1.07–1.23, *p*-value =  $2.25 \times 10^{-4}$ ; OR<sub>SBP</sub> = 1.18, 95% CI 1.10–1.27, *p*-value =  $4.42 \times 10^{-6}$ ; OR<sub>PP</sub> = 1.14, 95% CI 1.05–1.23, *p*-value =  $1.77 \times 10^{-3}$ ). In the cross-ancestry meta-analysis, all three PGSs were highly significant (Table 3; Figs. 1, 2, 3).

### Prediction models

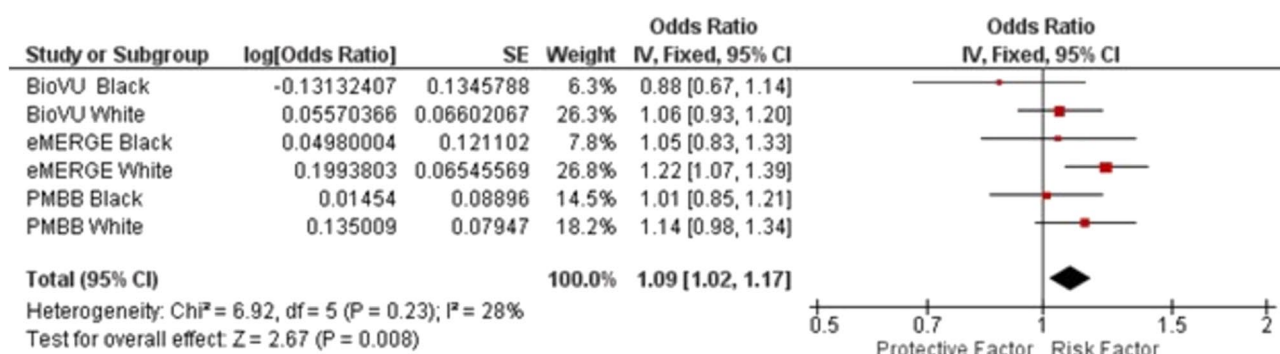
Models composed of solely of all three PGSs and 10 principal components had extremely poor discriminatory abilities, with area under the receiver operating characteristics curve (AUCs) ranging from 0.49 to 0.53 (Table 4). Clinical models had slightly better discrimination (AUCs: 0.59–0.67). The addition of the PGSs to the clinical models decreased the predictive ability of the models. Brier scores for PGS, clinical, and full models demonstrate low accuracy in the testing population. In general, lower AUCs and higher Brier scores were seen in PGS models compared to models composed of clinical risk factors and in EHR-identified non-Hispanic Black populations compared to White or cross-ancestry populations. Overall, inclusion of PGSs in a predictive model composed of significant clinical risk factors did not provide an improvement in predictive performance. Calibration and discrimination of the predictive models indicated poor fits in the testing population (Table 4).

	Black		White	
	OR (95% CI)	P-value	OR (95% CI)	P-value
DBP	0.99 (0.88, 1.12)	0.89	1.14 (1.05, 1.23)	<b>1.21 × 10<sup>-3</sup></b>
SBP	1.10 (0.98, 1.22)	0.11	1.18 (1.10, 1.27)	<b>4.24 × 10<sup>-6</sup></b>
PP	1.16 (1.03, 1.30)	<b>0.01</b>	1.13 (1.05, 1.23)	<b>2.34 × 10<sup>-3</sup></b>

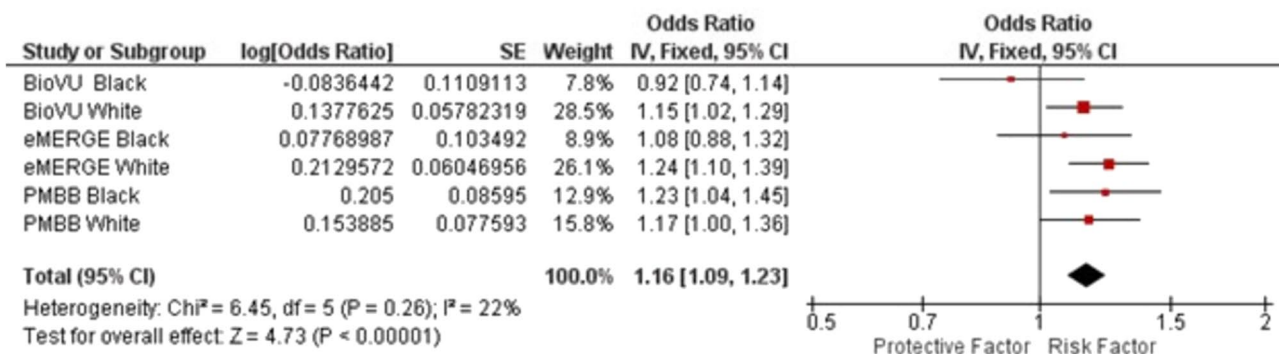
**Table 2.** Association between blood pressure PGSs and preeclampsia by self-reported race. OR odds ratio, CI confidence interval, DBP diastolic blood pressure, SBP systolic blood pressure, PP pulse pressure. Significant values are in bold.

