

# 1 Clotting factor genes are associated with preeclampsia in high altitude 2 pregnant women in the Peruvian Andes

3 **Running Title:** Genetics of preeclampsia in the Peruvian Andes

4

5 Keyla M. Badillo Rivera<sup>1,†</sup>, Maria A. Nieves-Colón<sup>2,3,4,†</sup>, Karla Sandoval Mendoza<sup>3</sup>, Vanessa  
6 Villanueva Dávalos<sup>5</sup>, Luis E. Enriquez Lencinas<sup>5</sup>, Jessica W. Chen<sup>1</sup>, Elisa T. Zhang<sup>1</sup>, Alexandra  
7 Sockell<sup>1</sup>, Patricia Ortiz Tello<sup>1</sup>, Gloria Malena Hurtado<sup>6</sup>, Ramiro Condori Salas<sup>6</sup>, Ricardo  
8 Cebreco<sup>6</sup>, José C. Manzaneda Choque<sup>7</sup>, Franz P. Manzaneda Choque<sup>7</sup>, Germán P. Yábar Pilco<sup>7</sup>,  
9 Erin Rawls<sup>4</sup>, Celeste Eng<sup>8</sup>, Scott Huntsman<sup>8</sup>, Esteban González Burchard<sup>8</sup>, Giovanni Poletti<sup>6</sup>,  
10 Carla Gallo<sup>6</sup>, Carlos D. Bustamante<sup>1,11</sup>, Julie C. Baker<sup>1</sup>, Christopher R. Gignoux<sup>9</sup>, Genevieve L.  
11 Wojcik<sup>10</sup>, Andrés Moreno-Estrada<sup>3,\*</sup>

12

13 <sup>1</sup>*Department of Genetics, Stanford School of Medicine, Stanford, CA, United States, 94305*

14 <sup>2</sup>*Department of Anthropology, University of Minnesota Twin Cities, Minneapolis, MN, United  
15 States, 55455*

16 <sup>3</sup>*Laboratorio Nacional de Genómica para la Biodiversidad (UGA-LANGEBIO), CINVESTAV,  
17 Irapuato, GTO, Mexico, 36821*

18 <sup>4</sup>*School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, United  
19 States, 85281*

20 <sup>5</sup>*Departamento de Gineco-Obstetricia, Hospital Regional Manuel Nuñez Butrón, Puno, Peru,  
21 21002*

22 <sup>6</sup>*Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad*

23 *Peruana Cayetano Heredia, Lima, Peru, 15102*

24 <sup>7</sup>*Universidad Nacional del Altiplano, Puno, Peru, 21002*

25 <sup>8</sup>*Department of Bioengineering and Therapeutic Sciences, University of California San Francisco,*

26 *San Francisco, CA, United States, 94143*

27 <sup>9</sup>*University of Colorado Anschutz Medical Campus, Aurora, CO, United States, 80045*

28 <sup>10</sup>*Bloomberg School of Public Health, John Hopkins University, Baltimore, MD, United States,*

29 *21205*

30 <sup>11</sup>*Department of Biomedical Data Science, Stanford School of Medicine, Stanford, CA, United*

31 *States, 94305*

32

33 † The authors consider that the first two authors should be regarded as joint First authors.

34

35 **\*Correspondence address:** Laboratorio Nacional de Genómica para la Biodiversidad,

36 CINVESTAV, Km 2.6 Libramiento Norte Carretera Irapuato-León, C.P. 36281, Irapuato,

37 Guanajuato, Mexico. Tel: +52 (462) 166-3000 Ext. 3141, E-mail: andres.moreno@civestav.mx

38

39

40

41

42

43

44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86

## Abstract

**Study question:** What is the genetic basis of preeclampsia in Andean families residing at high altitudes?

**Summary answer:** A top candidate region associated with preeclampsia containing clotting factor genes *PROZ*, *F7* and *F10* was found on chromosome 13 of the fetal genome in affected Andean families.

**What is known already:** Preeclampsia, a multi-organ complication of pregnancy, is a leading cause of maternal morbidity and mortality worldwide. Diagnosed by the onset of maternal hypertension and proteinuria after 20 weeks of gestation, this disorder is a common cause of preterm delivery and affects approximately 5-7% of global pregnancies. The heterogeneity of preeclampsia has posed a challenge in understanding its etiology and molecular basis. However, risk for the condition is known to increase in high altitude regions such as the Peruvian Andes.

