Clotting factor genes are associated with preeclampsia in high-altitude pregnant women in the Peruvian Andes

Authors

Maria A. Nieves-Colón, Keyla M. Badillo Rivera, Karla Sandoval, ..., Christopher R. Gignoux, Genevieve L. Wojcik, Andrés Moreno-Estrada

Correspondence

mnievesc@umn.edu (M.A.N.-C.),
andres.moreno@cinvestav.mx (A.M.-E.)

Nieves-Colón et al., 2022, The American Journal of Human Genetics *109*, 1117–1139
June 2, 2022
https://doi.org/10.1016/j.ajhg.2022.04.014

CellPress

Clotting factor genes are associated with preeclampsia in high-altitude pregnant women in the Peruvian Andes

Maria A. Nieves-Colón,[1,2,3,22,*] Keyla M. Badillo Rivera,[4,22] Karla Sandoval,[1] Vanessa Villanueva Dávalos,[5] Luis E. Enriquez Lencinas,[5] Javier Mendoza-Revilla,[6,7] Kaustubh Adhikari,[8,9] Ram González-Buenfil,[1] Jessica W. Chen,[4] Elisa T. Zhang,[4] Alexandra Sockell,[4] Patricia Ortiz-Tello,[4] Gloria Malena Hurtado,[6] Ramiro Condori Salas,[6] Ricardo Cebrecos,[6] José C. Manzaneda Choque,[10] Franz P. Manzaneda Choque,[10] Germán P. Yábar Pilco,[10] Erin Rawls,[2] Celeste Eng,[11] Scott Huntsman,[11] Esteban Burchard,[11] Andrés Ruiz-Linares,[9,12,13] Rolando González-José,[14] Gabriel Bedoya,[15,24] Francisco Rothhammer,[16,17] Maria Cátira Bortolini,[18] Giovanni Poletti,[6] Carla Gallo,[6] Carlos D. Bustamante,[4,19] Julie C. Baker,[4] Christopher R. Gignoux,[20,23] Genevieve L. Wojcik,[21,23] and Andrés Moreno-Estrada[1,23,*]

Summary

Preeclampsia is a multi-organ complication of pregnancy characterized by sudden hypertension and proteinuria that is among the leading causes of preterm delivery and maternal morbidity and mortality worldwide. The heterogeneity of preeclampsia poses a challenge for understanding its etiology and molecular basis. Intriguingly, risk for the condition increases in high-altitude regions such as the Peruvian Andes. To investigate the genetic basis of preeclampsia in a population living at high altitude, we characterized genome-wide variation in a cohort of preeclamptic and healthy Andean families (n = 883) from Puno, Peru, a city located above 3,800 meters of altitude. Our study collected genomic DNA and medical records from case-control trios and duos in local hospital settings. We generated genotype data for 439,314 SNPs, determined global ancestry patterns, and mapped associations between genetic variants and preeclampsia phenotypes. A transmission disequilibrium test (TDT) revealed variants near genes of biological importance for placental and blood vessel function. The top candidate region was found on chromosome 13 of the fetal genome and contains clotting factor genes *PROZ*, *F7*, and *F10*. These findings provide supporting evidence that common genetic variants within coagulation genes play an important role in preeclampsia. A selection scan revealed a potential adaptive signal around the *ADAM12* locus on chromosome 10, implicated in pregnancy disorders. Our discovery of an association in a functional pathway relevant to pregnancy physiology in an understudied population of Native American origin demonstrates the increased power of family-based study design and underscores the importance of conducting genetic research in diverse populations.

Introduction

Preeclampsia (MIM: 189800) is a hypertensive disorder of pregnancy that is a leading cause of morbidity and mortality for mothers and infants worldwide. The disorder complicates 5%–7% of global pregnancies, causes nearly 40% of all premature births, and is associated with 10%–15% of all maternal deaths.[1–3] This morbidity is even higher in developing countries and among communities with limited access to healthcare.[4] Despite posing a significant global disease burden, the heterogeneity of preeclampsia remains a major challenge for understanding its etiology

[1]Laboratorio Nacional de Genómica para la Biodiversidad (UGA-LANGEBIO), CINVESTAV, Irapuato, Guanajuato 36821, México; [2]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85281, USA; [3]Department of Anthropology, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA; [4]Department of Genetics, Stanford School of Medicine, Stanford, CA 94305, USA; [5]Departmento de Gineco-Obstetricia, Hospital Regional Manuel Nuñez Butrón, Puno 21002, Peru; [6]Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima 15102, Peru; [7]Human Evolutionary Genetics Unit, Institut Pasteur, UMR 2000, CNRS, Paris 75015, France; [8]School of Mathematics and Statistics, Faculty of Science, Technology, Engineering and Mathematics, The Open University, Milton Keynes MK7 6AA, UK; [9]Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, WC1E 6BT London, UK; [10]Universidad Nacional del Altiplano, Puno 21002, Peru; [11]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94143, USA; [12]Aix-Marseille Université, CNRS, EFS, ADES, 13005 Marseille, France; [13]Ministry of Education Key Laboratory of Contemporary Anthropology and Collaborative Innovation Center of Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Yangpu District, Shanghai, China; [14]Instituto Patagónico de Ciencias Sociales y Humanas, Centro Nacional Patagónico-CONICET y Programa Nacional de Referencia y Biobanco Genómico de la Población Argentina (PoblAr), Ministerio de Ciencia, Tecnología e Innovación, Puerto Madryn, Chubut, Argentina; [15]Genética Molecular (GENMOL), Universidad de Antioquía, Medellin, Colombia; [16]Instituto de Alta Investigación Universidad de Tarapacá, Tarapacá, Chile; [17]Programa de Genética Humana, ICBM Facultad de Medicina, Universidad de Chile, Santiago, Chile; [18]Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, Rio Grande do Sul, Brazil; [19]Department of Biomedical Data Science, Stanford School of Medicine, Stanford, CA 94305, USA; [20]University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA; [21]Department of Epidemiology, Bloomberg School of Public Health, John Hopkins University, Baltimore, MD 21205, USA
[22]These authors contributed equally
[23]These authors contributed equally
[24]Deceased
*Correspondence: mnievesc@umn.edu (M.A.N.-C.), andres.moreno@cinvestav.mx (A.M.-E.)
https://doi.org/10.1016/j.ajhg.2022.04.014.

and genetic basis.[3,5] Clinical and pathological research suggests a major role of the fetal placental organ in preeclampsia, where shallow invasion of fetal cells into the maternal endometrium results in insufficient remodeling of the maternal vasculature.[6,7] While it roots in early placental development, preeclampsia is usually not detected until the third trimester of pregnancy (>20 weeks of gestation), when it is identified by a sudden onset of hypertension and signs of organ damage, typically proteinuria (excess protein in the urine). The severity of preeclampsia is determined by gestational age at onset, as well as the magnitude of hypertension and organ damage.[8] The disorder is known to be heritable with multicomponent risk determined by maternal, fetal, and paternal factors.[3,5,9–11] Estimates suggest the overall heritability of preeclampsia is approximately 55%–60%, with maternal, fetal, and paternal contributions making up 30%–35%, 20%, and 13%, respectively.[5,12,13] Several lines of evidence, including pregnant mothers' changing risk of developing preeclampsia with different male partners[13,14] and the identification of polymorphisms in the paternal genome that impact preeclampsia risk,[15] highlight the importance of including family trios in the study of this condition (as reviewed in Galaviz-Hernandez et al.[16]). Other risk factors include family history,[17,18] socioeconomic status,[19] and chronic hypertension or diabetes.[2] Residence at high altitudes above 2,500 meters (m) above sea level also contributes considerably to the risk of developing preeclampsia.[20]

High altitude increases the risk of preeclampsia and other hypertensive pregnancy disorders at least 2- to 3-fold.[21] For example, Bolivian communities living at 3,500 m of altitude have an incidence of preeclampsia of up to 20%,[22] about three times higher than the world average.[23] In neighboring Peru, preeclampsia complicates up to 22% of all pregnancies and is the second leading cause of maternal deaths.[24,25] Previous research suggests that chronic hypoxia due to altitude is the major component underlying this increased risk.[26,27] For instance, retrospective cohort studies of high-altitude women living at 3,100 m in Colorado, USA suggest that chronic hypoxia exposure may lead to irregular adjustments of the maternal vascular architecture during pregnancy and improper placentation.[27,28] As discussed by Zamudio,[20] the potential mechanisms by which hypoxia may increase preeclampsia risk include reduced uteroplacental blood flow, placental oxidative stress, increased maternal vascular reactivity, genetic factors, and circulating angiogenic growth factors of placental origin. While changes in blood flow may have shown the strongest effect,[26] no single one of these mechanisms is solely responsible for the increased risk of developing preeclampsia at high altitudes. Instead, the strong impact of hypoxia on several physiological systems likely causes many factors to act together to increase individual preeclampsia risk.

Due to the high incidence of preeclampsia among high-altitude resident populations, highland pregnancy studies have been proposed as a natural setting to elucidate genetic factors involved in preeclampsia and other hypertensive pregnancy complications.[20,27–31] Native Andean populations are of particular interest for this research due to their unique physiological adaptations to chronic high-altitude hypoxia, such as enhanced pulmonary volumes and elevated blood hemoglobin concentrations.[32] These physiological adaptations are not simply the result of local acclimatization, but instead arose over thousands of years of evolution under chronic exposure to high-altitude hypoxia since the initial peopling of the Andes.[33,34] Genomic analyses of Native Andean populations have identified several candidate genes involved in these adaptations, including *EGLN1* (MIM: 606425), *NOS2* (MIM: 163730), and several genes involved in the hypoxia-inducible factor 1 (*HIF1* [MIM: 614529]) pathway.[32,35,36] Ancient DNA analyses of pre-Columbian Andean genomes indicate that some of these genes were already under positive selection at least 8,500 years before the present.[37] Although neither of these loci have been conclusively linked to health outcomes, altered HIF-regulatory gene expression patterns are found in placentas from preeclamptic individuals.[30] Moreover, measured basal arterial pressure levels among Andeans adults living at high altitudes are lower than those observed in low-altitude cohorts, possibly due to altitude adaptation.[38]

Previous research has found that highland Andean ancestry and long-term, multi-generational residence at high altitude are associated with lower rates of hypoxia induced pregnancy complications among high-altitude-resident women.[21,30,39] These findings suggest that Andeans with Native American ancestry may carry distinct adaptive variants or a unique repertoire of genetic risk factors for preeclampsia—distinct from other populations previously studied.[40] Characterizing fine-scale ancestry and genetic variation patterns in high-altitude-resident Andeans may uncover preeclampsia-associated variants found at higher frequencies in affected individuals, possibly as a result of adaptive processes to high altitude.[41,42]

To this end, here we analyze genome-wide genotyping data from a large cohort of preeclamptic Andean families from Puno, Peru (Figure 1). This city, located at 3,830 m of altitude, has a population with one of the highest incidences of preeclampsia and associated maternal mortality in the world.[24,43,44] The study cohort was recruited at the Puno regional hospital (Hospital Manuel Nuñez Butrón), where severe preeclampsia is the third leading cause of hospitalization for women of reproductive age.[45] We note that in addition to genetic and altitude-dependent factors, this high incidence may also be exacerbated by limited access to healthcare. As of 2017, the department of Puno, where Puno city is located, had only 5.1 obstetric professionals per 10,000 inhabitants and the highest rate of infant mortality in Peru.[46,47] A revision of current literature indicates that Peru ranks fifth globally in preeclampsia prevalence, and fourth compared to other middle-income countries in the Americas (Table S1).

Our work takes a comprehensive approach to the genetic study of preeclampsia in a population adapted to high altitude by employing a family-based study design within a

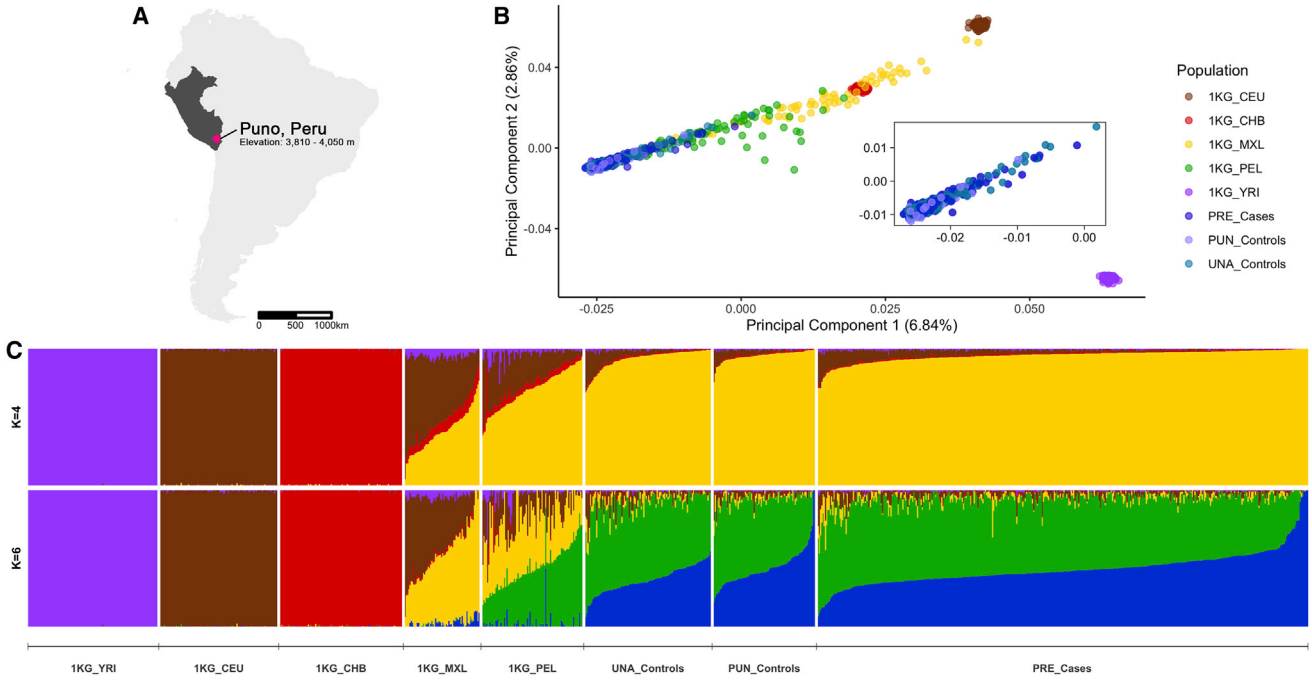

Figure 1. Location and population structure of the Puno preeclampsia cohort
(A) Approximate location and altitude of Puno, Peru.
(B) Principal components analysis including PRE-affected individuals, PUN and UNA control subjects, and five continental reference populations from the 1000 Genomes. Inset zoom shows Puno cohort individuals only.
(C) ADMIXTURE analysis results showing unsupervised clustering models assuming K = 4 and K = 6. At K = 6 a Puno-specific sub-continental ancestry component not shared with 1KG Peruvians from Lima appears in the Puno cohort, shown here in dark blue.

case-control cohort. Our trio sampling approach allows us to overcome confounding factors due to unaccounted population structure and enables identification of genetic regions that influence preeclampsia considering each of the family members that affect disease risk—mothers, fathers, and offspring. This is unlike most genome-wide studies focused on pregnancy disorders which tend to solely include maternal or fetal genomes.[48] We also characterize genetic diversity and admixture patterns in the Puno cohort, and we investigate whether signals of selection are present in the regions surrounding candidate associated genes. Additionally, because preeclampsia unfolds in a spectrum of severity based on gestational age, organ damage, and hypertension, we take advantage of extensive cohort phenotyping to study associations of genetic variants with disease severity. Lastly, we re-analyze our case-control test with an external control cohort of Latin American populations.[49] Our findings have implications for the general understanding of preeclampsia etiology, and the etiology of human pregnancy hypertensive disorders more broadly, while also shedding light on the genetic factors that underlie human adaptations for successful reproduction at high altitudes.

Material and methods

Puno cohort

Preeclampsia-affected families (PRE) were recruited between 2011 and 2016 in the Puno regional hospital (Hospital Regional Manuel Nuñez Butrón) after their preeclampsia diagnosis. Expecting parents (mothers and fathers) had to be at least 18 years of age and report at least two generations of parents from Puno or nearby Andean regions. Recruited families and participants included 136 trios (mother, father, and fetal umbilical cord), 197 duos (190 mother and fetal umbilical cord duos, and 7 mother and father pairs), and 14 singletons (mother or umbilical cord only). 100 healthy same-population control families from Puno (PUN) were also recruited at the hospital at their time of admission for labor. These included 4 trios and 96 duos (mother and fetal umbilical cord). Lastly, 110 unrelated male and female healthy population control subjects were recruited at the local university, Universidad Nacional del Altiplano (UNA) in Puno. In total, 1,129 samples were collected, including 815 PRE-affected individuals, 204 PUN hospital control subjects, and 110 UNA population control subjects (Table S2).