	OR (95% CI)	P-value
DBP	1.10 (1.02, 1.17)	<b>7.68 × 10<sup>-3</sup></b>
SBP	1.16 (1.09, 1.23)	<b>2.23 × 10<sup>-6</sup></b>
PP	1.14 (1.07, 1.27)	<b>9.86 × 10<sup>-5</sup></b>

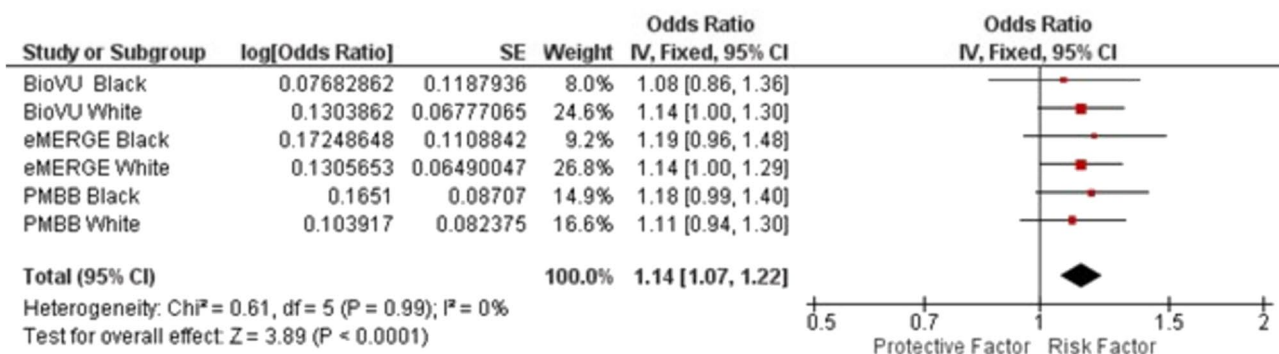
**Table 3.** Association between blood pressure PGSs and preeclampsia (cross-ancestry). OR odds ratio, CI confidence interval, DBP diastolic blood pressure, SBP systolic blood pressure, PP pulse pressure. Significant values are in bold.



**Figure 1.** Cross-ancestry meta-analysis of the association of diastolic blood pressure polygenic risk score and preeclampsia.



**Figure 2.** Cross-ancestry meta-analysis of the association of systolic blood pressure polygenic risk score and preeclampsia.



**Figure 3.** Cross-ancestry meta-analysis of the association of pulse pressure polygenic risk score and preeclampsia.

	Clinical model		DBP		SBP		PP		PGS Model		Full Model	
	AUC (95% CI)	Brier Score	AUC (95% CI)	Brier Score	AUC (95% CI)	Brier Score	AUC (95% CI)	Brier Score	AUC (95% CI)	Brier Score	AUC (95% CI)	Brier Score
Black	0.59 (0.54–0.63)	0.32	0.48 (0.43, 0.52)	0.62	0.48 (0.43, 0.52)	0.55	0.48 (0.43, 0.52)	0.64	0.52 (0.47–0.56)	0.64	0.53 (0.49–0.57)	0.68
White	0.67 (0.64–0.71)	0.30	0.51 (0.48, 0.55)	0.09	0.53 (0.49, 0.56)	0.31	0.53 (0.49, 0.57)	0.19	0.53 (0.49–0.56)	0.24	0.66 (0.62–0.69)	0.52
All individuals	0.64 (0.62–0.67)	0.29	0.51 (0.48, 0.54)	0.29	0.52 (0.49, 0.55)	0.29	0.51 (0.48, 0.54)	0.28	0.51 (0.48–0.54)	0.30	0.57 (0.54–0.60)	0.37

**Table 4.** Prediction models in testing (BioVU) population.

### Discussion

We evaluated the potential overlap between genetic risk factors for blood pressure traits and preeclampsia. Utilizing PGSs for DBP, SBP, and PP, we found significant associations between polygenic risk for these blood pressure traits and preeclampsia. These findings support shared genetic architecture across preeclampsia risk and blood pressure traits. Overall, these findings suggest genetic factors contributing to blood pressure regulation in the general population also predispose to preeclampsia. However, polygenic risk scores for blood pressure traits had poor predictive ability in our study population. Furthermore, the addition of PGSs to a risk score composed of known clinical risk factors for preeclampsia did not result in a significant improvement in the prediction of preeclampsia. The poor predictive ability of the PRSs should be viewed with caution, as the analyses were likely influenced by small sample size and limited power.