**Study design, size, duration:** To investigate the genetic basis of preeclampsia in a high-altitude resident population, we characterized genetic diversity in a cohort of Andean families (N=883) from Puno, Peru, a high-altitude city above 3,500 meters. Our study collected DNA samples and medical records from case-control trios and duos between 2011-2016, thus allowing for measurement of maternal, paternal, and fetal genetic factors influencing preeclampsia risk.

**Participants/materials, setting, methods:** We generated high-density genotype data for 439,314 positions across the genome, determined ancestry patterns and mapped associations between genetic variants and preeclampsia phenotype. We also conducted fine mapping of potential causal variants in a subset of family participants and tested ProZ protein levels in post-partum maternal and cord blood plasma by ELISA.

**Main results and the role of chance:** A transmission disequilibrium test (TDT) revealed variants near genes of biological importance in pregnancy physiology for placental and blood vessel function. The most significant SNP in this cluster, rs5960 ( $p < 6 \times 10^{-6}$ ) is a synonymous variant in the clotting factor *F10*. Two other members of the coagulation cascade, *F7* and *PROZ*, are also in the top associated region. However, we detected no difference of PROZ levels in maternal or umbilical cord plasma.

**Limitations, reasons for caution:** Our genome-wide association analysis (GWAS) was limited by a small sample size and lack of functional follow up. Our ELISA was limited to post-natal blood sampling (only samples collected immediately after birth). But, despite a small sample size, our family based GWAS design permits identification of novel significant and suggestive associations with preeclampsia. Further longitudinal studies could analyze clotting factor levels

87 and activity in other pregnant cohorts in Peru to assess the impact of thrombosis in  
88 preeclampsia risk among Andean highlanders.

89  
90 **Wider implications of the findings:** These findings support previous evidence suggesting that  
91 coagulation plays an important role in the pathology of preeclampsia and potentially underlies  
92 susceptibility to other pregnancy disorders exacerbated at high altitudes. This discovery of a  
93 novel association related to a functional pathway relevant to pregnancy biology in an  
94 understudied population of Native American origin demonstrates the increased power of  
95 family-based study design and underscores the importance of conducting genetic research in  
96 diverse populations.

97  
98 **Study funding/competing interest(s):**

99 This work was supported in part by the National Science Foundation (NSF) Graduate Research  
100 Fellowship Program Grant No. DGE-1147470 awarded to K.M.B.R. (fellow no. 2014187481);  
101 NSF SBE Postdoctoral Research Fellowship Award No. 1711982 awarded to M.N.C.; an A.P.  
102 Giannini Foundation postdoctoral fellowship, a Stanford Child Health Research Institute  
103 postdoctoral award, and a Stanford Dean's Postdoctoral Fellowship awarded to E.T.Z.; the Chan  
104 Zuckerberg Biohub Investigator Award to C.D.B; a Burroughs Wellcome Prematurity Initiative  
105 Award to J.C.B.; the George Rosenkranz Prize for Health Care Research in Developing Countries,  
106 and the International Center for Genetic Engineering and Biotechnology (ICGEB, Italy) grant  
107 CRP/ MEX15-04\_EC, and Mexico's CONACYT grant FONCICYT/50/2016, each awarded to A.M.E.  
108 Further funding was provided by the Sandler Family Foundation, the American Asthma  
109 Foundation, the RWJF Amos Medical Faculty Development Program, Harry Wm. and Diana V.  
110 Hind Distinguished Professor in Pharmaceutical Sciences II, National Institutes of Health,  
111 National Heart, Lung, and Blood Institute Awards R01HL117004, R01HL128439, R01HL135156,  
112 R01HL141992, National Institute of Health and Environmental Health Sciences Awards  
113 R01ES015794, R21ES24844, the National Institute on Minority Health and Health Disparities  
114 Awards R01MD010443, and R56MD013312, and the National Human Genome Research  
115 Institute Award U01HG009080, each awarded to E.G.B. Author J.W.C. is currently a full-time  
116 employee at Genentech, Inc. and hold stocks in Roche Holding AG. Author E.G.B. reports grants  
117 from the National Institute of Health, Lung, Blood Institute, the National Institute of Health,  
118 General Medical Sciences, the National Institute on Minority Health and Health Disparities, the  
119 Tobacco-Related Disease Research Program, the Food and Drug Administration, and the  
120 Sandler Family Foundation, during the conduct of the study.