Ethical approval

All participants were recruited with informed consent and IRB approval by the Stanford University Institutional Review Board eProtocols 20782 (Investigating the Genetic Basis of Preeclampsia in Populations Adapted to High Altitude) and 20839 (Population and Functional Genomics of the Americas). Local IRB approvals were obtained from the ethics committee of the Manuel Nuñez Butrón Regional Hospital (01541-11-UADI-HRMNB-RED-PUNO) and the Peruvian National Institute of Health (213-2011-CIEI/INS).

Phenotypic data

Preeclampsia was defined as new onset of hypertension with presence of proteinuria in the urine of pregnant mothers after 20 weeks

of gestation (Table S3). Hypertension in the mother was defined as systolic blood pressure 30 mmHg higher than basal level, and diastolic blood pressure at least 15 mmHg higher over basal level. If no prior blood pressure measurements were available, average basal levels were used as prior (85/55 mmHg). Note that measured basal arterial pressure levels in pregnant women of the Puno cohort are around 80/50–90/60 mmHg (systolic/diastolic), much lower than the U.S. standards.[38] Proteinuria levels were confirmed to be at least 30 mg/dL by dipstick in two tests 24 h apart. Severity of preeclampsia was defined by the attending physician and categorized into mild or severe based on the physician's diagnostic criteria which included assessment of blood pressure and proteinuria measurements, together with additional symptoms (such as vomiting or blurred vision) not specified on the charts used for our analyses. We acknowledge that physician-level variation and ascertainment bias at the time of diagnoses may introduce some variation in the categorization of preeclampsia severity, as has been noted for other health conditions.[50] Gestational time was self-reported by the mother (by date of last menstrual period [LMP]) or determined by the neonate Capurro test.[51,52] Blood pressure and proteinuria measurements were not collected for fathers and newborns. A non-parametric Wilcoxon rank test was used to test for statistical differences in blood pressure levels between affected mothers with mild and severe preeclampsia diagnoses with the wilcox.test function in R v.4.1.2.

Blood and tissue collection

Whole blood from the mothers was collected within a few hours post-partum by venipuncture into EDTA tubes and frozen at −20°C. Umbilical cord blood was collected by venipuncture following cord clamping immediately after delivery. Blood from fathers and from UNA control subjects was obtained at recruitment sites (hospital or university, respectively) upon consent. For plasma, EDTA tubes were spun within 60 min of collection at 1,200 × *g* for 10 min in a tabletop centrifuge. Separated plasma was transferred to Eppendorf tubes, spun again under the same conditions for better purity, then stored in cryovials at −20°C.

Genotypic data

DNA was obtained from whole blood with the Promega (USA) Wizard Genomic DNA Purification Kit following manufacturer's instructions. Samples that had both >10 ng/μL of DNA concentration, as quantified via fluorometer, and visible bands on a 1% agarose gel were selected for genotyping. In total, 950 samples were genotyped in two batches using the Affymetrix (USA) Axiom Genome-wide LAT 1 array (supplemental material and methods). Batch 1 was genotyped in 2014 at the University of California San Francisco, Gladstone Genomics Core in San Francisco, CA. In this batch, 360 PRE-affected, 10 PUN control, and 110 UNA control individuals (n = 480), and 813,366 variants were successfully genotyped. Batch 2 was genotyped in 2018 at Affymetrix Research Services Laboratories, Thermo Fisher Scientific in Santa Clara, CA. This batch included 324 PRE-affected and 146 PUN control individuals (n = 470), as well as 10 controls added by the genotyping facility. Three samples failed the genotyping facility filtering thresholds, so a total of 477 samples and 818,154 variants were successfully genotyped with batch 2.

Quality control

We performed separate quality control assessments for the genotype data generated in batch 1 and batch 2 (see supplemental material and methods for more details). For both datasets, we first restricted the variants to a list of recommended SNPs provided by Affymetrix. Next, we used Plink v.1.9[53] to remove variants with duplicate marker names, structural variants, variants with no physical position in the NCBI Build GRCh37 human reference, and the genotyping controls. Lastly, we flipped all SNPs to the forward strand using snpflip (see web resources). After QC, the batch 1 dataset included 713,667 bi-allelic SNPs and 480 individuals, and the batch 2 dataset included 777,946 bi-allelic SNPs and 467 individuals. We next intersected batch 1 and 2 datasets at overlapping sites and did another round of quality control on the merged dataset. We removed SNPs with genotype missing rate >5%, minor allele frequency (MAF) <0.5%, and failing Hardy-Weinberg equilibrium at 10E−10. We also filtered for individuals with call rate <90%, duplicate individuals, and cryptic relatedness. In total, 31 individuals were removed, and 56 family pedigrees were updated after this second round of QC on the merged dataset (see Table S4 for list of individuals assigned as unrelated after pedigree revision). We found that for some affected and control families, offspring sex was missing, or some individuals' sex was misassigned in the medical records at the time of data collection (e.g., mothers noted as male, fathers noted as female). Therefore, chromosomal sex was estimated using Plink v.1.9 for 176 individuals whose biological sex was either not previously recorded (n = 124, mostly offspring) or was incorrectly recorded (n = 52) during sample collection. After QC, the merged dataset (batch 1 + 2) included 504,475 genome-wide distributed SNPs and 914 individuals (Figure S1).

Finally, we tested for batch effects on the merged dataset by calculating principal components analysis (PCA) in Plink after filtering for linkage disequilibrium (LD) and removing related offspring. We initially identified a strong batch effect with the top principal components statistically significantly associated with batch ($p < 0.05$) (Figure S2). To correct this effect, we conducted an additional round of site and sample-specific filtering, again filtering for excess heterozygosity and cryptic relatedness. This resulted in a further 31 individuals being removed from the dataset. We then repeated the PCA as detailed above. The final dataset after batch effect correction included 439,314 SNPs and 883 individuals.

Population structure

We intersected our dataset with reference panels including five populations from 1000 Genomes Phase 3 (1KG): Yoruba from Ibadan, Nigeria (YRI), Utah residents with Northern and Western European ancestry (CEU), Han Chinese from Beijing, China (CHB), Mexican Americans from Los Angeles, USA (MXL), and Peruvians from Lima, Peru (PEL). After merging, we removed offspring and related individuals, restricted to autosomal markers and re-applied quality filters. The filtered, merged dataset consisted of 422,224 variants and 1,057 individuals. The unsupervised clustering algorithm ADMIXTURE[54] was run on this dataset to explore global patterns of population structure. After LD pruning, 45,496 variants remained for analysis. Ten ancestral clusters (K = 2 through K = 10) were tested and the best fit model was selected after examining cross-validation errors. To account for possible convergence variation, we performed 10 additional runs using different random seeds per run and estimated parameter standard errors using 200 bootstrap replicates per run. ADMIXTURE results were plotted with the R pophelper package.[55] PCA was applied to the LD pruned dataset using EIGENSOFT v.7.2.1[56] and plots were generated using the ggplot2 package in R v.4.0.3.[57,58]

Table 1. Phenotypic characteristics measured in mothers of analyzed affected families with preeclampsia (duos and trios)

	Total	Trios only
Diagnosis, n	mild (n = 119)	mild (n = 47)
	severe (n = 106)	severe (n = 41)
Mode of delivery, n	C-section (n = 92)	C-section (n = 42)
	vaginal (n = 132)	vaginal (n = 45)
Maternal age, mean years (median)	26.64 (26.00)	27.44 (26.00)
Gestational age (maternal), mean weeks (median)	38.11 (39.00)	37.84 (38.00)
Gestational age (fetal), mean weeks (median)	38.29 (39.00)	38.17 (39.00)
Newborn weight, mean grams (median)	601.3 (600.0)	620.6 (620.0)
Systolic BP at admission, mean mmHg (median)	133.0 (130.0)	130.9 (130.0)
Diastolic BP at admission, mean mmHg (median)	89.82 (90.00)	87.1 (85.0)
Parity, n	nulliparous (n = 128)	nulliparous (n = 45)
	1 or more (n = 97)	1 or more (n = 43)
Sex of newborn, n	female (n = 100)	female (n = 35)
	male (n = 125)	male (n = 53)
Proteinuria, n	+ (n = 124)	+ (n = 49)
	++/+++ (n = 101)	++/++ (n = 39)

"Trios only" identifies the subset from the total that are in whole trio units; the rest are mother-offspring duos. Average systolic/diastolic blood pressure in preeclampsia cohort mothers is 130/90 mmHg, approximately 30 mmHg and 15 mmHg higher than basal levels.

Phasing and local ancestry estimation

We used RFMix v.1.5.4[59] to infer genome-wide local ancestry calls for the Puno cohort founders, assuming a model of K = 3 ancestral populations. Given that local ancestry methods have decreased accuracy when attempting to distinguish between closely related ancestries,[59,60] in this analysis we focused on the three major continental ancestries identified in the Puno cohort through the ADMIXTURE analysis. The reference panel included 108 YRI and 94 CEU individuals from 1KG, and 94 native individuals from Mexico (30 Mixe, 15 Zapotec, 49 Nahua) genotyped as part of the GALA II study.[61] These reference samples were used as proxies for African, European, and Native American ancestral source populations, respectively. After merging, the working dataset consisted of 420,105 overlapping variants and 899 individuals. The data were phased with SHAPEIT2[62] and RFMix was run with default parameters and EM = 2 iterations. Ancestry call cutoffs were determined with a 0.9 posterior probability threshold as recommended in Kidd et al.[63]

Ancestry proportions analysis

We tested for significant differences in proportions of Native American, European, and African ancestry components between PRE-affected individuals and PUN and UNA control subjects. We applied the Wilcoxon signed ranks test in R v.3.5.1 (pairwise.wilcox.test function) with Bonferroni correction for multiple testing. This non-parametric test assesses whether significant differences exist between two distributions.[64] Our null hypothesis was that the distribution of each ancestry proportion was identical between PRE-affected individuals and PUN and UNA control subjects.

Selection scan analysis

We computed the integrated haplotype score statistic (iHS)[65] using Selscan[66] to find genomic regions with signatures compatible with positive selection in the founders of PRE-affected individuals and UNA control subjects. After removing variants and individuals with missing genotype rates more than 5%, this analysis included 405 PRE-affected and 106 UNA control individuals and 253,896 SNPs (see supplemental material and methods). Selscan was run with default parameters. Unstandardized iHS values were normalized by allele frequency bins. Since positive selection decreases genetic diversity in the region around the selected allele (leading to longer haplotypes),[67] we considered as candidate signals those regions encompassing an SNP in the top 1% of the iHS distribution (99th percentile) and that had at least one other flanking SNPs, also in the top 1%, within a 50 kb window (see supplemental material and methods for more details).

Statistical analysis of clinical phenotypes

We assessed batch bias of clinical phenotypes measured in affected mothers and correlation with each other by statistical analysis in R v.3.4.0. The following dichotomous phenotypes were tested for batch association with a chi-squared test: severity of diagnosis (mild or severe), proteinuria (+/++ or +++), parity (nulliparous or more than one birth), newborn sex, and delivery mode (vaginal or C-section). The following continuous phenotypes were tested for batch association by t test: gestational time measured by the mother (date of last menstrual period [LMP]) and by the fetus (Capurro test), neonate weight, systolic and diastolic blood pressure measurements, and maternal age. We provide summaries of relevant phenotypic data for all case pregnancies in Table 1. In addition to the data shown in Table 1, we also note that most mothers (>98%) had no history of chronic hypertension or diabetes and all were non-smokers. By statistical comparison, we found that there was moderate batch bias in approximately half of the measured phenotypes in affected mothers (e.g., batch 2 had significantly more vaginal deliveries than C-sections, when compared to batch

1, $p < 0.04$), but none likely to influence the analysis when supported by batch correction. We also explored potential correlations between the secondary phenotypes of preeclampsia (severity, proteinuria, gestational age, and blood pressure) and all additional phenotype variables collected from the clinical records of the preeclamptic individuals, including maternal age, offspring sex, gravidity, and parity. To this end, we generated a Spearman correlation coefficient matrix of all available variables for the PRE-affected individuals. This analysis was performed in R v.4.0.0 using the function 'rcorr' from the 'Hmisc' library, with type = 'spearman' (Figure S3).

Transmission-disequilibrium test (TDT) and parent of origin (TDT-POO) tests

Leveraging the trio family structure, we applied the transmission disequilibrium test (TDT) and parent-of-origin (TDT-POO) test on all 87 parent-offspring case trios (preeclamptic families with offspring) in Plink v.1.9 using the –tdt flag, with and without the 'poo' modifier. Variants were then filtered by MAF > 0.05 within the analyzed cohort. The TDT assumes Mendelian rules for transmission of alleles and tests if the queried allele is being transmitted/untransmitted disproportionately from parents to the affected offspring population.[68,69] The POO analysis is part of the TDT and separately queries transmission from each parent individually to assess paternal- or maternal-specific transmission. This test self-corrects for covariate effects by treating each trio as a separate unit.

To obtain an appropriate significance threshold for the TDT, we used the 'max(T) permutation test' option implemented in Plink. This option permutes the transmitted/non-transmitted allele from each parent and re-calculates the test statistic. The maximum value of the test statistic across all genome-wide SNPs was taken, and the distribution of this maximum value across 30,000 permutations was considered. To take LD between SNPs into account, permutations were conducted considering "strong LD" blocks, as implemented in the Plink –block feature, representing the block definition suggested by Galanter et al.[61] The top 1% significance thresholds determined by this permutation approach were $p = 8E{-}6$ and $p = 8E{-}5$ for the TDT and TDT-POO analysis, respectively. The top 5% significance thresholds, which were used as suggestive thresholds, were $p = 3E{-}5$ and $p = 2E{-}4$ for TDT and TDT-POO, respectively. Note that although the classical threshold of $p = 5E{-}8$ corresponds to a 5% error rate, we decided to be conservative and use 1% as the genome-wide significance threshold and 5% as suggestive (see supplemental material and methods for more details).

GWAS for case-control association and additional analysis using external controls

Puno cohort individuals recruited at the hospital were divided into offspring and mothers for two separate case-control GWAS analyses using logistic regression in Plink (flag: –logistic) with the first 3 PCs, sequencing batch, and maternal age included as covariates. The analysis on the mothers includes 254 PRE-affected individuals and 70 PUN control subjects. The offspring analysis includes 225 PRE-affected individuals and 60 PUN control subjects. These analyses included individuals in trios, duos, and singletons. Variants were filtered by MAF > 0.05 within the analyzed cohort.

To further investigate the associated variants, we conducted two additional case-control GWAS analyses including 551 individuals from the CANDELA Consortium as external control subjects.[49] We note that these additional analyses do not constitute independent replication, as the CANDELA cohort does not include preeclamptic individuals, so a CANDELA-only replication analysis could not be performed. Yet, incorporating external controls can serve as a valid strategy to increase statistical power, especially when affected individuals are rare or difficult to sample. The SNP array-based CANDELA dataset includes multiple populations of diverse origin across Latin America. Thus, to mitigate the effects of population structure and include control subjects with ancestry related to the Puno cohort, we selected Peruvian and Chilean individuals with >70% Native American ancestry and only Andean Native American ancestry (i.e., related to Andean native populations such as Aymara, Quechua, or Uros and <1% lowland Mapuche Native American ancestry) from the CANDELA dataset.[70] We next merged these two datasets at overlapping autosomal sites and performed an additional round of quality control after which we excluded a single PUN control individual and retained a total of 381,675 autosomal SNPs (supplemental material and methods). Visual inspection of the scree plots and PCA scatterplots resulting from the merging of the Puno cohort and CANDELA datasets did not show any apparent batch effect (Figures S4–S6), only some variation due to varying levels of European and regional Native American ancestry. Finally, reproducing the earlier analyses, we conducted the two separate case-control GWAS analyses of mothers and offspring as affected individuals with this additional set of 551 Peruvian and Chileans from CANDELA as control subjects (i.e., including PRE, PUN, and CANDELA individuals), using logistic regression in Plink with the first 3 PCs as covariates. Maternal age was not included in this analysis as this information was not available for the CANDELA dataset.

GWAS of additional phenotypes

Multiple phenotypes measured and captured as part of the recruited patient's medical history allow for testing of additional genetic associations in the Puno cohort. We performed genome-wide association analyses of endophenotypes in the PRE-affected mothers (n = 253) and offspring (n = 225), separately. These analyses included individuals in case trios, duos, and singletons. Control individuals without a preeclampsia diagnosis were not included in this analysis. The endophenotypes tested for each were (1) gestational age, maternal measurement; (2) gestational age, fetal measurement (see phenotypic data section above for measurement description); (3) systolic blood pressure at diagnosis of preeclampsia; (4) diastolic blood pressure at diagnosis of preeclampsia; (5) proteinuria at diagnosis; and (6) severity of diagnosis. The first four were treated as continuous variables and analyzed by linear regression in Plink (flag: –linear). Proteinuria and severity of diagnosis were dichotomous variables analyzed in Plink by logistic regression (flag: –logistic), with proteinuria reduced to + and ++ vs. +++. Genotyping batch was included as a discrete covariate, while the first 3 PCs and maternal age were included as continuous covariates. Several of these analyses included less individuals due to missing data. Specifically, GWASs with systolic and diastolic blood pressure included 252 PRE-affected mothers and 224 PRE-affected offspring and GWASs with maternal measurement of gestational age included 251 PRE-affected mothers and 223 PRE-affected offspring.