Pre-pregnancy hypertension is a major risk factor for preeclampsia<sup>8</sup>. The development of preeclampsia also increases the risk of persistent hypertension outside of pregnancy<sup>7</sup>. This overlapping risk is likely due in part to shared underlying genetic mechanisms. Several preeclampsia GWAS support this idea, finding polymorphisms previously associated with hypertension are also associated with preeclampsia<sup>14,15</sup>. A recent large-scale meta-analysis of European and Central Asian women further reinforced the concept that genetic predisposition to hypertension contributes to the risk of hypertensive disorders of pregnancy<sup>14</sup>. While an association between a hypertension PGS and hypertensive disorders in pregnancy were observed, the effect was stronger for gestational hypertension compared to preeclampsia, suggesting that, while there is overlap in genetic susceptibility, additional factors are involved in preeclampsia.

It is unclear whether blood pressure complications and preeclampsia are comorbid conditions or if complications lead to preeclampsia or vice versa. Maternal cardiovascular systems undergo significant changes during pregnancy<sup>23,24</sup>. Individuals who experience preeclampsia may undergo further physiological changes. For example, previous research has noted significant increases in arterial stiffness in those with preeclampsia compared to individuals with normotensive pregnancy or those with gestational hypertension<sup>25</sup>. Our study provides further evidence of the role of arterial stiffness in preeclampsia, given that the PGS for PP, a marker of arterial stiffness, was strongly associated with preeclampsia in both Blacks and Whites. Additional pathways involved in preeclampsia could also be uniquely triggered by pregnancy-related changes and the fetal genome. Blood pressure regulation through genetic mechanisms may also be perturbed during pregnancy, with interactions resulting in those with increased genetic risk becoming more likely to develop preeclampsia. Additional research is needed to disentangle the causal origins of shared risk factors between blood pressure traits, complications, and preeclampsia and to identify variants that uniquely contribute to preeclampsia risk. Our cohort is also significantly younger than most cohorts used in blood pressure research. Age-related changes in blood pressure regulation have also been reported previously and could account for some of the variation seen between genetic risk factors for blood pressure traits and preeclampsia<sup>26</sup>.

This study builds upon previous research by investigating the traits underlying hypertension and further disentangling how genetic risk factors for these traits impact preeclampsia risk. Two recent large scale GWAS have been conducted for preeclampsia and hypertensive disorders of pregnancy. Honigberg et al. used 20,064 cases of preeclampsia and 703,117 control individuals to perform a GWAS. Results from the GWAS were then used to create a PRS for preeclampsia<sup>16</sup>. They also used the Giri et al. study<sup>27</sup> to derive a PRS for SBP; however, they use different methods to create the score. The result was a score contain 1,064,898 HapMap3 variants compared to our score using 4294 single nucleotide polymorphisms (SNPs). The conclusion of the association analysis was the same as our study: preeclampsia is associated with SBP PGS. Though our study and Honigberg et al. both utilized PGS constructed from summary statistics from Giri et al. and found associations, the populations that were used to train and test these models are important considerations. Honigberg et al. trained their PRS on the UK Biobank European LD panel and validated the models in two other cohorts that were either completely or predominantly (>73.0%) of European descent<sup>16</sup>. We utilized eMERGE and BioVU populations that had a large proportion of non-Hispanic Black individuals.

Previous research has demonstrated that predictive performance of European ancestry derived PGS is lower in non-European ancestry samples<sup>28</sup>. Thus, the differences in methods, training, and testing populations make further comparison of results across studies difficult. Additional studies have also reported associations between blood pressure PGS (SBP and DBP) and preeclampsia<sup>17–19</sup>. Research utilizing blood pressure PGS to predict hypertensive disorders of pregnancy is conflicting. Several studies report improvement in predictive ability when PGSs were added to a model of clinical risk factors<sup>16,21</sup>, while others report no significant improvement in performance<sup>19</sup>. However, clinical risk factors and population characteristics used in prediction modeling were inconsistent across studies. Utilization of PGS for predicting preeclampsia warrants further investigation.

Our results, which support the conclusion that genetic risk factors for blood pressure have a small contribution to the risk of preeclampsia, have several key strengths. First, previous genetic studies of pregnancy-specific phenotypes, including preeclampsia, have included all females from the source population, regardless of pregnancy status. Results of these studies could be biased as a proportion of their controls do not have the ability to be classified as cases. Our study cohort was also restricted to females with a pregnancy, ensuring all individuals possessed the appropriate window during which the outcome could have occurred. Previous research, including two recent studies which found associations between hypertensive disorders of pregnancy and polygenic risk for gestational hypertension/preeclampsia or blood pressure traits<sup>21,22</sup>, has been largely limited to Asian and White women or those of European ancestry. Here, we investigate preeclampsia genetic risk factors in EHR-identified Black individuals, a population known to be at increased risk of preeclampsia<sup>11</sup>.