121  
122 **Trial registration number:** N/A

123  
124 \* for MESH terms see PubMed at <http://www.ncbi.nlm.nih.gov/pubmed/>  
125 *Version 2.6 25/01/2013*

126  
127 **Keywords:** genetics, preeclampsia, pregnancy, Andean, Peru, *PROZ*, GWAS, family trio,  
128 offspring genome

129

## 130 **Introduction**

131

132 Preeclampsia is a hypertensive disorder of pregnancy that is a leading cause of morbidity and  
133 mortality for mothers and infants worldwide. The disorder complicates 5-7% of global  
134 pregnancies, causes nearly 40% of all premature births, and is associated with 10-15% of all  
135 maternal deaths (Duley, 2009, Rana et al., 2019, Valenzuela et al., 2012). This morbidity is even  
136 higher in developing countries and among communities with limited access to healthcare  
137 (Osungbade and Ige, 2011). Despite posing a significant global disease burden, the  
138 heterogeneity of preeclampsia has posed a major challenge for understanding its etiology and  
139 genetic basis (Phipps et al., 2019, Valenzuela et al., 2012).

140

141 Clinical and pathological research suggests a major role for the placenta in preeclampsia, where  
142 shallow invasion of fetal cells into the maternal endometrium results in insufficient remodeling  
143 of the maternal vasculature (Yong et al., 2018). While it roots in early placental development,  
144 preeclampsia is usually not detected until the third trimester of pregnancy (>20 weeks  
145 gestation), when it is identified by a sudden onset of hypertension and signs of organ damage,  
146 typically proteinuria (excess protein in the urine). The severity of preeclampsia is determined by  
147 gestational age at onset, as well as the magnitude of hypertension and organ damage  
148 (American College of Obstetricians and Gynecologists, 2013). The disorder is known to be  
149 heritable with multicomponent risk determined by maternal, fetal, and paternal factors  
150 (McGinnis et al., 2017, Pappa et al., 2011, Phipps et al., 2019, Valenzuela et al., 2012). Other risk

151 factors include family history (Boyd et al., 2013, Cincotta and Brennecke, 1998), socioeconomic  
152 status (Silva et al., 2008) and chronic hypertension or diabetes (Rana, et al., 2019). Residence at  
153 high altitudes above 2,500 meters (m) also contributes considerably to risk of developing  
154 preeclampsia (Zamudio, 2007).

155

156 Residence at high altitudes increases the risk for preeclampsia and other hypertensive  
157 pregnancy disorders at least two to threefold (Moore et al., 2011). For example, Bolivian  
158 communities living at 3,500 m altitude have an incidence of preeclampsia of up to 20% (Keyes  
159 et al., 2003), about three times higher than the world average (Abalos et al., 2013). In  
160 neighboring Peru, preeclampsia complicates up to 22% of all pregnancies and is the second  
161 leading cause of maternal deaths (Gil Cipirán, 2017, Guevara Ríos and Meza Santibáñez, 2014).  
162 Due to this high incidence, highland pregnancy studies have been proposed as a natural  
163 experiment to elucidate genetic factors involved in preeclampsia and other hypertensive  
164 pregnancy complications (Moore et al., 1982, Moore et al., 2004, Palmer et al., 1999, Tissot van  
165 Patot et al., 2009, Zamudio, 2007). Native Andean populations are of particular interest for this  
166 research due to their unique physiological adaptations to chronic high-altitude hypoxia, such as  
167 enhanced pulmonary volumes and elevated blood hemoglobin concentrations (Bigham et al.,  
168 2013). Candidate genes involved in these adaptations include EGLN1, NOS2 and the hypoxia-  
169 inducible factor 1 (HIF1) pathway, among others (Beall, 2014, Bigham, et al., 2013).

170

171 Previous research has found that Highland Andean ancestry and long term, multi-generational  
172 residence at altitude are associated with lower rates of hypoxia induced pregnancy

173 complications among high altitude resident women (Julian et al., 2009, Moore, et al., 2011,  
174 Moore, et al., 2004). Because preeclampsia risk increases with altitude (Palmer, et al., 1999),  
175 these findings suggest that Andeans with Native American ancestry may carry rare adaptive  
176 variants or a unique repertoire of genetic risk factors for preeclampsia—distinct from other  
177 populations previously studied (Michita et al., 2018). Characterizing fine-scale ancestry and  
178 genetic structure patterns in native Andeans may uncover preeclampsia relevant genetic  
179 variation found at higher frequencies due to selection for altitude adaptation (Bigham and Lee,  
180 2014, Tishkoff, 2015).