PROZ ELISA

PROZ levels in post-partum maternal and cord blood plasma were assayed using the human-PROZ ELISA kit from MyBioSource (USA,

Cat. No. MBS765710), following manufacturer's instructions. Maternal and fetal plasma samples were diluted at 1:400 and washes were performed manually with a multichannel pipet. Final optical density absorbance at 450 nm was read using the Bio Rad (USA) iMarkTM Microplate Absorbance reader. A 4-Parameter curve fit was applied to the standards, and the resulting equation was used to calculate concentration in the experimental samples. Boxplots and t tests were done in R v.3.4.0.

GWAS data visualization

All genome-wide analyses were filtered by MAF $\geq$ 0.05 within the analyzed cohorts and visualized by Manhattan plots using the qqman R package v.0.1.4.[71] QQ plots were generated with the same package to confirm no effects from population structure or other confounders. Top suggestive SNPs and their genomic regions were plotted using LocusZoom.[72] Maps displaying the geographic distribution of candidate associated variants were produced using the Geography of Genetic Variants (GGV) browser (web resources).[73]

Functional annotation data

Functional annotation of candidate associated variants was reviewed using the Open Targets Genetics Platform.[74] We queried for significant associations linking these variants to changes in gene expression (eQTLs) or protein abundance (pQTLs) in blood based on the eQTLGen Consortium[75] and Sun et al.,[76] respectively. We note that unlike the eQTLGen Consortium data, the study by Sun et al.[76] contained blood level measurements of approximately 3,000 proteins. To prioritize candidate causal genes, we assessed whether the expression of the genes within the region was controlled by the associated SNP (i.e., is an eGene[75]) or whether the gene had the highest overall Variant-to-Gene (V2G) scores from the Open Targets Genetic Platform. The overall V2G score aggregates differentially weighted evidence of variant-gene association from several sources, including *cis*-QTL data, chromatin interaction experiments, *in silico* function predictions (e.g., Variant Effect Predictor from Ensembl), and distance between the variant and each gene's canonical transcription starting site.

Sequencing capture

We conducted fine mapping of potential causal variants in a subset of families genotyped in batch 1 prior to batch 2 genotyping. Preliminary data obtained from batch 1 genotypes were analyzed using standard family-based TDT on Plink for preeclampsia associations (as above), and regression analysis on secondary phenotypes was conducted using linear mixed models in GTCA[77] (flag: –mlma-loco). Based on these preliminary results, we designed a targeted capture assay including windows around top hits for preeclampsia and secondary phenotypes, as well as several genes previously suggested to be associated with preeclampsia in the GWAS catalog (release 2.0.5).[78] The total capture size was approximately 10 Mb (see supplemental material and methods).

We next selected families from batch 1 with the strongest associations on the preliminary TDT analysis (n = 86 individuals, Table S2). Genomic DNA from 86 individuals (Figure S7) was fragmented by mechanical shearing (Covaris) and prepared using the KAPA Hyperprep library preparation kit (Kapa Biosystems). DNA capture was performed on the libraries using the Agilent (USA) SureSelect platform following manufacturer's instructions. Paired-end sequencing of captured libraries was performed on Illumina's NextSeq platform. Sequence data were analyzed through a standard FASTQC-BWA-GATK pipeline following published guidelines.[79] We then performed the same GWAS analyses listed above (TDT test for the preeclampsia phenotype and linear regressions for continuous phenotypes) in the captured regions for a limited set of individuals: 25 trios, 4 duos (3 mother-offspring, 1 father-offspring), and 3 singletons (1 offspring and 2 mothers). Candidate loci identified in these analyses were individually merged and annotated with ANNOVAR[80] and overlapped with GTEx single-tissue *cis*-eQTL data (version V6p) from the online database[81,82] to find relevant GTEx annotations in our data set.

Results

We obtained blood samples and maternal clinical records from consented families at the Hospital Regional Manuel Nuñez Butrón and blood alone from individuals recruited at the Universidad Nacional del Altiplano (UNA). At the time of recruitment, mothers from affected families (labeled PRE throughout this study) were at the hospital experiencing pregnancy with a preeclampsia diagnosis, defined as hypertension and proteinuria after 20 weeks of gestation. Rather than based on a hard cutoff, hypertension was defined as a systolic measurement 30 mmHg higher than basal and diastolic at least 15 mmHg higher than estimated basal levels for each individual (see material and methods for more details). We note that, on average, systolic/diastolic blood pressure in the preeclampsia cohort mothers is approximately 130/90 mmHg, approximately 40 mmHg and 30 mmHg higher than their estimated basal levels (80/50–90/60) mmHg, and well over the diagnostic thresholds noted above. This measurement is also comparable to the high blood pressure cutoff previously reported in the literature for adult residents of the Andean highlands (134/89 mmHg).[38] We further note that maternal blood pressure levels differed somewhat by severity in the Puno preeclampsia cohort. Average systolic/diastolic blood pressure measurements in mothers with a severe preeclampsia diagnosis (138/94 mmHg) were approximately 9/7 mmHg higher than in mothers with a mild preeclampsia diagnosis (129/87 mmHg). This difference between mild and severe cases was statistically significant in a Wilcoxon test for both systolic ($W = 8660$, $p = 9.2E{-}12$) and diastolic ($W = 8345.5$, $p = 4.34E{-}13$) blood pressure measurements.

For consistency, and to control for other hypertensive complications of pregnancy, we included proteinuria in the diagnosis, despite this factor not being currently required in many diagnostic guidelines.[83] Mothers from control families (labeled PUN) were experiencing a pregnancy without complications at time of hospital recruitment. 88 PRE-affected families and two PUN control families were collected as complete trios—including both biological parents and offspring; the rest are duos (one parent and offspring) or single individuals (mothers) (Table 2). Overall, the Puno cohort collected for this study includes 815 individuals from the hospital case group (PRE), 204 from the

Table 2. All individuals genotyped by group (case/control) and batch after QC filtering

	PUN and UNA control subjects				PRE-affected subjects				
Family category[1]	Batch 1 individuals	Batch 2 individuals	Batch 1 + 2 individuals	Batch 1 + 2 family units	Batch 1 individuals	Batch 2 individuals	Batch 1 + 2 individuals	Batch 1 + 2 family units	Total no. of individuals
Trios[a] (M + F + UC)	0	6	6	2	241	21	262	88[c]	268
Duos	10	106	*116*	58	110	192	302	151	418
M+UC	*8*	*106*	*114*	*57*	*66*	*188*	*254*	*127*	*368*
P+UC	*2*	*0*	*2*	*1*	*18*	*2*	*20*	*10*	*22*
M+F	*0*	*0*	*0*	*0*	*26*	*2*	*28*	*14*	*28*
Singletons [b]	106	24	130	–	19	48	67	–	197
All individuals	116	136	252	60	370	261	631	239	883

Text in italics indicates the number of individuals genotyped within three different types of family duos.
[a]M, mother; F, father; UC, offspring umbilical cord.
[b]Includes one trio with two offspring from family PRE061.
[c]Includes individuals coded as UNR (unrelated and no longer connected to medical records), UNA (university controls), and some individuals collected as part of PRE/PUN (still connected to medical records).

hospital control group (PUN), and 110 from the university (UNA) as population controls. We extracted DNA from blood and genotyped PRE, PUN, and UNA individuals in two batches on the Affymetrix Axiom LAT array. Our final dataset after quality filtering included 439,314 SNPs genome wide and 883 individuals (see Tables 2 and S2 for breakdown of PRE, PUN, and UNA groups). Our sample represents 0.7% of the population of Puno city which was estimated at approximately 128,637 inhabitants in the 2017 census (Censo Nacional de Población y Vivienda) of Peru.[84]

Puno individuals have high proportions of native American ancestry

We sought to understand the ancestral background of our test population by characterizing patterns of genetic diversity and population structure in the Puno cohort. To this end we intersected the entirety of the Puno cohort dataset (n = 883) with a reference panel including five continental populations from the 1KG panel: YRI, CEU, MXL, CHB, and PEL. Using PCA, we find that individuals from Puno (either PRE, PUN, or UNA) cluster together in PC space and are distributed in a clinal pattern alongside Peruvians from Lima (PEL) who have high proportions of Native American ancestry (Figures 1 and S8).

We next investigated admixture patterns in the Puno population with the goal of estimating proportions of Native versus non-Native genomic ancestry. Using the clustering algorithm ADMIXTURE,[54] we explored unsupervised models assuming K = 2 through K = 10 ancestral clusters (Figure S9). Cross-validation errors for each K cluster are shown in Figure S10. At K = 4, we observe a clear separation of continental ancestry components. We find that Puno individuals have large proportions of Native American ancestry and small proportions of European ancestry, represented by yellow and brown in Figure 1, respectively. At the best fit model of K = 6, ADMIXTURE analysis finds substructure within the Native American ancestry component of the Peruvian population. Specifically, we observe a Puno-specific ancestry component (shown in dark blue in Figure 1) which is not present within the Native American ancestry components of 1KG Peruvians from Lima (PEL) and Mexicans (MXL), which show yet another local component. This substructure may derive from an Andean-specific ancestry component that has been previously identified among Indigenous and mestizo communities from the Andean highlands.[85,86] Overall, we find that individuals in the Puno cohort are predominantly of Native American ancestry (95.7% on average) and have low levels of non-Native American admixture (approximately 4.2% on average; Table S5). We further find that the Puno population carries a highland-specific Native American sub-continental ancestry component, as noted in previous research.[85–87]

We tested for significant differences in continental ancestry proportions between affected individuals (PRE) and control subjects (PUN, UNA) in the Puno cohort. Given that local ancestry methods have decreased accuracy when attempting to distinguish between closely related ancestries,[59,60] in this analysis we focused on the main continental ancestries identified in the ADMIXTURE analysis. Thus, we used RFMix to determine local ancestry proportions in the Puno cohort assuming a model of K = 3 ancestral components (e.g., Native American, European, and African ancestries). We note that this model is also consistent with those used by previous research investigating Peruvian population structure.[86,88] We then estimated average ancestry proportions per individual from the RFMix local ancestry calls (Tables S6 and S7). The results of this estimation further confirm the predominantly Native American ancestry background and highlight the small proportion of European admixture present in our sample.

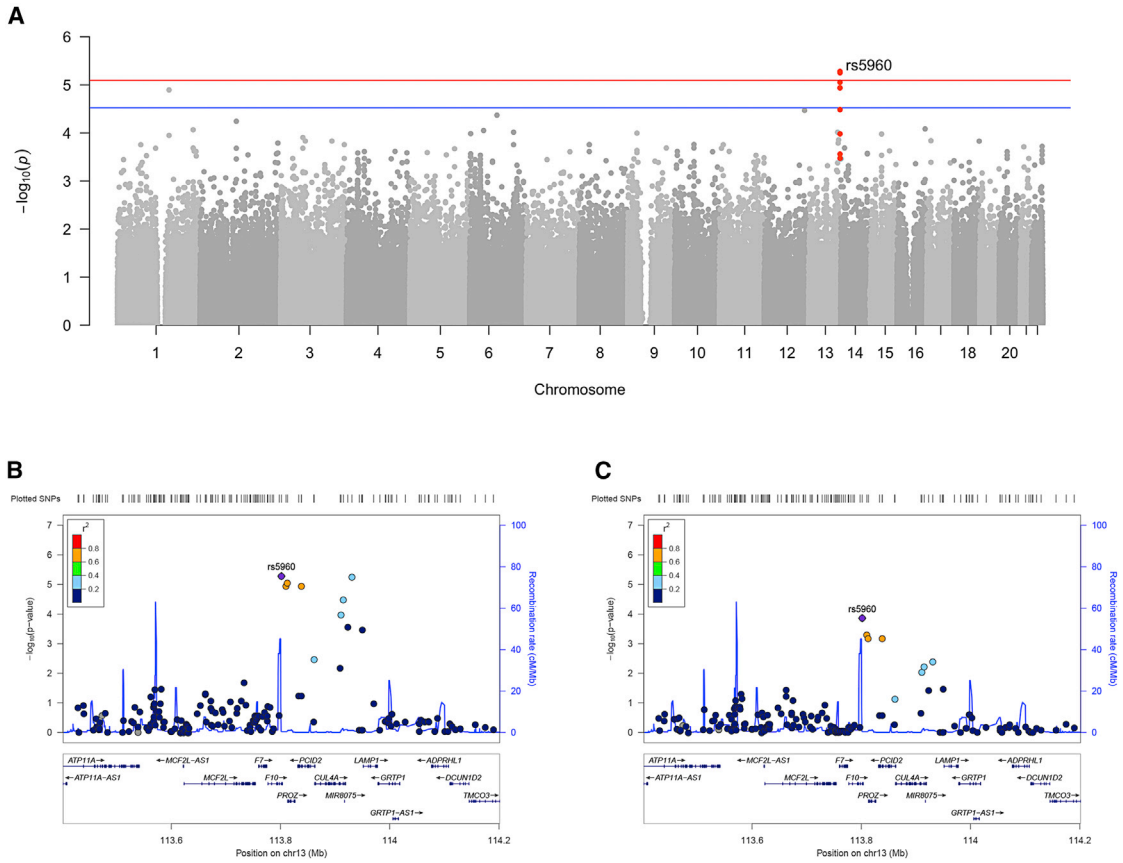

Figure 2. Top associations from trio analyses by TDT and TDT-POO

(A) Manhattan plot showing top association with preeclampsia in the offspring genome: SNP rs5960 on chromosome 13 at p < 10E−05 suggestive of significance shown in red. Horizonal lines indicate top 1% (red) and 5% (blue) significance thresholds of p = 8.E−06 and p = 3.E−05, respectively, as determined by permutation approach.

(B) Locus Zoom plot depicting the top associated SNP cluster from the TDT on chromosome 13.

(C) Locus Zoom plot depicting the top paternal region from TDT-POO analysis on chromosome 13.

Finally, we performed a Wilcoxon rank test to contrast ancestry proportions between PRE, PUN, and UNA. This test identified a small but significant difference in European ancestry proportions between PRE and UNA (estimated European ancestry PRE 3.61% and UNA 6.07%, Wilcoxon rank test p = 0.015) but found no significant differences in Native American or African ancestry proportions (see Table S8 for p values and Figure S11 for box plots). Overall, UNA individuals have slightly higher proportions of European ancestry than PRE and PUN individuals. However, proportions of Native American ancestry are not significantly different between affected individuals and control subjects (Native American ancestry PRE 96.37%, PUN 95.98%, and UNA 93.90%; Wilcoxon rank test PRE vs PUN: p = 0.378; Wilcoxon rank test PRE vs UNA p = 0.763). These findings confirm the results of the ADMIXTURE analysis and further support the study design focused on a cohort of primarily Native American ancestry background.