Our study has several considerations. The overall rate of preeclampsia in our study population was higher than expected based on previous observational data. The biobanks in our study all contain data from hospital systems, as opposed to population-based samples used to obtain estimates of incidence of preeclampsia. The populations served by these hospitals may be different from the general population. For example, Vanderbilt University Medical Center (VUMC) has a high-risk obstetric team which can provide advanced care. Individuals with preeclampsia may have been sent to VUMC by other health systems that are unable to provide care for these complex cases. The racial diversity in our study cohort could also contribute to the higher rates of preeclampsia. This study included a large proportion of Black individuals. Epidemiologic research has revealed that individuals who identify as Black or African American are at increased risk of preeclampsia compared to those from other racial and ethnic groups. While our sample size is small, results from our association analyses align with previous research which has found an association between polygenic risk scores for blood pressure traits and both gestational hypertension and preeclampsia<sup>21</sup>.

Our study found small differences in the associations of blood pressure PGSs and preeclampsia between the EHR-identified races, which cannot be corroborated by other research as previous studies have limited diversity or did not validate the PGS associations across other, non-European populations. Our stratified results should be interpreted carefully, especially in the non-Hispanic Black group which had small sample sizes and displayed a larger degree of uncertainty in effect estimates. Additionally, the less significance associations and poor transferability of the PGS in the non-Hispanic Black group may be related to the populations used to create them. Each PGS was created using summary statistics from a cross-ancestry GWAS and the DBP and SBP PGS were subsequently validated on a cross-ancestry population. However, both populations were predominately those of European ancestry or non-Hispanic White individuals. The differences in the ancestry of the populations used to discover significant loci, those used to validate the PGS, and our study cohort could have biased the results and contributed to the lack of significance in the non-Hispanic Black populations. It is also notable that the PGS

performed better in White compared to Black populations generally<sup>27</sup>, which may also impair the quality of their prediction of preeclampsia. To ensure health equity and transferability of PGS between populations of different ancestries, future genetic research of preeclampsia should aim to increase diversity of their cohorts.

Prediction modeling using the PGS models was also limited due to many individuals in the study cohort missing BMI and being excluded during predictive model development and testing. Prediction models composed of all three PGS had poor predictive ability. Adding the PGS to the clinical model resulted in poorer performance than the clinical model alone. This could be due to a portion of individuals with higher genetic burden being classified as controls, as the study cohort did not exclude those with hypertension diagnoses and the timing of hypertension development and diagnoses relative to pregnancy could not be fully determined without prospective blood pressure measurement. A previous study found that incorporating an SBP PGS into a clinical model resulted in a small improvement in the risk prediction and discrimination of preeclampsia<sup>21</sup>. However, the study also included a preeclampsia PGS and additional variables (e.g., in-vitro fertilization, renal insufficiency, multifetal pregnancy, etc.). Results from our prediction modeling should also be considered cautiously as clinical and environmental risk factors, including family history, were incomplete in our datasets. The addition of other known risk factors in development and testing of prediction models in larger datasets with more complete BMI information may increase their predictive ability and clinical value.

While this study further endorses the connection between genetic risk factors for blood pressure and preeclampsia, results suggest that many variants, each with small effect sizes, likely affect genetic predisposition to preeclampsia. Though many of these variants may mainly impact preeclampsia through regulation of blood pressure, other molecular pathways are likely involved. Future research should continue to develop and validate genome informed clinical models for preeclampsia across global populations and conduct additional analyses of preeclampsia shared genetic architecture with other conditions. These analyses will expand our understanding of preeclampsia etiology and potentially inform future treatments by providing evidence for repurposing treatments for conditions that share a biological basis with preeclampsia.

## Methods

### Study populations

Our study populations were obtained from three sources: BioVU, the Electronic Medical Record and Genomics Network (eMERGE), and the Penn Medicine Biobank (PMBB) at the University of Pennsylvania<sup>29–31</sup>. BioVU is Vanderbilt University Medical Center's biorepository linking DNA samples to deidentified electronic health records (EHR). eMERGE, a national network which also combines DNA biorepositories with EHR systems, contains data from individuals at more than 10 sites around the U.S. Though Vanderbilt is an eMERGE site, data from eMERGE is independent of BioVU. Like BioVU, the PMBB connects de-identified EHR data to a biobank from a single academic healthcare system. The data and source populations from BioVU, eMERGE, and PMBB are similar. All individuals were recruited at their corresponding healthcare system, largely in an outpatient or hospital setting, and have linked genotype and de-identified EHR data. Each source contains a subset of the active patients within their system. Most individuals are of European ancestry or identify as non-Hispanic White, though a sizable proportion identify as African American or Black, Hispanic, and Asian. Mirroring the U.S. population, a slightly higher proportion of individuals in BioVU, eMERGE, and the PMBB are female<sup>29–31</sup>.