181  
182 To this end, here we analyze genotype data from a large cohort of preeclamptic Andean  
183 families from Puno, Peru (Figure 1A). This city, located at 3,830 m altitude, has a population  
184 with one of the highest incidences of preeclampsia and associated maternal mortality in the  
185 world (Bristol, 2009, Gil Cipirán, 2017). Our work takes a comprehensive approach to the  
186 genetic study of preeclampsia in a population adapted to high-altitude by employing a family-  
187 study design within a case-control cohort. This enables identification of genetic regions that  
188 influence preeclampsia considering each of the family members that affect disease risk—  
189 mothers, fathers, and offspring—unlike most genome-wide studies focused on pregnancy  
190 disorders which tend to solely include maternal or fetal genomes (Williams and Broughton  
191 Pipkin, 2011). We also aim to understand the role of ancestry-related susceptibility in this  
192 disorder by characterizing genetic diversity and admixture patterns in the Puno cohort.  
193 Additionally, because preeclampsia presents in a spectrum of severity based on gestational age,  
194 organ damage, and hypertension, we take advantage of extensive cohort phenotyping to study

195 associations of genetic variants with disease severity. Our findings have implications for general  
196 understanding of preeclampsia, and human pregnancy hypertensive disorders more broadly,  
197 while also shedding light on the genetic factors that underlie human adaptations for successful  
198 reproduction at high altitudes.

199

200

## 201 **Materials & Methods**

202

### 203 **Puno cohort**

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



215 Ethical approval

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

222

223 Phenotypic data

224 Preeclampsia was defined as new onset of hypertension with presence of proteinuria in urine  
225 after 20 weeks of gestation. Hypertension was defined as systolic blood pressure 30 mmHg  
226 higher than basal level, and diastolic blood pressure at least 15 mmHg higher over basal level. If  
227 no prior blood pressure measurements were available, average basal levels were used as prior  
228 (85/55 mmHg). Note that measured basal arterial pressure levels in pregnant women in Puno  
229 are around 80/50 – 90/60 mmHg (systolic/diastolic), much lower than the U.S. standards,  
230 possibly due to altitude adaptation (Segura-Vega, 2019). Proteinuria levels were confirmed to  
231 be at least 30mg/dL by dipstick in two tests 24 hours apart. Severity of preeclampsia was  
232 defined by the attending physician and categorized into mild or severe. Gestational time was  
233 self-reported by the mother (by date of last menstrual period: LMP) or determined by the  
234 neonate Capurro test.

235

236 Blood and tissue collection

237 Whole blood from the mothers was collected within a few hours post-partum by venipuncture  
238 into EDTA tubes and frozen at -20C. Umbilical cord blood was collected by venipuncture  
239 following clamping of the cord immediately after delivery. Paternal blood, and blood from UNA  
240 controls, was obtained upon consent. For plasma, EDTA tubes were spun within 60min of  
241 collection at 1,200g for 10min in a tabletop centrifuge. Separated plasma was transferred to  
242 Eppendorf tubes, spun again under the same conditions for better purity, then stored at -20C in  
243 cryovials.

244

245 Genotypic data

246 DNA was obtained from whole blood with the Promega (USA) Wizard<sup>®</sup> Genomic DNA  
247 Purification Kit following manufacturer's instructions. DNA extracts were initially quantified  
248 with the Nanodrop. DNA content and quality were further assessed through quantification with  
249 the Qubit<sup>®</sup> Broad Range Assay and by visualizing on a 1% agarose gel, respectively. Samples  
250 that had both >10 ng/uL of DNA concentration and visible bands on the gel were selected for  
251 genotyping. Genotype data at over 800,000 sites across the genome were generated with the  
252 Affymetrix (USA) Axiom Genome-wide LAT 1 array for 950 samples in two batches. Batch 1 was  
253 genotyped in February 2014 at the University of California San Francisco, Gladstone Genomics  
254 Core in Mission Bay, San Francisco, CA. This batch included 360 PRE, 10 PUN and 110 UNA  
255 individuals (n=480). A total of 813,366 variants were successfully genotyped with Batch 1. Batch  
256 2 was genotyped in November 2018 at Affymetrix Research Services Laboratories, Thermo  
257 Fisher Scientific in Santa Clara, CA. This batch included 324 PRE and 146 PUN individuals

258 (n=470), as well as 10 controls added by the genotyping facility. Three samples failed the  
259 genotyping facility filtering metrics, therefore a total of 477 samples and 818,154 variants were  
260 successfully genotyped with Batch 2.