Family-based analysis reveals association of a cluster of clotting factor genes (*PROZ*, *F7*, *F10*) with preeclampsia

Next, we sought to identify genetic loci associated with the risk of preeclampsia in the Puno cohort. We first performed a parent-offspring trio GWAS analysis, or transmission-disequilibrium test (TDT), in the 88 affected (PRE) trios. The TDT offers a robust association test of genotype to phenotype in affected families by measuring over-transmission of alleles from heterozygous parents to the offspring. With this analysis, we identified a group of SNPs in LD over a cluster of blood clotting factor genes with a high odds ratio for preeclampsia (Figure 2; Table 3; Figure S12). The most significant SNP in this cluster, rs5960 (OR 3.05, 95% CI 1.841–5.054, p = 5.23E−06; global 1KG MAF 0.377), is a synonymous variant in the clotting factor *F10* (MIM: 613872). This variant is genome-wide significant based on the 1% cutoff (p = 8E−06) calculated from the permutation analysis to determine p value thresholds (see material and methods). Two other members of the coagulation cascade, *F7* (MIM: 613878) and *PROZ* (MIM: 176895), are also in this region, as part of a gene cluster within 50 kb up- and downstream of the *F10* locus. Another top hit in the TDT, SNP rs553316 (OR 0.339, 95% CI 0.2041–0.5629, p = 1.15E−05; global 1KG MAF 0.408), is suggestive based on the 5% permutation threshold. Interestingly, rs553316 is in high LD with rs5960 in 1KG Peruvian populations (R^2 = 0.7476).[89]

Table 3. GWAS statistics and genomic annotations of top hits (p < 5E−04) from the TDT

						TDT GWAS stats					
Chr	**BP**	**cytoBand**	**rsID**	**Ref**	**Alt**	**OR**	**95% CI**	**P**	**Function**	**Genes in region**	**Puno MAF**
13	113801737	13q34	rs5960	C	T	3.05	1.841–5.054	5.23E−6	exonic	**F10**	0.4929
13	113930853	13q34	rs9549724	C	T	0.2963	0.1696–0.5176	5.58E−6	intergenic	CUL4A, LAMP1	0.4278
13	113812962	13q34	rs2273971	A	G	0.3276	0.1951–0.55	8.81E−6	upstream	**PROZ**	0.499
13	113838015	13q34	rs553316	G	A	0.339	0.2041–0.5629	1.15E−5	intronic	PCID2	0.4827
13	113810186	13q34	rs7335409	T	C	2.95	1.777–4.899	1.15E−5	intergenic	**F10, PROZ**	0.4939
13	113915303	13q34	rs3814260	G	A	0.3396	0.199–0.5797	3.27E−5	intronic	CUL4A	0.4236
2	109581319	2q12.3	rs260692	C	T	19	2.544–141.9	5.70E−5	intronic	**EDAR**	0.0551
1	228715705	1q42.13	rs11586639	G	A	0.4	0.2492–0.6422	8.57E−5	intergenic	BTNL10, MIR7641-2	0.4587
6	42252385	6p21.1	rs9471831	A	G	0.4242	0.2727–0.6601	8.88E−5	intronic	**TRERF1**	0.4562
13	113910926	13q34	rs3861723	A	G	0.36	0.2101–0.617	1.04E−4	intronic	CUL4A	0.4228
1	228805855	1q42.13	rs765070	C	T	0.4107	0.2528–0.6673	2.05E−4	intergenic	DUSP5P1, **RHOU**	0.4347
2	109557099	2q12.3	rs260711	T	C	9.5	2.213–40.78	2.08E−4	intronic	EDAR	0.05612
1	229076157	1q42.13	rs10916389	G	A	0.325	0.1738–0.6076	2.08E−4	intergenic	**RHOU, MIR4454**	0.2368
13	113923202	13q34	rs77626225	A	G	0.3878	0.2283–0.6586	2.75E−4	intergenic	CUL4A, LAMP1	0.2307
13	113949751	13q34	rs9549380	G	A	0.4	0.2382–0.6718	3.36E−4	intergenic	CUL4A, LAMP1	0.3201
2	109513601	2q12.3	rs3827760	A	G	9	2.088–38.79	3.47E−4	exonic	EDAR	0.05793

eGenes (based on eQTL analyses using blood-derived expression through the eQTLGen Consortium) are shown in bold, and genes with the highest overall V2G scores (see material and methods) are underlined.

This variant is annotated in GTEx as an eQTL for *PROZ* on mammary tissue (note that, as of our analysis, no placental or pregnancy blood data were available on GTEx). To further support our findings of an association with genes related to coagulation factors, we queried for significant eQTL and pQTL associations in blood. Notably, we found that several of our preeclampsia-associated variants strongly regulate the expression of these genes, particularly of *PROZ*, including rs5960 (eQTL p = 2.8E−56) and rs553316 (eQTL p = 3.4E−93) (Tables S9 and S10). The global distribution of allele frequencies for rs5960 and rs553316 in 1KG populations are shown in Figure S13 and noted in Table S11.

Given the importance of clotting genes in pregnancy, we sought to complement the genotype analysis by performing deep sequencing of targeted genomic regions surrounding rs5960 in a subset of cohort participants (Tables S12 and S13, Figure S7). To fine-map potential causal variants, we repeated the same TDT analysis described above in the fine-mapped individuals and cross-referenced with the GTEx database for expression phenotypes. This analysis found an association of preeclampsia with several eQTLs for *PROZ* (Table S14). Other top hits from the SNP array-based TDT that were recapitulated in this sequencing-based analysis include variants in *SLC46A3* (MIM: 616764) and *CUL4A* (MIM: 603137), also located on chromosome 13 (Table S14). Both genes have been previously associated with preeclampsia risk in clinical studies.[9,90] These data suggest that clotting factors on chromosome 13 may play an important role in preeclamptic pregnancies.

Next, we asked whether this *PROZ* eQTL resulted in differential PROZ protein expression between PRE-affected individuals and PUN control subjects. Since the TDT identifies associated variants in the offspring, we analyzed the umbilical cord plasma of 8 PUN control subjects and 16 PRE-affected individuals by ELISA. In this limited sample, we detected no difference of PROZ protein levels in umbilical cord plasma collected after delivery (difference in means = 41.550 μg/mL, 95% CI −342.758 to 425.858, p = 0.85; Table S15, Figure S14). We note that this result may reflect our previous finding in which the *PROZ* locus reached a genome-wide suggestive, but not significant, association with preeclampsia, likely due to sample size limitations. However, it is possible that variants in *PROZ* may be involved in determining risk of preeclampsia through mechanisms that are not measurable by blood protein levels, such as molecular interactions. Future testing could evaluate PROZ, F10, and F7 levels in the placenta, where interaction with the maternal environment is more significant to the preeclampsia phenotype than in umbilical cord blood.

Selection scan reveals signal around *ADAM12* on chromosome 10 in case founders

We investigated whether our top hits in the TDT analysis were found in genetic regions under positive selection by computing genome-wide integrated haplotype scores (iHS)[65] in the PRE case founders. We identified 1,416

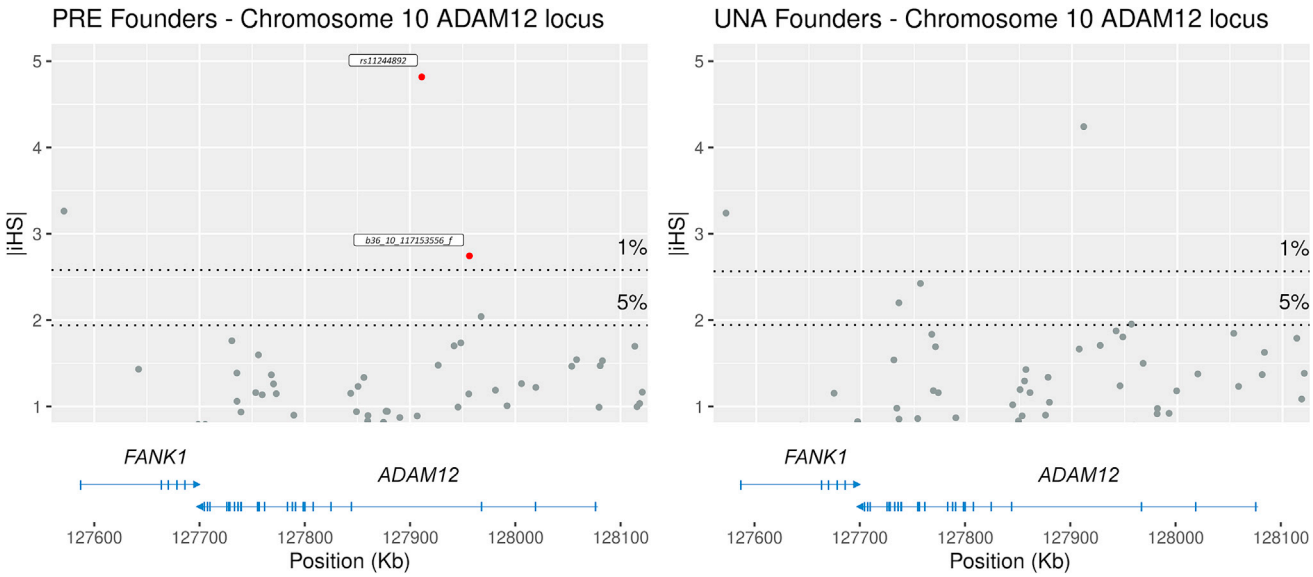

Figure 3. iHS hits across *ADAM12* and its closest neighbor, *FANK1*, on chromosome 10
Genetic positions, exons (vertical bars), introns (horizontal bars), and the direction of transcription (arrows) are shown in the panel below. A signal of selection encompassed by two SNPs (red dots) falls in an intronic region on *ADAM12* in case founders (PRE). This signal is weaker in control individuals (UNA).

genomic candidate regions in the top 1% (99th percentile) of the iHS distribution. None of the signals in this percentile were associated with the top TDT candidate region on chromosome 13 where *PROZ, F7, F10, SLC46A3*, and *CUL4A* are located (Table S16; Figures S15 and S16). Our selection scan revealed the top three highest genome-wide iHS values among affected individuals were in *AP1S1* (MIM: 603531) (top SNP KG_7_100587025, |iHS| = 5.32) and *SERPINE1* (MIM: 607378) (top SNP rs1050955, |iHS| = 4.91), both on chromosome 7; and on the *ADAM12* (MIM: 602714) (top SNP rs11244892, |iHS| = 4.82) gene locus chromosome 10. *ADAM12* has been previously associated with pregnancy disorders including ectopic pregnancy and preeclampsia.[91] The encoded metalloprotease ADAM12 has been proposed as a biomarker for preeclampsia[92,93] and according to GTEx and UniProt/SwissProt mRNA expression profiling is overexpressed in ovary, placenta, and muscle. When we repeated the same selection scan in the UNA population control subjects, the *ADAM12* signal is weaker as only one SNP in the region passes the top 1% threshold (Figure 3). The global distribution of allele frequencies for all three top iHS hits in 1KG populations is shown in Figure S17. Interestingly, the minor alleles for the top hit SNPs on *AP1S1* and *SERPINE1* are found at high frequencies in Peruvians (PEL MAF 0.74) compared to other 1KG populations (global 1KG MAF 0.34).

Clotting factor locus shows paternal inheritance

We next examined whether there were loci associated with preeclampsia that were disproportionately inherited either maternally or paternally. To this end, we performed a parent-of-origin TDT GWAS in the same 88 trios tested above. This test investigates whether any of the associated SNPs are disproportionately inherited from fathers versus mothers, and vice versa. The most significant SNP from the TDT analysis, rs5960 in *F10*, is suggested to be paternally inherited more often than expected by chance under the 5% permutation-based significance threshold (p = 1.38E−04, Figure 2, Table 4, Figure S18). Another locus showing suggestive evidence of paternal inheritance is rs79278805 (p = 1.77E−04), located within *SPAG6* (MIM: 605730) on chromosome 10. Similarly, we find suggestive evidence of maternal origin bias for locus rs130121 (p = 1.91E−04) on chromosome 22 in *FAM19A5/TAFA5* (MIM: 617499) . Both variants are suggestive based on the 5% cutoff calculated from the permutation analysis (Table 4, Figures S19–S21). At least one gene in the vicinity of these SNPs has been implicated in reproduction. *SPAG6* is recognized by anti-sperm antibodies and might be involved in infertility.[94,95] Overall, these parent-of-origin effects support the hypothesis that maternal and/or paternal bias might contribute to the development of preeclampsia.

Case-control analysis, placental gene *S100P* is associated with preeclampsia in the offspring

While the TDT identifies preeclampsia risk variants from inheritance analysis, a more common way to test for disease risk variants is to compare affected individuals and control subjects. The collection of control (PUN) mother-offspring duos allowed us to compare preeclamptic to healthy pregnancies in both the mothers and the offspring. To this end, we performed two case-control GWASs of preeclampsia (see material and methods): (1) 254 PRE-affected vs. 70 PUN control mothers and (2) 225 PRE-affected and 60 PUN control offspring. Several genetic regions showed suggestive association with preeclampsia in both test groups (Table S17; Figures S22 and S23). The

Table 4. GWAS p values and genomic annotations of top hits (p < 5E−04) from the TDT-POO

Chr	BP	cytoBand	rsID	Ref	Alt	P	Function	Genes in region	Puno MAF
Paternal TDT-POO									
13	113801737	13q34	rs5960	C	T	1.38E-4	exonic	F10	0.4929
10	22686205	10p12.2	rs79278805	G	A	1.77E-4	intronic	SPAG6	0.05499
6	142668901	6q24.1	rs9399401	C	T	2.76E-4	intronic	**ADGRG6**	0.4575
11	115757874	11q23.3	rs4938220	C	T	3.86E−4	intergenic	**LINC00900,** LOC101929011	0.3585
11	115761165	11q23.3	rs639053	C	T	4.99E−4	intergenic	LINC00900, LOC101929011	0.3344
Maternal TDT-POO									
22	49095071	22q13.32	rs130121	G	A	1.91E−4	intronic	FAM19A5	0.2912
8	98411402	8q22.1	rs10282765	C	T	2.39E−4	ncRNA_intronic	LOC101927066	0.1194
8	98428772	8q22.1	rs2331465	A	G	2.39E−4	ncRNA_intronic	LOC101927066	0.122
22	49099888	22q13.32	rs4925446	C	T	3.86E−4	intronic	FAM19A5	0.2827
8	98432618	8q22.1	rs4588816	C	T	3.93E−4	ncRNA_intronic	LOC101927066	0.122

eGenes (based on eQTL analyses using blood-derived expression through the eQTLGen Consortium) are shown in bold, and genes with the highest overall V2G scores (see material and methods) are underlined.

most interesting association was the top SNP in the offspring, rs34360485 on chromosome 4 (p = 5.9E−06, OR 4.379, 95% CI 2.311 to 8.295, global 1KG MAF 0.166; Table 5), which contains the placental gene *S100P* (MIM: 600614). *S100P* is a calcium-binding protein strongly expressed in the placenta[96] that promotes trophoblast proliferation in culture.[97]

To further investigate these associations, we repeated these two case-control GWAS analyses using an additional set of 551 individuals from the CANDELA consortium as external control subjects (see supplemental material and methods). With an increased sample size for the controls, the top SNP rs34360485 in the offspring case-control GWAS reached genome-wide suggestive significance (p = 1E−05, OR 2.755, 95% CI 1.757 to 4.321; Figure S24). We were unable to include any additional affected individuals, however, which could have provided further power to our analyses. The global distribution of allele frequencies for rs34360485 in 1KG populations is shown in Figure S25 and noted in Table S11.

Associations of secondary phenotypes reveal loci with roles in placental biology

Preeclampsia is a heterogeneous disease with varying potential markers of severity. For instance, the earlier in gestation preeclampsia occurs, the more severe it is considered to be.[98,99] Likewise, all the characteristic clinical features associated with preeclampsia (such as proteinuria and elevated blood pressure) can present at varying levels of severity. Harnessing the availability of clinical records for all individuals in the PRE cohort, we next performed GWAS tests on six secondary phenotypes of preeclampsia measured at the time of diagnosis: (1) gestational age, maternal measurement; (2) gestational age, fetal measurement; (3) diastolic blood pressure; (4) systolic blood pressure; (5) proteinuria; and (6) severity of diagnosis as stated by the clinician. It is worth clarifying that gestational age (the time of the fetus in the womb) was measured in two different ways throughout the study. The fetal measurement was done by the "Capurro" test,[51,52] which combines five different measurements in the neonate, while the maternal measurement relies on the date of the mother's last menstrual period before pregnancy.

To investigate possible genetic associations with secondary phenotypes of preeclampsia, we performed GWAS analyses by logistic and linear regression for each of the six phenotypes in 254 mothers and 225 offspring, separately. In total, we ran 12 GWAS tests. Logistic regression was applied to binary phenotypes (proteinuria and severity of diagnosis), while linear regression was applied to continuous phenotypes (gestational age and blood pressure measurements). All analyses were corrected for batch, plus the first three PCs and maternal age were included as continuous covariates. With this analysis we found several suggestive associations of SNPs to secondary maternal phenotypes (Table 5; Tables S18–S20). These findings point to several genetic regions containing relevant genes associated with pregnancy and the complex biology of preeclampsia, as detailed below. Due to the low sample size of this unique dataset, only two of these associations reached standard genome-wide significance (p = 5E−08).