This study was deemed non-human subjects research and approved by Vanderbilt University Medical Center Institutional Review Board. All methods were carried out in accordance with relevant guidelines and regulations. Individuals were eligible for inclusion in the study if they were females who had documented pregnancies or deliveries, were of reproductive age at last known EHR entry (18 to 45 years old), identified in the EHR as non-Hispanic Black or non-Hispanic White and had genotype information available. Identification of individuals who are or were pregnant and/or had deliveries was accomplished using International Classification of Disease, 9<sup>th</sup> and 10<sup>th</sup> revisions (ICD-9/ICD-10) codes. Presence of any of the following codes indicated potential eligibility for inclusion in the study: 631, 633, 634.3, 635 through 679, 760–779, 796.5, V22–V24, V27, V28, V72.42, V82.4, V89, V91 (ICD-9), O09–O16, O20–O26, O28–O36, O40–O48, O60–O77, O80, O82, O85–O92, O94, O98, O99, O9A, Z03.7, Z32.01, Z33, Z34, Z36, Z37, Z39, or Z3A (ICD-10). Preeclampsia cases were defined using the presence of any one of codes 642.4, 642.5, 642.6, 642.7 (ICD-9 codes), O11, O14, and O15 (ICD-10).

### Genotyping and quality control

Samples from BioVU were genotyped using the custom Multi-ethnic Genotyping Array (MEGA-Ex, Illumina) and imputed to the Haplotype Reference Consortium (HRC) panel v1.1 using the Michigan Imputation Server (MIS)<sup>30,32,33</sup>. As described previously, individuals from all eMERGE sites were genotyped via Illumina 660WQuad array for those reported as European or unknown ancestry and the Illumina 1M-Duo array for those of African ancestry and were similarly imputed to the HRC genotype reference panel using the MIS<sup>34</sup>. Standard quality control procedures were performed on cohorts from BioVU and eMERGE, including filtering based on sample quality and composition (i.e., checking for sex inconsistencies, sample relatedness, and sample genotyping call rate) and marker quality (i.e., marker genotype call rate, concordance, minor allele frequency, and Hardy-Weinberg Equilibrium)<sup>29,30</sup>. The eMERGE cohort included additional quality control procedures, such as strand orientation analysis, before genotype data from the contributing sites were merged<sup>34</sup>. PMBB data was genotyped on an Illumina Global Screening Array v.1 and v.2.0 (GSAv1, GSAv2). Sample-level quality control was performed before imputation was performed using Minimac5 version 2.0.0 software on the TOPMed Imputation Server. As with the cohorts from BioVU and eMERGE, post-imputation quality control was performed and included imputation score filter ( $R^2 > 0.7$ ), removal of palindromic variants, biallelic variant check, genotype call rate ( $> 99\%$ ), minor allele frequency ( $> 1\%$ ), sex check, sample call rate ( $> 99\%$ ) filtering, and a Hardy-Weinberg equilibrium test ( $p\text{-value} > 1.0 \times 10^{-8}$ )<sup>31</sup>.

## Blood pressure traits polygenic scores (PGS) development

We evaluated the association between PGS for three blood pressure traits (DBP, SBP, and PP) and preeclampsia. Individually, DBP, SBP, and PP are highly heritable and common methods to capture blood pressure in genetic studies<sup>35</sup>. We evaluated these traits independently because the genetic and biologic mechanisms underlying these traits vary. They capture different domains of blood pressure: SBP is the maximal arterial pressure exerted when the heart beats while DBP is the (arterial) pressure between heartbeats, and PP is the difference between the two and indicative of arterial stiffness<sup>27,35</sup>. We utilized previously developed PGS for these three blood pressure traits<sup>36</sup>. Briefly, summary statistics for genetic associations with DBP, SBP, and PP were obtained from a previously published cross-ancestry genome-wide association study ( $n_{\max} = 760,226$  subjects)<sup>27</sup>. In this large meta-analysis for blood pressure phenotypes, the measurements for SBP, DBP, and PP were obtained from electronic health records. For the discovery cohort, containing individuals from the Million Veterans Program, these phenotypes were defined as the median value of all eligible measurements of SBP and corresponding DBP. To be eligible for inclusion in calculation of the median and final blood pressure phenotypes, these measurements could not come from Emergency Department outpatient visits and could not have occurred at or after an ICD-9 code for chronic kidney disease (code group 585), secondary hypertension (code group 405), or heart failure (code group 428). For those with available pain scores, blood pressure measurements taken during encounters where individuals had a pain score greater than or equal to 5 were also ruled ineligible and taken out of median calculation. Individuals on an antihypertensive medication, they added 15 mmHg and 10 mmHg to SBP and DBP, respectively. Finally, if median values were documented multiple times on distinct dates in an individual's electronic health record, the earliest measurement was used to identify DBP, antihypertensive treatment status, and age of the individual. In the UK Biobank cohort, SBP, DBP, and PP phenotypes were averaged over two measures and similarly adjusted for antihypertensive drug status<sup>27</sup>.

The PGSs for DBP, SBP, and PP were developed using these summary statistics and the resulting SBP and DBP PGSs were validated in BioVU ( $n_{\max} = 37,132$ , DBP PGS  $p = 7.63 \times 10^{-113}$ , SBP PGS  $p = 4.18 \times 10^{-132}$ ) for association with blood pressure measurements, adjusting for age, sex, BMI, and the top ten principal components of ancestry<sup>36</sup>. The BioVU genetic data was pruned for linkage disequilibrium at an  $r^2$  threshold of 0.1 with a maximum distance of 250 kilobases from associated SNPs ( $p < 1 \times 10^{-5}$ ) in the blood pressure summary statistics. Weighted scores were calculated in PLINK using p-value thresholding ( $p \leq 0.00001$ )<sup>37</sup>. The DBP, SBP, and PP scores contained 4194, 4294, and 3345 SNPs respectively (Supplementary Tables S1–S3)<sup>36</sup>. PGS were transformed by multiplying scores by the standard deviation of the PGS in the original population from which they were optimized.

## Blood pressure trait PGSs-preeclampsia association analyses

Logistic regression was used to evaluate the association between each of the three blood pressure traits and preeclampsia separately in the three cohorts (BioVU, eMERGE, and PMBB) stratified by race (Black and White). Analyses were adjusted for age, age squared, and the top ten principal components. Associations were considered statistically significant at  $p$ -values less than 0.05. Odds ratios (OR) and 95% confidence intervals (CI) are reported for each model as the increase in odds of preeclampsia per one standard deviation increase in each blood pressure polygenic risk score. Analyses were completed using R version 4.2.2<sup>38</sup>. To investigate whether trends were similar across different populations, we combined effect estimates and standard errors using fixed-effects meta-analyses for each EHR-identified race. Then, we meta-analyzed data overall to obtain a single value across the entire study for each of the three blood pressure PGSs. Fixed effects meta-analyses were performed in RevMan, version 5.4, using an inverse-variance approach<sup>39</sup>.

## Predictive modeling

To evaluate the performance of blood pressure PGS for the prediction of preeclampsia, we developed prediction models using tenfold cross validation. Joint prediction models composed of all three PGSs (DBP, SBP, and PP) were developed. Clinical models, composed of two major preeclampsia risk factors (BMI and age at last EHR entry), were constructed for comparison. A full model, which added the three PGSs to the clinical model, was used to investigate whether PGS improved prediction of preeclampsia beyond traditional risk factors. Models that included PGSs also adjusted for the top ten principal components.

Models were developed in the cohort from eMERGE and tested in BioVU. Synthetic minority over-sampling techniques (SMOTE) were used to attenuate the effects of class imbalance (number of cases and controls) due to the rarity of the outcome during model training<sup>30,40</sup>. Only those with complete covariate information were used in prediction modeling. All models were created and evaluated separately for each EHR-identified race (non-Hispanic Blacks and Whites) with subsequent development of cross-ancestry models. Model performance was assessed through discrimination and calibration using the area under the receiver operating characteristics curve (AUC) and the Brier Score. All prediction modeling was performed in R<sup>38</sup>.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. Data from the Penn Medicine Biobank is available from Dr. Verma (Shefali.SetiaVerma@pennmedicine.upenn.edu) upon reasonable request.