Gestational age

QA Gestational age was associated in mothers with one locus on chromosome 1 (rs952593, beta −1.621, 95% CI −2.241 to −1.001, p = 3.6E−07, global 1KG MAF 0.283). This region is near *TBX15* (Table 5; Table S18; Figure S26–S29), a t-box transcription factor shown to be downregulated in intrauterine growth restricted placentas.[100] The association held true with both measurements of gestational age (maternal last

Table 5. Statistics and annotations of the top SNPs (p < 5E−04) with biological relevance for preeclampsia of secondary phenotype and case-control GWAS analyses

							GWAS stats					
GWAS	**Chr**	**BP**	**cytoBand**	**rsID**	**Ref**	**Alt**	**BETA/ OR**	**95% CI**	**P**	**Function**	**Genes in region**	**Puno MAF**
GA, fetal measurement, maternal genome	1	119404210	1p12	rs952593	T	C	−1.621	−2.241 to −1.001	6.00E−7	intergenic	SPAG17, TBX15	0.13
	19	34438859	19q13.11	rs16960768*	C	T	−1.805	−2.426 to −1.184	3.49E−8	intergenic/ regulatory	none reported	0.12
GA, maternal measurement, maternal genome	1	119425100	1p12	–	A	C	−1.428	−2.003 to −0.8532	2.02E−6	downstream	TBX15	0.14
	11	57203942	11q12.1	rs2581927	C	T	−1.969	−2.826 to −1.112	1.05E−5	intergenic	SLC43A3, RTN4RL2	0.06
	19	34438859	19q13.11	rs16960768	C	T	−1.472	−2.039 to −0.905	7.21E−7	intergenic/ regulatory	none reported	0.1225
Diastolic BP, maternal genome	4	85200613	4q21.23	rs1874237*	G	A	−4.228	−5.676 to −2.78	3.03E−8	ncRNA intronic	LOC101928978	0.45
Severity, maternal genome	11	94360812	11q21	rs1940640	T	G	2.546	1.701–3.81	5.57E−6	ncRNA intronic	LOC105369438	0.43
Proteinuria, maternal genome	17	2028106	17p13.3	rs2760751	A	G	2.795	1.784–4.38	7.30E−6	intronic	SMG6	0.29
	11	94356914	11q21	rs12276362	C	T	0.3856	0.2574–0.5779	3.898E−6	ncRNA intronic	LOC105369438	0.49
Systolic BP, fetal genome	2	18099832	2p24.2	rs4553827	C	T	7.76	4.971–10.55	1.35E−7	intronic	KCNS3	0.25
Proteinuria, fetal genome	4	5341148	4p16.2	rs62297274	C	T	0.3483	0.2238"0.542	2.94E−6	intronic	STK32B	0.49
	3	25052754	3p24.2	rs4241542	C	T	0.2637	0.1467–0.4742	8.48E−6	intronic	RARB	0.21
Case-control, offspring	4	6671568	4p16.1	rs34360485	A	G	4.379	2.311–8.295	5.90E−6	downstream	LINC02482, **S100P**	0.36

All SNPs in this table are described in the text (for a complete list of regions at p < 5E−4, see supplemental tables). Beta values are reported for linear regressions and odds ratio (OR) for logistic regressions. GA, gestational age; BP, blood pressure. eGenes (based on eQTL analyses using blood-derived expression through the eQTLGen Consortium) are shown in bold, and genes with the highest overall V2G scores (see material and methods) are underlined. Genome-wide significant hits noted with asterisk (*).

period and neonate Capurro test). The maternal measurement, but not the fetal measurement, of gestational age was associated with a multigenic locus on chromosome 11 (top SNP rs2581927, beta −1.969, 95% CI −2.826 to −1.112, p = 1.05E−05, global 1KG MAF 0.109). A gene of interest in this locus is *APLNR* (MIM: 600052), the receptor to *ELABELA* (MIM: 615594), which causes preeclampsia-like symptoms in mice[101] (Figures S30 and S31). Finally, both measurements of gestational age were associated in mothers with an intergenic regulatory region on chromosome 19 upstream of *KCTD15*. This association reached genome-wide significance for the fetal measurement only (top SNP rs16960768, beta −1.805, 95% CI −2.426 to −1.184, p = 3.49E−08; Table S18; Figures S32 and S33). As of this writing, rs16960768 has no annotations on GTEx, no functional consequences noted in ANNOVAR, and no clinical significance reported in ClinVar.

Diastolic and systolic blood pressure

Diastolic blood pressure reached genome-wide significance for one association in the maternal genome on chromosome 4 (top SNP rs1874237, beta −4.228, 95% CI −5.676 to −2.78, p = 3.03E−08, global 1KG MAF 0.379; Table 5; Figure 4). This SNP is within an uncharacterized non-coding RNA locus near *NKX6-1* (MIM: 602563), a gene involved in β-cell development and function.[102] In the offspring, both systolic and diastolic blood pressure were associated with SNPs in *KCNS3/K(V)9.3* (MIM: 603888) (top SNP rs4553827, beta 7.76, 95% CI 4.971 to 10.55, p = 1.35E−07, global 1KG MAF 0.472), a voltage-gated potassium channel gene that is highly expressed in the human placenta, where it localizes to placental vascular tissues and syncytiotrophoblast cells[103] (Table S19; Figures S34–S37).

Proteinuria and severity of diagnosis

Proteinuria was associated in the mothers with rs2760751 on chromosome 17 (OR 0.3856, 95% CI 0.2574 to 0.5779, p = 3.89E−06, global 1KG MAF 0.319). This SNP is intronic to *SMG6* (MIM: 610963), a telomerase binding protein. A second association with proteinuria in the maternal genome was found with SNP rs12276362 (OR 0.3856, 95% CI 0.2574 to 0.5779, p = 3.89E−06, global 1KG MAF 0.297) in chromosome 11, by *PIWIL4* (MIM: 610315) (Figures S38–S41). This region is also correlated with severity of diagnosis in the mothers (rs1940640, OR 2.546, 95% CI 1.701 to 3.81, p = 5.57E−06, global 1KG MAF 0.216). It is

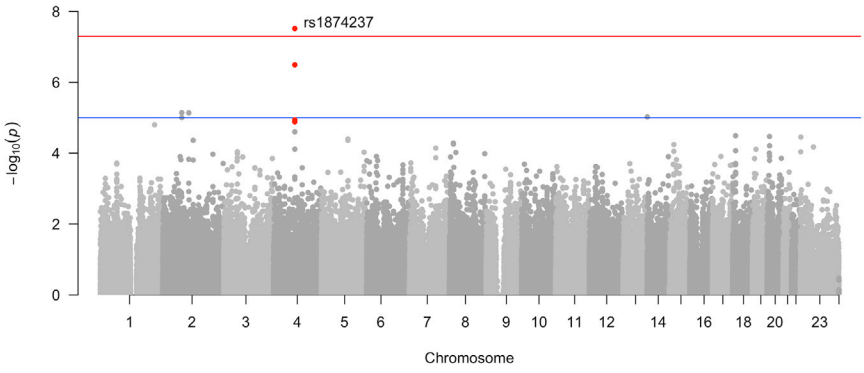

Figure 4. Manhattan plot showing top association in the maternal genome with diastolic blood pressure

SNP rs1874237 on chromosome 4 at p = 3.03E−08 shown in red. Horizonal lines indicate standard genome-wide significance (p = 5E−08 in red) and suggestive thresholds (p = 1E−05) commonly used in the literature.

not surprising that proteinuria and severity of diagnosis share a common association, since these two phenotypes are correlated—clinically severe cases generally have higher levels of protein in the urine. Aberrant PIWI proteins, which interact with pi-RNAs to drive post-transcriptional gene regulation, have been found in cancers,[104] and theoretical evidence from piRNA evolution suggests a role in placentation, although this has yet to be proven empirically.[105] In the offspring genome, proteinuria showed an association with placental gene *RARB* (MIM: 180220) or retinoic acid receptor beta (rs4241542, OR = 0.3745, 95% CI 0.1467 to 0.4742, p = 8.48E−08, global 1KG MAF 0.400),[106,107] and with an SNP intronic to *STK32B* (MIM: 119530) (rs62297274; OR 0.3483, 95% CI 0.2238 to 0.542, p = 2.94E−06, global 1KG MAF 0.037; Table S19; Figure S42–S44). Interestingly, the minor allele for SNP rs62297274 is found at high frequencies in Peruvians compared to other global populations. In the Puno cohort MAF for this variant is 0.49, slightly higher than among Peruvians from Lima sampled in the 1KG (PEL MAF 0.41) (Figure S45). In contrast, the minor allele is found at low frequencies in the rest of the Americas (1KG AMR MAF 0.19) and is rarely observed globally (global 1KG MAF <0.05) (Table S11).

Discussion

In this analysis, we investigate the genetic variation of a preeclampsia cohort of Andean families from Puno, Peru; a high-altitude population with one of the highest incidences of this disease in the world.[24,43] By characterizing patterns of genetic diversity and population structure in this cohort we found that study participants have high proportions of Native American ancestry and carry an Andean genetic component which is not shared with other admixed populations from other parts of Latin America. We harnessed the power of a trio study design to uncover maternal, paternal, and fetal genetic factors influencing the incidence and severity of preeclampsia in this cohort. In contrast to previous preeclampsia genome-wide association studies, which have been hampered by limited phenotyping and heterogeneous sampling,[48] the present work includes a case-control cohort sampled from a single population of Native American origin, treated at the same hospital, and exposed to similar selective pressures due to long-term residence at high altitude. Thus, despite the limited sample size, our family-based GWAS design circumvents potential confounding issues with population structure and permits identification of significant and suggestive associations with preeclampsia that would remain otherwise undiscovered.[42]

Most genetic studies on preeclampsia have not investigated whole family units,[9,17,18,108] despite evidence of a complex genetic risk profile involving contributions from both parents and the fetus.[3] For instance, higher rates of preeclampsia have been documented among families where either the mother or father were the products of a preeclamptic pregnancy,[11,109,110] and there is evidence that primiparous women with a recent partner change are at higher risk of developing the disorder.[111–113] Results from case-control studies involving parental couples suggest that genetic and immunological factors from fathers may also play an important role in mediating preeclampsia risk.[16] A study of Malay preeclamptic families found that paternal HLA-G factors significantly increased risk for preeclampsia in multigravida pregnancies.[114] A subsequent case-control study investigating preeclamptic mothers and their male partners in northern Mexico found that paternal variants at the rs5370 polymorphism in *EDN1* (MIM: 131240) decrease the risk of preeclampsia.[115] There is also mounting evidence that the offspring genome plays an important role, as demonstrated by several recent studies which have identified fetal variants[9,116] and fetal-maternal complotypes[117] influencing preeclampsia risk. Lastly, trio studies have also become an important tool for differentiating between maternal, paternal, and fetal contributions to preeclampsia risk[10,118,119] and for elucidating the role of parental imprinting in the expression of placental genes and associated risk of pregnancy complications.[120] Taken together, the evidence gathered by this large body of work reinforces the strength of our trio-sampling approach to investigating preeclampsia in the Puno cohort.

The top association with preeclampsia identified in our trio study was rs5960, a variant in the clotting factor gene *F10*, in a locus with two other clotting factors: *PROZ* and *F7*. *PROZ*, a vitamin K-dependent factor, is an anticoagulant protein with a role in factor X inhibition.[121] Several previous studies have suggested a hypercoagulative

state in preeclampsia (reviewed in Khadijah Ismail and Higgins[122]), as spiral arteries of preeclamptic pregnancies often present thrombosis and atherosis.[123] In fact, strong evidence supporting an effect of thrombotic processes on preeclampsia is based on the observation that aspirin, a known blood thinner, successfully delays preeclampsia onset.[124]

Low PROZ levels are associated with thrombotic disorders, and many adverse pregnancy outcomes have also been linked with maternal PROZ levels.[121] A small, prospective case-control study found low PROZ levels associated with intrauterine growth restriction (IUGR) and intrauterine fetal demise, but not preeclampsia.[125] In contrast, a larger cross-sectional study found lower median levels of PROZ in preeclampsia outcomes but not IUGR or fetal demise.[126] One study found a correlation between lower PROZ levels and severity of HELLP syndrome (MIM: 614985), a complication of preeclampsia that stands for hemolysis, elevated liver enzymes, and low platelets and which occurs in 10%–20% of preeclamptic pregnancies.[123,127] However, no study on PROZ or other clotting factors in preeclampsia has been successfully replicated, likely due to the extreme heterogeneity of the disease and the mix of populations studied.

Although previous studies on *PROZ* have focused on the mother's genome,[126,128] ours suggests a correlation between the fetal *PROZ/F7/F10* locus on chromosome 13 and preeclampsia. In a subset of our sample, we found no differences in protein plasma levels of PROZ between preeclamptic and healthy pregnancies in the mother or the offspring. However, this analysis was limited by small sample size and post-natal blood sampling. As blood samples were only collected immediately after birth, we were unable to monitor changes in PROZ protein levels throughout the pregnancy or in other relevant tissues. Further longitudinal studies could analyze several clotting factor levels and activity in this pregnant population to assess the impact of thrombosis in preeclampsia risk among Andean highlanders.

Expanding the TDT to a parent-of-origin analysis (POO), we found several associations to genetic regions with suggested paternal inheritance. For instance, the top TDT hit on *F10*, rs5960, is also the locus with the strongest paternal origin effect in the TDT-POO. Although future research examining variation at the *PROZ/F7/F10* region in a larger population will be needed to confirm this finding, our results are of interest to studies investigating the role of paternal genetic factors, genomic imprinting, and paternal-offspring conflict in preeclampsia and other pregnancy disorders.[13,16,120,129–131] Another top region in the TDT-POO includes the biologically relevant gene *SPAG6* previously described as being involved in infertility and the immune system.[94,95] Future work could investigate the potential role of this candidate gene in the maternal-fetal interface and its potential involvement in the pathophysiology of preeclampsia.

We also found several placental genes associated with secondary phenotypes that underline the severity of preeclampsia, such as hypertension, gestational age, and proteinuria. Differential expression of these genes may contribute to the insufficiency of placental development in early pregnancy that leads to hypertension and proteinuria in the third trimester. Some of our suggestive associations are near genes previously shown to have roles in pregnancy, vascular processes, and even preeclampsia. One such gene is *APLNR*, the receptor to ELABELA, which causes preeclampsia-like symptoms in mice[101] and is found expressed in lower levels in the serum and placentas of some women with late-onset, but not early-onset preeclampsia.[132] However, this gene is in a multigenic locus, and fine-mapping approaches with functional studies are required to discover the effect of this locus in our cohort.

Our study is one of few preeclampsia GWASs to include the offspring genome. One recent study with a large cohort found a gene, *sFLT1* (MIM: 165070), associated with late (but not early) preeclampsia,[9,116] suggesting that dysregulation of genes in the fetal genome contribute to preeclampsia. In our study, we found several associations with preeclampsia and its severity phenotypes in the fetal genome. For instance, we found an association between severity of hypertension (systolic and diastolic pressure measurements) and *KCNS3/K(V)9.3,* a gene that is highly expressed in the human placenta, where it localizes to placental vascular tissues and syncytiotrophoblast cells.[103] We also found an association of the retinoic acid (RA) signaling gene *RARB* and severity of proteinuria in the preeclamptic fetal genome. RA signaling is essential for healthy placental and fetal development in animal models, with evidence of similar requirement in humans (reviewed in Comptour et al.[106]). RARB is expressed in the extravillous part of the placenta and its activation induces *RARRES* (MIM: 605090), shown to be overexpressed in preeclamptic placentas.[107] Our study adds to this body of literature and highlights the role of RA in proper placentation. Lastly, the most interesting region in the offspring genome was identified in our case-control analysis: the *S100P* gene, a calcium-binding protein strongly expressed in the placenta[96] that promotes trophoblast proliferation in culture.[97] This finding suggests that fetal biology, and specifically placental development driven by fetal genes, highly contributes to the pathology of preeclampsia.

We examined the global distribution of allele frequencies for each of the candidate associated SNPs detailed above. Most alleles were shared among several global populations (see global distribution plots in supplemental information). A notable exception is SNP rs62297274, an intronic variant located in *STK32B* which is associated with proteinuria in the offspring genome. The minor allele reaches its highest global frequency in Peruvian populations (Figure S45). As of this writing, rs62297274 has no reported clinical significance in dbSNP or ANNOVAR. However, intronic variants are known to have functional impacts on RNA splicing patterns.[133] To elucidate the

functional significance of this variant, future research could evaluate its pathogenic potential in Peruvian populations.[134,135]

We further investigated whether signals of selection are present in the regions surrounding candidate preeclampsia-associated genes in this high-altitude-adapted cohort. We found association of the top three candidate signals with the highest iHS score to the genes *AP1S1, SERPINE1*, and *ADAM12*. *AP1S1*, located on chromosome 7, is involved in clathrin coat-assembly and trafficking between the trans-Golgi network, endosomes, and the plasma membrane.[136] *SERPINE1*, also on chromosome 7, encodes the protein plasminogen activator inhibitor 1 (PAI-1) and serves as the primary inhibitor of tissue and urinary plasminogen activators, constituting an important regulatory protein in fibrinolysis (breakdown of blood clots).[137] *ADAM12,* located in chromosome 10, is a member of the A Disentegrin and Metaloproteinase family of proteases and serves as a promoter of throboblast differentiation in placental development.[138] The mechanisms that regulate *ADAM12* are not well characterized, but it has been shown that Notch signaling, activated by hypoxia, can increase the levels and activity of *ADAM12* and lead to increased shedding of the epidermal growth factor (EGF).[139] *ADAM12* has also been described as playing an important role in placental development and function.[140,141] In preeclampsia, the role of *ADAM12* remains elusive and controversial since its expression levels have been positively and inversely correlated with the condition.[142–144] These results could suggest an ongoing selection sweep in genes involved in cell transport, blood clotting, and preeclampsia, particular to the cohort of preeclamptic parents. However, we note that a potential limitation of this analysis is that iHS, as a haplotype-based test,[65] can detect only relatively recent selection signals.[67,145] Thus, our analysis may not have sufficient power to detect more ancient patterns of adaptation. Additionally, this analysis did not test for other types of selective processes beyond selective sweeps, such as polygenic adaptation, which may also be acting on standing genetic variation in this population.