Received: 8 November 2023; Accepted: 24 July 2024

Published online: 30 July 2024

## References

1. Gestational Hypertension and Preeclampsia. ACOG Practice Bulletin, Number 222. *Obstet. Gynecol.* **135**(6), e237–e260 (2020).
2. Wallis, A. B. *et al.* Secular trends in the rates of preeclampsia, eclampsia, and gestational hypertension, United States, 1987–2004. *Am. J. Hypertens* **21**(5), 521–526 (2008).
3. Ananth, C. V., Keyes, K. M. & Wapner, R. J. Pre-eclampsia rates in the United States, 1980–2010: age-period-cohort analysis. *BMJ* **347**, f6564 (2013).
4. Stevens, W., *et al.* Short-term costs of preeclampsia to the United States health care system. *Am. J. Obstet. Gynecol.* **217**(3), 237–248 e16 (2017).
5. Goffin, S. M. *et al.* Maternal pre-eclampsia and long-term offspring health: Is there a shadow cast?. *Pregn. Hypertens* **12**, 11–15 (2018).
6. Staff, A. C. & Redman, C. W. G. The difference between early- and late-onset pre-eclampsia. In *Preeclampsia: Basic, genomic, and clinical* (ed. Saito, S.) 157–172 (Springer, 2018).
7. Bokslag, A. *et al.* Preeclampsia; short and long-term consequences for mother and neonate. *Early Hum. Dev.* **102**, 47–50 (2016).
8. Bartsch, E. *et al.* Clinical risk factors for pre-eclampsia determined in early pregnancy: Systematic review and meta-analysis of large cohort studies. *BMJ* **353**, i1753 (2016).
9. Conde-Agudelo, A. & Belizan, J. M. Risk factors for pre-eclampsia in a large cohort of Latin American and Caribbean women. *BJOG* **107**(1), 75–83 (2000).
10. Duckitt, K. & Harrington, D. Risk factors for pre-eclampsia at antenatal booking: Systematic review of controlled studies. *BMJ* **330**(7491), 565 (2005).
11. Shiozaki, A. & Saito, S. Risk factors for preeclampsia. In *Preeclampsia: basic, genomic, and clinical* (ed. Saito, S.) 3–25 (Springer, 2018).
12. Gray, K. J., Saxena, R. & Karumanchi, S. A. Genetic predisposition to preeclampsia is conferred by fetal DNA variants near FLT1, a gene involved in the regulation of angiogenesis. *Am. J. Obstet. Gynecol.* **218**(2), 211–218 (2018).
13. Johnson, M. P. *et al.* Genome-wide association scan identifies a risk locus for preeclampsia on 2q14, near the inhibin, beta B gene. *PLoS One* **7**(3), e33666 (2012).
14. Steinthorsdottir, V. *et al.* Genetic predisposition to hypertension is associated with preeclampsia in European and Central Asian women. *Nat. Commun.* **11**(1), 5976 (2020).
15. Zhao, L., Bracken, M. B. & DeWan, A. T. Genome-wide association study of pre-eclampsia detects novel maternal single nucleotide polymorphisms and copy-number variants in subsets of the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study cohort. *Ann. Hum. Genet.* **77**(4), 277–287 (2013).
16. Honigberg, M. C. *et al.* Polygenic prediction of preeclampsia and gestational hypertension. *Nat. Med.* **29**(6), 1540–1549 (2023).
17. Kivioja, A. *et al.* Increased risk of preeclampsia in women with a genetic predisposition to elevated blood pressure. *Hypertension* **79**(9), 2008–2015 (2022).
18. Gray, K. J. *et al.* Risk of pre-eclampsia in patients with a maternal genetic predisposition to common medical conditions: A case-control study. *BJOG* **128**(1), 55–65 (2021).
19. Kovacheva, V. P. *et al.* preeclampsia prediction using machine learning and polygenic risk scores from clinical and genetic risk factors in early and late pregnancies. *Hypertension* **81**(2), 264–272 (2024).
20. Tyrmi, J. S. *et al.* Genetic risk factors associated with preeclampsia and hypertensive disorders of pregnancy. *JAMA Cardiol.* **8**(7), 674–683 (2023).
21. Nurkkala, J. *et al.* Associations of polygenic risk scores for preeclampsia and blood pressure with hypertensive disorders of pregnancy. *J. Hypertens* **41**(3), 380–387 (2023).
22. Perisic, M. M. *et al.* Polygenic risk score and risk factors for preeclampsia and gestational hypertension. *J. Pers. Med.* **12**(11), 1 (2022).
23. Duvekot, J. J. & Peeters, L. L. Maternal cardiovascular hemodynamic adaptation to pregnancy. *Obstet. Gynecol. Surv.* **49**(12 Suppl), S1–14 (1994).
24. Shen, M. *et al.* Trajectory of blood pressure change during pregnancy and the role of pre-gravid blood pressure: A functional data analysis approach. *Sci. Rep.* **7**(1), 6227 (2017).
25. Hausvater, A. *et al.* The association between preeclampsia and arterial stiffness. *J. Hypertens.* **30**(1), 17–33 (2012).
26. Simino, J. *et al.* Gene-age interactions in blood pressure regulation: A large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia. *Am. J. Hum. Genet.* **95**(1), 24–38 (2014).
27. Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **51**(1), 51–62 (2019).
28. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**(1), 3328 (2019).
29. Gottesman, O. *et al.* The electronic medical records and genomics (eMERGE) network: Past, present, and future. *Genet. Med.* **15**(10), 761–771 (2013).
30. Roden, D. M. *et al.* Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharmacol. Ther.* **84**(3), 362–369 (2008).
31. Verma, A. *et al.* The Penn Medicine BioBank: Towards a genomics-enabled learning healthcare system to accelerate precision medicine in a diverse population. *J. Pers. Med.* **12**(12), 1 (2022).
32. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**(10), 1284–1287 (2016).
33. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**(10), 1279–1283 (2016).
34. Zuvich, R. L. *et al.* Pitfalls of merging GWAS data: Lessons learned in the eMERGE network and quality control procedures to maintain high data quality. *Genet. Epidemiol.* **35**(8), 887–898 (2011).
35. Keaton, J. M. *et al.* Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nat. Genet.* **56**(5), 778–791 (2024).
36. Breeyear, J. H. *et al.* Diastolic blood pressure alleles improve congenital heart defect repair outcomes. *Circ. Res.* **130**(7), 1030–1037 (2022).
37. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**(3), 559–575 (2007).
38. R Core Team, *R: A language and environment for statistical computing*. 2019, R Foundation for Statistical Computing: Vienna, Austria.
39. The Cochrane Collaboration, *Review Manager (RevMan)* (2020).
40. Chawla, N. V. *et al.* SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).