As discussed, several genes found in our analyses are involved in placental function. Interestingly, morphological studies comparing placentas from Andean-descent and European-descent individuals in Bolivia, at both low and high altitudes, describe differences in placental composition.[146,147] Highland placentas from individuals of both ancestries show more intervillous space but less villi, and the Andean highland placenta, compared to the European, have more trophoblast and villous stroma on average. Differences in placental morphology suggest an adaptive mechanism to the lower oxygen pressure at high altitude, but one that does not lower the risk of preeclampsia.

In conclusion, this study investigates a cohort of preeclamptic highland Andean families from Puno, Peru to elucidate the genetic basis of this pregnancy disorder at high altitudes. We generated genotype data at more than 400,000 positions across the genome and used these data to determine ancestry patterns and map associations between genetic variants and preeclampsia phenotypes. Our trio-based recruitment strategy, including genotype data from mothers, fathers, and offspring, allowed us to identify genetic regions not previously reported in preeclampsia genome-wide association studies. Specifically, we identified suggestive associations with several variants near genes involved with placental and blood vessel function, and therefore, of functional importance for human pregnancy biology. The strongest association hit involves a cluster of clotting factor genes on chromosome 13 including *PROZ*, *F7*, and *F10* in the fetal genome. This finding provides supporting evidence that coagulation plays an important role in the pathology of preeclampsia and potentially underlies other pregnancy disorders exacerbated at high altitude. Additionally, our findings provide a list of candidate genetic loci with suggestive associations for future replication and functional validation.

A major limitation of our study is the small sample size of the recruited cohort which resulted in reduced power for statistical analyses. Although GWASs typically require large sample sizes for robust identification of candidate genetic variants involved in complex or polygenic traits, it is known that functionally relevant alleles may segregate at higher frequencies in some populations due to microevolutionary processes such as genetic drift, a history of founder effects, genetic isolation, or local adaptation.[42,148,149] Many of the same processes have occurred throughout the evolutionary history of Andean populations.[34,86–88,150] Moreover, depending on the effect size of functional or medically relevant variants, we have found in our own previous research that sample sizes as low as a few hundred individuals are enough to detect associations in isolated populations.[151]

Another possible limitation is that, as noted in Table 1, most of the affected mothers in the recruited case trios were diagnosed with mild cases of preeclampsia. This bias in our cohort may contribute to our limited power as we would expect to see stronger association signals in analyses such as the TDT with more severe cases. As a family-based test, TDT uses data from a single cohort of affected trios and is therefore more robust against artifacts such as those caused by population substructure. Still, any statistical test will have some chance of false positives, even though we control for this stringently through permutations. We acknowledge that a larger sample size and especially additional recruitment of trio families would have likely improved our statistical power and ability to detect relevant genetic variation, in both the TDT and case-control analysis. While we tried to increase the sample size in the case-control analysis by including additional controls (CANDELA), we were unable to recruit additional affected individuals. Future efforts could focus on recruiting participants from a larger cohort that could also be stratified by different levels of preeclampsia severity and

explore interactions between blood pressure, proteinuria, and altitude, to discern whether certain genetic signatures impact the severity of preeclampsia.

A further limitation of our study is that blood pressure measurements were unavailable for many PUN control mothers (Table S3). In these instances, we used average basal levels (85/55 mmHg) as prior and relied upon the attending physician's preeclampsia diagnosis. Individuals with hypertension at admission but no diagnosis of preeclampsia were also excluded as control subjects. Therefore, while we have limited clinical data for this cohort, female participants were declared by the clinicians to be women in labor with normal blood pressure (i.e., not significantly above the baseline for this population), no proteinuria, and no other signs of organ damage. We acknowledge that physician-level variation at the time of diagnoses may introduce some variation in the diagnosis and categorization of preeclampsia severity. However, since the clinical phenotypes of PUN control mothers were not considered in most of our statistical analyses (e.g., TDT, TDT-POO, secondary phenotypes GWAS) we believe this limitation is unlikely to strongly impact our conclusions. We also note that since the present study was focused on pregnant mothers—the only member of the family triad affected by preeclampsia symptoms—we did not collect measurements of secondary phenotypes such as blood pressure or urine proteinuria levels from fathers or offspring. We note that proteinuria in urine can occur in other health conditions beyond preeclampsia, including some that impact both men and women, and disproportionately occur among high-altitude-resident populations, such as high-altitude renal syndrome (HARS).[152–154] However, since proteinuria is not regularly measured in non-pregnant members of the family triad throughout pregnancy or at the time of birth, we were unable to test for the presence of this condition in the present study.

Finally, we acknowledge that because our study is primarily based on analysis of SNP array data, we may be limited in our ability to detect causal alleles. However, we sought to address this limitation by complementing the genotype-based analyses with a targeted sequencing approach to fine-map genomic regions of interest in a subset of our samples. This follow-up analysis provided us with increased genomic coverage surrounding several top candidate regions, albeit with a smaller sample size. Future follow-up efforts could rely upon imputation to boost power or deep sequencing of candidate loci identified in this study.

Considering that this is a severely understudied population with a large disease burden, this work represents a preliminary and seminal effort toward exploring potential associations which can inform future validation and functional studies, while also highlighting the need for further research with such underrepresented ancestry groups. Studying diverse human populations with unique genetic adaptations enables identification of the primary genetic factors underlying complex phenotypes and gene function.[42] This research examined Andean populations as a model to understand human pregnancy biology in hypoxic conditions. This natural experimental setting provides a unique opportunity to understand the genetic factors influencing human reproductive fitness in challenging environments worldwide and to discover population-specific variants underlying biomedical traits. Our work also underscores the importance of including diverse populations in genome wide association studies and functional variant discovery efforts to better understand human physiology and disease globally.

Data and code availability

The datasets generated during this study are deposited in the European Genome-Phenome Archive (EGA): Study EGAS00001004625 and can be accessed with Data Access Committee approval.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.04.014.

Acknowledgments

We extend our deepest gratitude to the people of Puno, Peru who participated in this study at Hospital Regional Manuel Nuñez Butrón and Universidad del Altiplano. We are tremendously grateful to the Mendoza Revilla family, who provided lodging and logistics support in Lima during fieldwork seasons of the technical team. We thank Victor Acuña Alonzo from the Instituto Nacional de Antropología e Historia and Samuel Canizales from the Instituto Nacional de Medicina Genómica of Mexico for providing access to the CANDELA cohort data and Colleen Julian for providing helpful commentary on this manuscript. This work was supported in part by the National Science Foundation (NSF) SBE Postdoctoral Research Fellowship Award No. 1711982 awarded to M.A.N.C.; NSF Graduate Research Fellowship Program Grant No. DGE–1147470 awarded to K.M.B.R. (fellow no. 2014187481); an A.P. Giannini Foundation postdoctoral fellowship, a Stanford Child Health Research Institute postdoctoral award, and a Stanford Dean's Postdoctoral Fellowship awarded to E.T.Z.; the Chan Zuckerberg Biohub Investigator Award to C.D.B.; a Burroughs Welcome Prematurity Initiative Award to J.C.B.; Mexico CONACyT fellowship PNPC 003649 to R.G.B.; and the George Rosenkranz Prize for Health Care Research in Developing Countries, the International Center for Genetic Engineering and Biotechnology (ICGEB, Italy) grant CRP/MEX15-04_EC, and Mexico's CONACYT grant FONCICYT/50/2016, each awarded to A.M.E. Further funding was provided by the Sandler Family Foundation, the American Asthma Foundation, the RWJF Amos Medical Faculty Development Program, Harry Wm. and Diana V. Hind Distinguished Professor in Pharmaceutical Sciences II, National Institutes of Health, National Heart, Lung, and Blood Institute Awards R01HL117004, R01HL128439, R01HL135156, R01HL141992, National Institute of Environmental Health Sciences Awards R01ES015794, R21ES24844, the National Institute on Minority Health and Health Disparities Awards R01MD010443, and R56MD013312, and the National Human Genome Research Institute Award

U01HG009080, each awarded to E.B. In addition, E.B. reports grants from General Medical Sciences, the Tobacco-Related Disease Research Program, and the Food and Drug Administration.

Declaration of interests

J.W.C. is currently a full-time employee at Genentech, Inc. and hold stocks in Roche Holding AG.

Received: June 26, 2021
Accepted: April 25, 2022
Published: May 18, 2022

Web resources

EGA, https://ega-archive.org/
GGV, https://popgen.uchicago.edu/ggv/
GTEx, https://gtexportal.org/home/datasets
Online Mendelian Inheritance in Man, https://www.omim.org/
Open Targets Genetics, https://genetics.opentargets.org/
Snpflip, https://github.com/biocore-ntnu/snpflip

References

1. Duley, L. (2009). The global impact of pre-eclampsia and eclampsia. Semin. Perinatol. *33*, 130–137. https://doi.org/10.1053/j.semperi.2009.02.010.
2. Rana, S., Lemoine, E., Granger, J.P., and Karumanchi, S.A. (2019). Preeclampsia: pathophysiology, challenges, and perspectives. Circ. Res. *124*, 1094–1112. https://doi.org/10.1161/circresaha.118.313276.
3. Valenzuela, F.J., Perez-Sepulveda, A., Torres, M.J., Correa, P., Repetto, G.M., and Illanes, S.E. (2012). Pathogenesis of preeclampsia: the genetic component. J. Pregnancy *2012*, 1–8. https://doi.org/10.1155/2012/632732.
4. Osungbade, K.O., and Ige, O.K. (2011). Public health perspectives of preeclampsia in developing countries: implication for health system strengthening. J. Pregnancy *2011*, 1–6. https://doi.org/10.1155/2011/481095.
5. Phipps, E.A., Thadhani, R., Benzing, T., and Karumanchi, S.A. (2019). Pre-eclampsia: pathogenesis, novel diagnostics, and therapies. Nat. Rev. Nephrol. *15*, 275–289. https://doi.org/10.1038/s41581-019-0119-6.
6. Burton, G.J., and Fowden, A.L. (2015). The placenta: a multifaceted, transient organ. Phil. Trans. R. Soc. B *370*, 20140066. https://doi.org/10.1098/rstb.2014.0066.
7. Yong, H.E.J., Murthi, P., Brennecke, S.P., and Moses, E.K. (2018). Genetic approaches in preeclampsia. Methods Mol. Biol. *1710*, 53–72. https://doi.org/10.1007/978-1-4939-7498-6_5.
8. American College of Obstetricians and Gynecologists. (2013). Hypertension in pregnancy. Report of the American college of obstetricians and gynecologists' task force on hypertension in pregnancy. Obstet. Gynecol. *122*, 1122–1131. https://doi.org/10.1097/01.AOG.0000437382.03963.88.
9. McGinnis, R., Steinthorsdottir, V., Williams, N.O., Thorleifsson, G., Shooter, S., Hjartardottir, S., Bumpstead, S., Stefansdottir, L., Hildyard, L., Sigurdsson, J.K., et al. (2017). Variants in the fetal genome near FLT1 are associated with risk of preeclampsia. Nat. Genet. *49*, 1255–1260. https://doi.org/10.1038/ng.3895.
10. Pappa, K.I., Roubelakis, M., Vlachos, G., Marinopoulos, S., Zissou, A., Anagnou, N.P., and Antsaklis, A. (2011). Variable effects of maternal and paternal-fetal contribution to the risk for preeclampsia combining GSTP1, eNOS, and LPL gene polymorphisms. J. Matern. Fetal Neonatal. Med. *24*, 628–635. https://doi.org/10.3109/14767058.2010.511351.
11. Esplin, M.S., Fausett, M.B., Fraser, A., Kerber, R., Mineau, G., and Varner, M.W. (2001). Paternal and maternal components of the predisposition to preeclampsia. New Engl. J. Med. *344*, 867–872. https://doi.org/10.1056/nejm200103223441201.
12. Cnattingius, S., Reilly, M., Pawitan, Y., and Lichtenstein, P. (2004). Maternal and fetal genetic factors account for most of familial aggregation of preeclampsia: a population-based Swedish cohort study. Am. J. Med. Genet. *130A*, 365–371. https://doi.org/10.1002/ajmg.a.30257.
13. Wikström, A.K., Gunnarsdottir, J., and Cnattingius, S. (2012). The paternal role in pre-eclampsia and giving birth to a small for gestational age infant; a population-based cohort study. BMJ Open *2*, e001178. https://doi.org/10.1136/bmjopen-2012-001178.
14. Lie, R.T., Rasmussen, S., Brunborg, H., Gjessing, H.K., Lie-Nielsen, E., and Irgens, L.M. (1998). Fetal and maternal contributions to risk of pre-eclampsia: population based study. BMJ. *316*, 1343–1347. https://doi.org/10.1136/bmj.316.7141.1343.
15. Zusterzeel, P.L.M. (2002). Paternal contribution to the risk for pre-eclampsia. J. Med. Genet. *39*, 44–45. https://doi.org/10.1136/jmg.39.1.44.
16. Galaviz-Hernandez, C., Sosa-Macias, M., Teran, E., Garcia-Ortiz, J.E., and Lazalde-Ramos, B.P. (2018). Paternal determinants in preeclampsia. Front. Physiol. *9*, 1870. https://doi.org/10.3389/fphys.2018.01870.
17. Boyd, H.A., Tahir, H., Wohlfahrt, J., and Melbye, M. (2013). Associations of personal and family preeclampsia history with the risk of early-intermediate- and late-onset preeclampsia. Am. J. Epidemiol. *178*, 1611–1619. https://doi.org/10.1093/aje/kwt189.
18. Cincotta, R.B., and Brennecke, S.P. (1998). Family history of pre-eclampsia as a predictor for pre-eclampsia in primigravidas. Int. J. Gynaecol. Obstet. *60*, 23–27. https://doi.org/10.1016/s0020-7292(97)00241-5.
19. Silva, L.M., Coolman, M., Steegers, E.A., Jaddoe, V.W., Moll, H.A., Hofman, A., Mackenbach, J.P., and Raat, H. (2008). Low socioeconomic status is a risk factor for preeclampsia: the Generation R Study. J. Hypertens. *26*, 1200–1208. https://doi.org/10.1097/hjh.0b013e3282fcc36e.
20. Zamudio, S. (2007). High-altitude hypoxia and preeclampsia. Front. Biosci. *12*, 2967–2977. https://doi.org/10.2741/2286.
21. Moore, L.G., Charles, S.M., and Julian, C.G. (2011). Humans at high altitude: hypoxia and fetal growth. Respir. Physiol. Neurobiol. *178*, 181–190. https://doi.org/10.1016/j.resp.2011.04.017.
22. Keyes, L.E., Armaza, F.J., Niermeyer, S., Vargas, E., Young, D.A., and Moore, L.G. (2003). Intrauterine growth restriction, preeclampsia, and intrauterine mortality at high altitude in Bolivia. Pediatr. Res. *54*, 20–25. https://doi.org/10.1203/01.pdr.0000069846.64389.dc.
23. Abalos, E., Cuesta, C., Grosso, A.L., Chou, D., and Say, L. (2013). Global and regional estimates of preeclampsia and eclampsia: a systematic review. Eur. J. Obstet. Gynecol.

Reprod. Biol. *170*, 1–7. https://doi.org/10.1016/j.ejogrb.2013.05.005.

24. Gil Cipirán, F. (2017). Situación epidemiológica de la mortalidad materna en el Perú Boletín Epidemiológico del Perú (Centro Nacional de Epidemiología, Prevención y Control de Enfermedades, Ministerio de Salud), pp. 1514–1516.
25. Guevara Ríos, E., and Meza Santibáñez, L. (2014). Manejo de la preeclampsia/eclampsia en el Perú. Rev. Peru. Ginecol. Obstet. *60*, 385–393. https://doi.org/10.31403/rpgo.v60i163.
26. Moore, L.G. (2021). Hypoxia and reproductive health: reproductive challenges at high altitude: fertility, pregnancy and neonatal well-being. Reproduction *161*, F81–F90. https://doi.org/10.1530/rep-20-0349.
27. Palmer, S.K., Moore, L.G., Young, D.A., Cregger, B., Berman, J.C., and Zamudio, S. (1999). Altered blood pressure course during normal pregnancy and increased preeclampsia at high altitude (3100 meters) in Colorado. Am. J. Obstet. Gynecol. *180*, 1161–1168. https://doi.org/10.1016/s0002-9378(99)70611-3.
28. Bailey, B., Euser, A.G., Bol, K.A., Julian, C.G., and Moore, L.G. (2020). High-altitude residence alters blood-pressure course and increases hypertensive disorders of pregnancy. J. Maternal Fetal Neonatal Med. *35*, 1264–1271. https://doi.org/10.1080/14767058.2020.1745181.
29. Moore, L.G., Hershey, D.W., Jahnigen, D., and Bowes, W., Jr. (1982). The incidence of pregnancy-induced hypertension is increased among Colorado residents at high altitude. Am. J. Obstet. Gynecol. *144*, 423–429. https://doi.org/10.1016/0002-9378(82)90248-4.
30. Moore, L.G., Shriver, M., Bemis, L., Hickler, B., Wilson, M., Brutsaert, T., Parra, E., and Vargas, E. (2004). Maternal adaptation to high-altitude pregnancy: an experiment of nature–a review. Placenta *25*, 60–71. https://doi.org/10.1016/j.placenta.2004.01.008.
31. Tissot van Patot, M.C., Murray, A.J., Beckey, V., Cindrova-Davies, T., Johns, J., Zwerdlinger, L., Jauniaux, E., Burton, G.J., and Serkova, N.J. (2009). Human placental metabolic adaptation to chronic hypoxia, high altitude: hypoxic preconditioning. Am. J. Physiol. Regul. Integr. Comp. Physiol. *298*, 166–172. https://doi.org/10.1152/ajpregu.00383.2009.
32. Bigham, A.W., Wilson, M.J., Julian, C.G., Kiyamu, M., Vargas, E., Leon-Velarde, F., Rivera-Chira, M., Rodriquez, C., Browne, V.A., Parra, E., et al. (2013). Andean and Tibetan patterns of adaptation to high altitude. Am. J. Hum. Biol. *25*, 190–197. https://doi.org/10.1002/ajhb.22358.
33. Gómez-Carballa, A., Pardo-Seco, J., Brandini, S., Achilli, A., Perego, U.A., Coble, M.D., Diegoli, T.M., Álvarez-Iglesias, V., Martinón-Torres, F., Olivieri, A., et al. (2018). The peopling of South America and the trans-Andean gene flow of the first settlers. Genome Res. *28*, 767–779. https://doi.org/10.1101/gr.234674.118.
34. Bigham, A.W. (2019). Natural selection and adaptation to extreme environments. In A Companion to Anthropological Genetics, D.H. O'Rourke, ed. (Wiley Blackwell), pp. 219–232.
35. Beall, C.M. (2014). Adaptation to high altitude: phenotypes and genotypes. Annu. Rev. Anthropol. *43*, 251–272. https://doi.org/10.1146/annurev-anthro-102313-030000.
36. Brutsaert, T.D., Kiyamu, M., Elias Revollendo, G., Isherwood, J.L., Lee, F.S., Rivera-Ch, M., Leon-Velarde, F., Ghosh, S., and Bigham, A.W. (2019). Association of *EGLN1* gene with high aerobic capacity of Peruvian Quechua at high altitude. Proc. Natl. Acad. Sci. U S A *116*, 24006–24011. https://doi.org/10.1073/pnas.1906171116.
37. Fehren-Schmitz, L., and Georges, L. (2016). Ancient DNA reveals selection acting on genes associated with hypoxia response in pre-Columbian Peruvian Highlanders in the last 8500 years. Sci. Rep. *6*, 23485. https://doi.org/10.1038/srep23485.
38. Vega, L.S. (2019). New blood pressure levels in Peruvian high altitude populations and the new North American high blood pressure guidelines. J. Cardiol. Curr. Res. *12*, 84–87. https://doi.org/10.15406/jccr.2019.12.00446.
39. Julian, C.G., Wilson, M.J., and Moore, L.G. (2009). Evolutionary adaptation to high altitude: a view from in utero. Am. J. Hum. Biol. *21*, 614–622. https://doi.org/10.1002/ajhb.20900.
40. Michita, R.T., Kaminski, V.d.L., and Chies, J.A.B. (2018). Genetic variants in preeclampsia: lessons from studies in Latin-American populations. Front. Physiol. *9*, 1771. https://doi.org/10.3389/fphys.2018.01771.
41. Bigham, A.W., and Lee, F.S. (2014). Human high-altitude adaptation: forward genetics meets the HIF pathway. Genes Dev. *28*, 2189–2204. https://doi.org/10.1101/gad.250167.114.
42. Tishkoff, S. (2015). Strength in small numbers. Science *349*, 1282–1283. https://doi.org/10.1126/science.aad0584.
43. Bristol, N. (2009). Dying to give birth: fighting maternal mortality in Peru. Health Aff. *28*, 997–1002. https://doi.org/10.1377/hlthaff.28.4.997.
44. Chirinos Cáceres, J. (1995). Incidencia y caracteristicas de la enfermedad hipertensiva en el embarazo: estudio retrospectivo a nivel del mar y en la altura. Acta Andina *4*, 25–34.
45. Castillo Apaza, Y.P. (2017). Factores de riesgo asociados con preeclampsia en gestantes atendidas en el Hospital Regional Manuel Nuñez Butrón en el periodo enero-diciembre 2017. Medical Surgeon (Universidad Nacional del Altiplano, Facultad de Medicina Humana).
46. Dirección Regional de Salud Puno; and Oficina de Epidemiología (2001). Análisis de la situación de salud del Departamento de Puno 2001 (Ministerio de Salud, Región de Puno)).
47. de Salud, Ministerio (2019). Análisis de situación de salud del Perú 2018 (Centro Nacional de Epidemiología, Prevención y Control de Enfermedades).
48. Williams, P.J., and Broughton Pipkin, F. (2011). The genetics of pre-eclampsia and other hypertensive disorders of pregnancy. Best Pract. Res. Clin. Obstet. Gynaecol. *25*, 405–417. https://doi.org/10.1016/j.bpobgyn.2011.02.007.
49. Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de Avila, F., et al. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. PLoS Genet. *10*, e1004572. https://doi.org/10.1371/journal.pgen.1004572.
50. Winchester, D.E., Agarwal, N., Burke, L., Bradley, S., Schember, T., and Schmalfuss, C. (2016). Physician-level variation in the diagnosis of myocardial infarction and the use of angiography among Veterans with elevated troponin. Mil. Med. Res. *3*, 22. https://doi.org/10.1186/s40779-016-0090-5.
51. Pereira, A.P.E., Dias, M.A.B., Bastos, M.H., da Gama, N.g.N., Leal, M.d.C., and Leal, M. do C. (2013). Determining gestational age for public health care users in Brazil: comparison

of methods and algorithm creation. BMC Res. Notes *6*, 60. https://doi.org/10.1186/1756-0500-6-60.

52. Capurro, H., Konichezky, S., Fonseca, D., and Caldeyro-Barcia, R. (1978). A simplified method for diagnosis of gestational age in the newborn infant. J. Pediatr. *93*, 120–122. https://doi.org/10.1016/s0022-3476(78)80621-0.
53. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.
54. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.
55. Francis, R.M. (2017). Pophelper: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. *17*, 27–32. https://doi.org/10.1111/1755-0998.12509.
56. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*. e190–2093. https://doi.org/10.1371/journal.pgen.0020190.
57. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer International Publisher).
58. R Core Team (2018). R: A Language and Environment for Statistical Computing.
59. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288. https://doi.org/10.1016/j.ajhg.2013.06.020.
60. Mazandu, G.K., Geza, E., Seuneu, M., and Chimusa, E.R. (2019). Orienting future trends in local ancestry deconvolution models to optimally decipher admixed individual genome variations. In Bioinformatics Tools for Detection and Clinical Interpretation of Genomic Variations, A. Samadikuchaksaraei and M. Seifi, eds. (IntechOpen)), 10.5772/intechopen.82764.
61. Galanter, J.M., Gignoux, C.R., Torgerson, D.G., Roth, L.A., Eng, C., Oh, S.S., Nguyen, E.A., Drake, K.A., Huntsman, S., Hu, D., et al. (2014). Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study. J. Allergy Clin. Immunol. *134*, 295–305. https://doi.org/10.1016/j.jaci.2013.08.055.
62. O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J.E., Rudan, I., et al. (2014). A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet. *10*, 1004234. https://doi.org/10.1371/journal.pgen.1004234.
63. Kidd, J.M., Gravel, S., Byrnes, J., Moreno-Estrada, A., Musharoff, S., Bryc, K., Degenhardt, J.D., Brisbin, A., Sheth, V., Chen, R., et al. (2012). Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. Am. J. Hum. Genet. *91*, 660–671. https://doi.org/10.1016/j.ajhg.2012.08.025.
64. Moore, D.S., McCabe, G.P., and Craig, B.A. (2009). Introduction to the Practice of Statistics (W.H. Freedman)).
65. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol. *4*, e72. https://doi.org/10.1371/journal.pbio.0040072.
66. Szpiech, Z.A., and Hernandez, R.D. (2014). Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol. Biol. Evol. *31*, 2824–2827. https://doi.org/10.1093/molbev/msu211.
67. Vitti, J.J., Grossman, S.R., and Sabeti, P.C. (2013). Detecting natural selection in genomic data. Annu. Rev. Genet. *47*, 97–120. https://doi.org/10.1146/annurev-genet-111212-133526.
68. Purcell, S., Sham, P., and Daly, M.J. (2005). Parental phenotypes in family-based association analysis. Am. J. Hum. Genet. *76*, 249–259. https://doi.org/10.1086/427886.
69. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.
70. Chacón-Duque, J.-C., Adhikari, K., Fuentes-Guajardo, M., Mendoza-Revilla, J., Acuña-Alonzo, V., Barquera, R., Quinto-Sánchez, M., Gómez-Valdés, J., Everardo Martínez, P., Villamil-Ramírez, H., et al. (2018). Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. Nat. Commun. *9*, 5388. https://doi.org/10.1038/s41467-018-07748-z.
71. Turner, S.D. (2017). Qqman: Q-Q and Manhattan Plots for GWAS Data.
72. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337. https://doi.org/10.1093/bioinformatics/btq419.
73. Marcus, J.H., and Novembre, J. (2017). Visualizing the geography of genetic variants. Bioinformatics *33*, 594–595. https://doi.org/10.1093/bioinformatics/btw643.
74. Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., et al. (2021). Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res. *49*, D1311–D1320. https://doi.org/10.1093/nar/gkaa840.
75. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Genet. *53*, 1300–1310. https://doi.org/10.1038/s41588-021-00913-z.
76. Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., et al. (2018). Genomic atlas of the human plasma proteome. Nature. *558*, 73–79. https://doi.org/10.1038/s41586-018-0175-2.
77. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.
78. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*, 1005–1012. https://doi.org/10.1093/nar/gky1120.
79. Koboldt, D.C. (2020). Best practices for variant calling in clinical sequencing. Genome Med. *12*, 91. https://doi.org/10.1186/s13073-020-00791-w.

80. Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat. Protoc. *10*, 1556–1566. https://doi.org/10.1038/nprot.2015.105.
81. Carithers, L.J., and Moore, H.M. (2015). The genotype-tissue expression (GTEx) project. Biopreserv. Biobank *13*, 307–308. https://doi.org/10.1089/bio.2015.29031.hmm.
82. Carithers, L.J., Ardlie, K., Barcus, M., Branton, P.A., Britton, A., Buia, S.A., Compton, C.C., DeLuca, D.S., Peter-Demchok, J., Gelfand, E.T., et al.; on behalf of the GTEx Consortium (2015). A novel approach to high-quality postmortem tissue procurement: the GTEx project. Biopreserv. Biobank. *13*, 311–319. https://doi.org/10.1089/bio.2015.0032.
83. American College of Obstetricians and Gynecologists. (2020). Gestational hypertension and preeclampsia. ACOG practice bulletin. Number *222*, 237–260.
84. Instituto Nacional de Estadistica e Informatica (2018). Peru: Perfil Sociodemografico (Informe Nacional).
85. Barbieri, C., Barquera, R., Arias, L., Sandoval, J.R., Acosta, O., Zurita, C., Aguilar-Campos, A., Tito-Álvarez, A.M., Serrano-Osuna, R., Gray, R.D., et al. (2019). The current genomic landscape of western south America: Andes, amazonia, and pacific coast. Mol. Biol. Evol. *36*, 2698–2713. https://doi.org/10.1093/molbev/msz174.
86. Harris, D.N., Song, W., Shetty, A.C., Levano, K.S., Caceres, O., Padilla, C., Borda, V., Tarazona, D., Trujillo, O., Sanchez, C., et al. (2018). Evolutionary genomic dynamics of Peruvians before, during, and after the inca empire. Proc. Natl. Acad. Sci. U S A *115*, E6526–E6535. https://doi.org/10.1073/pnas.1720798115.
87. Borda, V., Alvim, I., Mendes, M., Silva-Carvalho, C., Soares-Souza, G.B., Leal, T.P., Furlan, V., Scliar, M.O., Zamudio, R., Zolini, C., et al. (2020). The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. Proc. Natl. Acad. Sci. U S A *117*, 32557–32565. https://doi.org/10.1073/pnas.2013773117.
88. Homburger, J.R., Moreno-Estrada, A., Gignoux, C.R., Nelson, D., Sanchez, E., Ortiz-Tello, P., Pons-Estel, B.A., Acevedo-Vasquez, E., Miranda, P., Langefeld, C.D., et al. (2015). Genomic insights into the ancestry and demographic history of south America. PLoS Genet. *11*, e1005602. https://doi.org/10.1371/journal.pgen.1005602.
89. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants: fig. 1. Bioinformatics *31*, 3555–3557. https://doi.org/10.1093/bioinformatics/btv402.
90. Tan, D., Liang, H., Cao, K., and Yi, Q. (2017). CUL4A enhances human trophoblast migration and is associated with pre-eclampsia. Int. J. Clin. Exp. Pathol. *10*, 10544–10551.
91. Horne, A.W., Brown, J.K., Tong, S., and Kaitu'u-Lino, T. (2012). Evaluation of ADAM-12 as a diagnostic biomarker of ectopic pregnancy in women with a pregnancy of unknown location. PLoS One *7*, e41442. https://doi.org/10.1371/journal.pone.0041442.
92. Gack, S., Marmé, A., Marmé, F., Wrobel, G., Vonderstraß, B., Bastert, G., Lichter, P., Angel, P., and Schorpp-Kistner, M. (2005). Preeclampsia: increased expression of soluble ADAM 12. J. Mol. Med. *83*, 887–896. https://doi.org/10.1007/s00109-005-0714-9.
93. El-Sherbiny, W., Nasr, A., and Soliman, A. (2012). Metalloprotease (ADAM12-S) as a predictor of preeclampsia: correlation with severity, maternal complications, fetal outcome, and Doppler parameters. Hypertens. Pregnancy. *31*, 442–450. https://doi.org/10.3109/10641955.2012.690059.
94. Cooley, L.F., El Shikh, M.E., Li, W., Keim, R.C., Zhang, Z., Strauss, J.F., Zhang, Z., and Conrad, D.H. (2016). Impaired immunological synapse in sperm associated antigen 6 (SPAG6) deficient mice. Sci. Rep. *6*, 25840. https://doi.org/10.1038/srep25840.
95. Neilson, L.I., Schneider, P.A., Van Deerlin, P.G., Kiriakidou, M., Driscoll, D.A., Pellegrini, M.C., Millinder, S., Yamamoto, K.K., French, C.K., and Strauss, J.F., 3rd. (1999). cDNA cloning and characterization of a human sperm antigen (SPAG6) with homology to the product of the Chlamydomonas PF16 locus. Genomics. *60*, 272–280. https://doi.org/10.1006/geno.1999.5914.
96. Zhu, H.Y., Tong, X.M., Lin, X.N., Jiang, L.Y., Wang, J.X., and Zhang, S.Y. (2015). Expression and distribution of calcium-binding protein S100P in human placenta during pregnancy. Int. J. Fertil. Steril. *8*, 445–452. https://doi.org/10.22074/ijfs.2015.4189.
97. Zhou, T., Wang, H., Zhang, S., Jiang, X., and Wei, X. (2016). S100P is a potential molecular target of cadmium-induced inhibition of human placental trophoblast cell proliferation. Exp. Toxicol. Pathol. *68*, 565–570. https://doi.org/10.1016/j.etp.2016.09.002.
98. Gong, Y.H., Jia, J., Lu, D.H., Dai, L., Bai, Y., and Zhou, R. (2012). Outcome and risk factors of early onset severe pre-eclampsia. Chin. Med. J. *125*, 2623–2627.
99. Wojtowicz, A., Zembala-Szczerba, M., Babczyk, D., Kolodziejczyk-Pietruszka, M., Lewaczynska, O., and Huras, H. (2019). Early- and late-onset preeclampsia: a comprehensive cohort study of laboratory and clinical findings according to the new ISHHP criteria. Int. J. Hypertens. *2019*, 4108271–4108279. https://doi.org/10.1155/2019/4108271.
100. Chelbi, S.T., Doridot, L., Mondon, F., Dussour, C., Rebourcet, R., Busato, F., Gascoin-Lachambre, G., Barbaux, S., Rigourd, V., Mignot, T.M., et al. (2011). Combination of promoter hypomethylation and PDX1 overexpression leads to TBX15 decrease in vascular IUGR placentas. Epigenetics *6*, 247–255. https://doi.org/10.4161/epi.6.2.13791.
101. Ho, L., van Dijk, M., Chye, S.T.J., Messerschmidt, D.M., Chng, S.C., Ong, S., Yi, L.K., Boussata, S., Goh, G.H.Y., Afink, G.B., et al. (2017). ELABELA deficiency promotes preeclampsia and cardiovascular malformations in mice. Science *357*, 707–713. https://doi.org/10.1126/science.aam6607.
102. Taylor, B.L., Liu, F.F., and Sander, M. (2013). Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells. Cell Rep. *4*, 1262–1275. https://doi.org/10.1016/j.celrep.2013.08.010.
103. Fyfe, G.K., Panicker, S., Jones, R.L., and Wareing, M. (2012). Expression of an electrically silent voltage-gated potassium channel in the human placenta. J. Obstet. Gynaecol. *32*, 624–629. https://doi.org/10.3109/01443615.2012.709288.
104. Wang, Z., Liu, N., Shi, S., Liu, S., and Lin, H. (2016). The role of PIWIL4, an argonaute family protein, in breast cancer. J. Biol. Chem. *291*, 10646–10658. https://doi.org/10.1074/jbc.m116.723239.
105. Chirn, G.W., Rahman, R., Sytnikova, Y.A., Matts, J.A., Zeng, M., Gerlach, D., Yu, M., Berger, B., Naramura, M., Kile, B.T., and Lau, N.C. (2015). Conserved piRNA expression from a

distinct set of piRNA cluster loci in eutherian mammals. PLoS Genet. *11*, 1005652. https://doi.org/10.1371/journal.pgen.1005652.