## Acknowledgements

Drs. Elizabeth A. Jasper and Jacklyn N. Hellwege are supported by the NIH Building Interdisciplinary Research Career's in Women's Health career development program (K12HD043483 PIs: K.E. Hartmann, A.S. Major, and D.R. Velez Edwards). Dr. Jasper was also supported by a National Human Genome Research Institute award for the Vanderbilt Genomic Medicine Training Program (T32HG008341 PIs: J.F. Peterson, N. Cox, and

D.M. Roden). Joseph H. Breeyear was supported by Vanderbilt University Medical Center's Clinical and Translational Award training program (TL1-TR002244 PI: K.E. Hartmann). Iftikhar J. Kullo is funded by grants HG006379 and HG011710 from the National Human Genome Research Institute and K24HL137010 from the National Heart Lung and Blood Institute. BioVU: The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH funded Shared Instrumentation Grant S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at <https://vict.vumc.org/biovu-funding/>. eMERGE Network (Phase III): This phase of the eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG008657 (Group Health Cooperative/University of Washington); U01HG008685 (Brigham and Women's Hospital); U01HG008672 (Vanderbilt University Medical Center); U01HG008666 (Cincinnati Children's Hospital Medical Center); U01HG006379 (Mayo Clinic); U01HG008679 (Geisinger Clinic); U01HG008680 (Columbia University Health Sciences); U01HG008684 (Children's Hospital of Philadelphia); U01HG008673 (Northwestern University); U01HG008701 (Vanderbilt University Medical Center serving as the Coordinating Center); U01HG008676 (Partners Healthcare/Broad Institute); U01HG008664 (Baylor College of Medicine); and U54MD007593 (Meharry Medical College). eMERGE Network (Phase IV): This phase of the eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG011172 (Cincinnati Children's Hospital Medical Center); U01HG011175 (Children's Hospital of Philadelphia); U01HG008680 (Columbia University); U01HG011176 (Icahn School of Medicine at Mount Sinai); U01HG008685 (Mass General Brigham); U01HG006379 (Mayo Clinic); U01HG011169 (Northwestern University); U01HG011167 (University of Alabama at Birmingham); U01HG008657 (University of Washington Medical Center); U01HG011181 (Vanderbilt University Medical Center); U01HG011166 (Vanderbilt University Medical Center serving as the Coordinating Center).

### Author contributions

E.A.J. conceptualized and designed the study, performed analyses, interpreted the results, and prepared the initial article, and reviewed and revised the manuscript. J.N.H. and J.H.B. contributed to the design of the study, analyses, interpretation, and preparation of the manuscript. B.X. performed analyses and revised the manuscript. G.P.J., I.B.S., K.A.L., G.C., M.G.H., O.D., I.J.K., and I.A.H. reviewed and substantially revised the manuscript. S.S.V. aided in data acquisition and revised the manuscript. T.L.E. and D.R.V.E. conceptualized and designed the study, coordinated, and supervised data acquisition and storage, made substantial contributions to the interpretation of results, and critically revised the manuscript. All authors approved the final article as submitted and agree to be accountable for all aspects of the work.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68469-6>.

**Correspondence** and requests for materials should be addressed to D.R.V.E.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024