106. Comptour, A., Rouzaire, M., Belville, C., Bouvier, D., Gallot, D., Blanchon, L., and Sapin, V. (2016). Nuclear retinoid receptors and pregnancy: placental transfer, functions, and pharmacological aspects. Cell. Mol. Life Sci. *73*, 3823–3837. https://doi.org/10.1007/s00018-016-2332-9.
107. Huebner, H., Hartner, A., Rascher, W., Strick, R.R., Kehl, S., Heindl, F., Wachter, D.L., Beckmann Md, M.W., Fahlbusch, F.B., and Ruebner, M. (2018). Expression and regulation of retinoic acid receptor responders in the human placenta. Reprod. Sci. *25*, 1357–1370. https://doi.org/10.1177/1933719117746761.
108. Salonen Ros, H., Lichtenstein, P., Lipworth, L., and Cnattingius, S. (2000). Genetic effects on the liability of developing pre-eclampsia and gestational hypertension. Am. J. Med. Genet. *91*, 256–260. https://doi.org/10.1002/(sici)1096-8628(20000410)91:4<256::aid-ajmg3>3.0.co;2-t.
109. Lie, R.T. (2007). Intergenerational exchange and perinatal risks: a note on interpretation of generational recurrence risks. Paediatr. Perinat Epidemiol. *21*, 13–18. https://doi.org/10.1111/j.1365-3016.2007.00832.x.
110. Heyborne, K. (2001). Paternal and maternal components of the predisposition to preeclampsia. New Engl. J. Med. *345*, 149–150.
111. Pipkin, F.B. (2001). Risk factors for preeclampsia. New Engl. J. Med. *344*, 925–926. https://doi.org/10.1056/nejm200103223441209.
112. Li, D.-K., and Wi, S. (2000). Changing paternity and the risk of preeclampsia/eclampsia in the subsequent pregnancy. Am. J. Epidemiol. *151*, 57–62. https://doi.org/10.1093/oxfordjournals.aje.a010122.
113. Tubbergen, P., Lachmeijer, A.M.A., Althuisius, S.M., Vlak, M.E.J., van Geijn, H.P., and Dekker, G.A. (1999). Change in paternity: a risk factor for preeclampsia in multiparous women? J. Reprod. Immunol. *45*, 81–88. https://doi.org/10.1016/s0165-0378(99)00040-6.
114. Tan, C.Y., Ho, J.F.V., Chong, Y.S., Loganath, A., Chan, Y.H., Ravichandran, J., Lee, C.G., and Chong, S.S. (2008). Paternal contribution of HLA-G*0106 significantly increases risk for pre-eclampsia in multigravid pregnancies. MHR: Basic Sci. Reprod. Med. *14*, 317–324. https://doi.org/10.1093/molehr/gan013.
115. Galaviz-Hernandez, C., Arámbula-Meraz, E., Medina-Bastidas, D., Sosa-Macías, M., Lazalde-Ramos, B.P., Ortega-Chávez, M., and Hernandez-García, L. (2016). The paternal polymorphism rs5370 in the EDN1 gene decreases the risk of preeclampsia. Pregnancy Hypertens. *6*, 327–332. https://doi.org/10.1016/j.preghy.2016.07.002.
116. Gray, K.J., Saxena, R., and Karumanchi, S.A. (2018). 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*, 211–218. https://doi.org/10.1016/j.ajog.2017.11.562.
117. Banadakoppa, M., Balakrishnan, M., and Yallampalli, C. (2020). Common variants of fetal and maternal complement genes in preeclampsia: pregnancy specific complotype. Sci. Rep. *10*, 4811. https://doi.org/10.1038/s41598-020-60539-9.
118. Vefring, H., Lie, R.T., Ødegård, R., Mansoor, M.A., and Nilsen, S.T. (2004). Maternal and fetal variants of genetic thrombophilias and the risk of preeclampsia. Epidemiology *15*, 317–322. https://doi.org/10.1097/01.ede.0000112217.33111.23.
119. Nevalainen, J., Ignatius, J., Savolainen, E.-R., Ryynanen, M., and Jarvenpaa, J. (2018). Placenta-mediated pregnancy complications are not associated with fetal or paternal factor V Leiden mutation. Eur. J. Obstet. Gynecol. Reprod. Biol. *230*, 32–35. https://doi.org/10.1016/j.ejogrb.2018.09.016.
120. Pilvar, D., Reiman, M., Pilvar, A., and Laan, M. (2019). Parent-of-origin-specific allelic expression in the human placenta is limited to established imprinted loci and it is stably maintained across pregnancy. Clin. Epigenetics *11*, 94. https://doi.org/10.1186/s13148-019-0692-3.
121. Almawi, W.Y., Al-Shaikh, F.S., Melemedjian, O.K., and Almawi, A.W. (2013). Protein Z, an anticoagulant protein with expanding role in reproductive biology. Reproduction *146*, 73–80. https://doi.org/10.1530/rep-13-0072.
122. Khadijah Ismail, S., and Higgins, J. (2011). Hemostasis in preeclampsia. Semin. Thromb. Hemost. *37*, 111–117. https://doi.org/10.1055/s-0030-1270336.
123. Haram, K., Mortensen, J.H., and Nagy, B. (2014). Genetic aspects of preeclampsia and the HELLP syndrome. J. Pregnancy *2014*, 910751–910813. https://doi.org/10.1155/2014/910751.
124. Wright, D., and Nicolaides, K.H. (2019). Aspirin delays the development of preeclampsia. Am. J. Obstet. Gynecol. *220*. P580.E1–580.E6. https://doi.org/10.1016/j.ajog.2019.02.034.
125. Bretelle, F., Arnoux, D., Shojai, R., D'Ercole, C., Sampol, J., Dignat, F., and Camoin-Jau, L. (2005). Protein Z in patients with pregnancy complications. Am. J. Obstet. Gynecol. *193*, 1698–1702. https://doi.org/10.1016/j.ajog.2005.04.006.
126. Erez, O., Hoppensteadt, D., Romero, R., Espinoza, J., Goncalves, L., Nien, J.K., Kusanovic, J.P., Fareed, J., Gotsch, F., Pineles, B., and Chaiworapongsa, T. (2007). Preeclampsia is associated with low concentrations of protein Z. J. Matern. Fetal Neonatal. Med. *20*, 661–667. https://doi.org/10.1080/14767050701495011.
127. Kaygusuz, I., Firatli-Tuglular, T., Toptas, T., Ugurel, V., and Demir, M. (2011). Low levels of protein Z are associated with HELLP syndrome and its severity. Clin. Appl. Thromb. Hemost. *17*, 214–219. https://doi.org/10.1177/1076029609357738.
128. Xu, Z., Zhang, Y., Liu, W., Liu, Y., Su, Y., Xing, Q., He, X., Wei, Z., Cao, Y., and Xiang, H. (2018). Polymorphisms of F2, PROC, PROZ, and F13A1 genes are associated with recurrent spontaneous abortion in Chinese Han women. Clin. Appl. Thromb. Hemost. *24*, 894–900. https://doi.org/10.1177/1076029617750487.
129. Christians, J.K., Leavey, K., and Cox, B.J. (2017). Associations between imprinted gene expression in the placenta, human fetal growth and preeclampsia. Biol. Lett. *13*, 20170643. https://doi.org/10.1098/rsbl.2017.0643.
130. Hollegaard, B., Byars, S.G., Lykke, J., and Boomsma, J.J. (2013). Parent-offspring conflict and the persistence of pregnancy-induced hypertension in modern humans. PLoS One *8*, 56821. https://doi.org/10.1371/journal.pone.0056821.
131. Zadora, J., Singh, M., Herse, F., Przybyl, L., Haase, N., Golic, M., Yung, H.W., Huppertz, B., Cartwright, J.E., Whitley, G., et al. (2017). Disturbed placental imprinting in preeclampsia leads to altered expression of DLX5, a human-specific early trophoblast marker. Circulation *136*, 1824–1839. https://doi.org/10.1161/circulationaha.117.028110.

132. Zhou, L., Sun, H., Cheng, R., Fan, X., Lai, S., and Deng, C. (2019). ELABELA, as a potential diagnostic biomarker of preeclampsia, regulates abnormally shallow placentation via APJ. Am. J. Physiol. Endocrinol. Metab. *316*, 773–781. https://doi.org/10.1152/ajpendo.00383.2018.
133. Cooper, D.N. (2010). Functional intronic polymorphisms: buried treasure awaiting discovery within our genes. Hum. Genomics *4*, 284–288. https://doi.org/10.1186/1479-7364-4-5-284.
134. Lin, H., Hargreaves, K.A., Li, R., Reiter, J.L., Wang, Y., Mort, M., Cooper, D.N., Zhou, Y., Zhang, C., Eadon, M.T., et al. (2019). RegSNPs-intron: a computational framework for predicting pathogenic impact of intronic single nucleotide variants. Genome Biol. *20*, 254. https://doi.org/10.1186/s13059-019-1847-4.
135. Joynt, A.T., Evans, T.A., Pellicore, M.J., Davis-Marcisak, E.F., Aksit, M.A., Eastman, A.C., Patel, S.U., Paul, K.C., Osorio, D.L., Bowling, A.D., et al. (2020). Evaluation of both exonic and intronic variants for effects on RNA splicing allows for accurate assessment of the effectiveness of precision therapies. Plos Genet. *16*, 1009100. https://doi.org/10.1371/journal.pgen.1009100.
136. Klee, K.M.C., Janecke, A.R., Civan, H.A., Rosipal, Š., Heinz-Erian, P., Huber, L.A., Müller, T., and Vogel, G.F. (2020). AP1S1 missense mutations cause a congenital enteropathy via an epithelial barrier defect. Hum. Genet. *139*, 1247–1259. https://doi.org/10.1007/s00439-020-02168-w.
137. Huang, J., Sabater-Lleal, M., Asselbergs, F.W., Tregouet, D., Shin, S.-Y., Ding, J., Baumert, J., Oudot-Mellakh, T., Folkersen, L., Johnson, A.D., et al. (2012). Genome-wide association study for circulating levels of PAI-1 provides novel insights into its regulation. Blood *120*, 4873–4881. https://doi.org/10.1182/blood-2012-06-436188.
138. Aghababaei, M., Perdu, S., Irvine, K., and Beristain, A.G. (2014). A disintegrin and metalloproteinase 12 (ADAM12) localizes to invasive trophoblast, promotes cell invasion and directs column outgrowth in early placental development. Mol. Hum. Reprod. *20*, 235–249. https://doi.org/10.1093/molehr/gat084.
139. Díaz, B., Yuen, A., Iizuka, S., Higashiyama, S., and Courtneidge, S.A. (2013). Notch increases the shedding of HB-EGF by ADAM12 to potentiate invadopodia formation in hypoxia. J. Cell Biol. *201*, 279–292. https://doi.org/10.1083/jcb.201209151.
140. Biadasiewicz, K., Fock, V., Dekan, S., Proestling, K., Velicky, P., Haider, S., Knöfler, M., Fröhlich, C., and Pollheimer, J. (2014). Extravillous trophoblast-associated ADAM12 exerts pro-invasive properties, including induction of integrin beta 1-mediated cellular Spreading1. Biol. Reprod. *90*, 1–10. https://doi.org/10.1095/biolreprod.113.115279.
141. Cowans, N.J., and Spencer, K. (2007). First-trimester ADAM12 and PAPP-A as markers for intrauterine fetal growth restriction through their roles in the insulin-like growth factor system. Prenatal Diagn. *27*, 264–271. https://doi.org/10.1002/pd.1665.
142. Laigaard, J., Christiansen, M., Fröhlich, C., Pedersen, B.N., Ottesen, B., and Wewer, U.M. (2005). The level of ADAM12-S in maternal serum is an early first-trimester marker of fetal trisomy 18: ADAM12 and TRISOMY 18. Prenat. Diagn. *25*, 45–46. https://doi.org/10.1002/pd.1029.
143. Bestwick, J.P., George, L.M., Wu, T., Morris, J.K., and Wald, N.J. (2012). The value of early second trimester PAPP-A and ADAM12 in screening for pre-eclampsia. J. Med. Screen. *19*, 51–54. https://doi.org/10.1258/jms.2012.011085.
144. Kulkarni, H. (2013). ADAM12: the usual suspect in pre-eclampsia. Reprod. Syst. Sex. Disord. *02*, 1000121. https://doi.org/10.4172/2161-038x.1000121.
145. Oleksyk, T.K., Smith, M.W., and O'Brien, S.J. (2010). Genome-wide scans for footprints of natural selection. Phil. Trans. R. Soc. B *365*, 185–205. https://doi.org/10.1098/rstb.2009.0219.
146. Jackson, M.R., Mayhew, T.M., and Haas, J.D. (1987). The volumetric composition of human term placentae: altitudinal, ethnic and sex differences in Bolivia. J. Anat. *152*, 173–187.
147. Jackson, M.R., Mayhew, T.M., and Haas, J.D. (1988). On the factors which contribute to thinning of the villous membrane in human placentae at high altitude. II. An increase in the degree of peripheralization of fetal capillaries. Placenta *9*, 9–18. https://doi.org/10.1016/0143-4004(88)90068-9.
148. Belbin, G.M., Nieves-Colón, M.A., Kenny, E.E., Moreno-Estrada, A., and Gignoux, C.R. (2018). Genetic diversity in populations across Latin America: implications for population and medical genetic studies. Curr. Opin. Genet. Dev. *53*, 98–104. https://doi.org/10.1016/j.gde.2018.07.006.
149. Abdelmoumen, I., Jimenez, S., Valencia, I., Melvin, J., Legido, A., Diaz-Diaz, M.M., Griffith, C., Massingham, L.J., Yelton, M., Rodríguez-Hernández, J., et al. (2020). Boricua founder variant in FRRS1L causes epileptic encephalopathy with hyperkinetic movements. J. Child. Neurol. *36*, 93–98. https://doi.org/10.1177/0883073820953001.
150. Asgari, S., Luo, Y., Akbari, A., Belbin, G.M., Li, X., Harris, D.N., Selig, M., Bartell, E., Calderon, R., Slowikowski, K., et al. (2020). A positively selected FBN1 missense variant reduces height in Peruvian individuals. Nature *582*, 234–239. https://doi.org/10.1038/s41586-020-2302-0.
151. Kenny, E.E., Timpson, N.J., Sikora, M., Yee, M.-C., Moreno-Estrada, A., Eng, C., Huntsman, S., Burchard, E.G., Stoneking, M., Bustamante, C.D., and Myles, S. (2012). Melanesian blond hair is caused by an amino acid change in TYRP1. Science *336*, 554. https://doi.org/10.1126/science.1217849.
152. Hurtado, A., Escudero, E., Pando, J., Sharma, S., and Johnson, R.J. (2012). Cardiovascular and renal effects of chronic exposure to high altitude. Nephrol. Dial. Transplant. *27*, iv11–iv16. https://doi.org/10.1093/ndt/gfs427.
153. Arestegui, A.H., Fuquay, R., Sirota, J., Swenson, E.R., Schoene, R.B., Jefferson, J.A., Chen, W., Yu, X.q., Kelly, J.P., Johnson, R.J., and Escudero, E. (2011). High altitude renal syndrome (HARS). JASN *22*, 1963–1968. https://doi.org/10.1681/asn.2010121316.
154. Jefferson, J.A., Escudero, E., Hurtado, M.-E., Kelly, J.P., Swenson, E.R., Wener, M.H., Burnier, M., Maillard, M., Schreiner, G.F., Schoene, R.B., et al. (2002). Hyperuricemia, hypertension, and proteinuria associated with high-altitude polycythemia. Am. J. Kidney Dis. *39*, 1135–1142. https://doi.org/10.1053/ajkd.2002.33380.