261

## 262 Quality control

263 *Batch 1 data.* The genotyping facility performed a first round of QC restricting the raw dataset  
264 to 713,709 recommended SNPs that passed filtering thresholds for heterozygous strength  
265 offset, cluster resolution, off-target variants, call rate and genotype quality. We further  
266 removed 42 variants with duplicate marker names and flipped 21 SNPs to the forward strand  
267 using snpflip (<https://github.com/biocore-ntnu/snpflip>) and Plink v1.9 (Chang et al., 2015). We  
268 revised that all variants had physical positions in the NCBI Build GRCh37 human reference (hg19  
269 assembly). After QC, Batch 1 dataset included 713,667 biallelic SNPs and 480 individuals.

270

271 *Batch 2 data.* We removed 214 variants with duplicate marker names, 4,233 structural variants  
272 and 540 variants with no physical position in the NCBI Build GRCh37 human reference. 64 SNPs  
273 were flipped to the forward strand as above. Additionally, we followed the genotyping facility  
274 recommendations to restrict this dataset to 777,946 recommended SNPs that passed filtering  
275 thresholds for cluster resolution, off-target variants, call rate and genotype quality. The 10  
276 genotyping controls were also removed. After QC, Batch 2 dataset included 777,946 biallelic  
277 SNPs and 467 individuals.

278

279 *Batch 1 and 2 merge.* We intersected Batch 1 and 2 datasets at overlapping sites using Plink  
280 v1.9. The merged dataset contained 689,528 SNPs and 947 individuals. Using Plink, we removed  
281 1,438 SNPs with genotype missing call frequency >5% (flag: --geno 0.05) and 183,054 SNPs with  
282 minor allele frequency (MAF) <0.5% (flag: --maf 0.005). We also excluded two individuals with  
283 missing call frequency <10% (flag: --mind 0.1). 561 SNPs failing Hardy-Weinberg equilibrium at  
284  $10e-10$  were also excluded. We next filtered our dataset for families with excess Mendelian  
285 errors, cryptic relatedness, and duplicate samples (see Supplementary Table 2 for list of  
286 individuals assigned as unrelated after pedigree revision). 31 individuals were removed, and 56  
287 pedigrees were updated. Chromosomal sex was estimated and sex misassignments were  
288 corrected for 176 individuals whose biological sex was either not recorded or incorrectly  
289 recorded during data collection. After QC, the merged Batch 1 + 2 dataset included 504,475  
290 genome wide SNPs and 914 individuals (Supplementary Figure 1).

291

#### 292 *Batch effect correction*

293 We tested for batch effects by calculating principal components analysis in Plink after filtering  
294 the dataset for linkage disequilibrium and removing related offspring (flags: --indep-pairwise  
295 100 10 0.1, --pca). We initially identified a strong batch effect with the top principal  
296 components statistically significantly associated with batch ( $P < 0.05$ ) (Supplementary Figure 2).  
297 To correct this effect, we conducted an additional round of site and sample-specific filtering.  
298 We removed symmetrical SNPs (AT, CG), excluded all sites not included in the “Best and  
299 Recommended” list provided by Affymetrix for this array, and filtered sites with genotype  
300 missingness <5% and MAF >0.5%. Additionally, we removed individuals with excess

301 heterozygosity (outliers >4SD), duplicate individuals and individuals with cryptic or unexpected  
302 relatedness. In total, 65,161 SNPs and 31 individuals were removed. We repeated the principal  
303 components calculation as above on the filtered dataset and found no statistically significant  
304 association between batch and the top principal components (Supplementary Figure 2). The  
305 final dataset after batch effect correction included 439,314 genome wide SNPs and 883  
306 individuals.

307

### 308 Population structure

309 We intersected our dataset with reference panels including five populations from 1000  
310 Genomes (1KG) Phase 3: Yoruba from Ibadan, Nigeria (YRI), Utah residents with Northern and  
311 Western European ancestry (CEU), Han Chinese from Beijing, China (CHB), Mexican Americans  
312 from Los Angeles, USA (MXL) and Peruvians from Lima, Peru (PEL). After merging, we removed  
313 offspring and related individuals, restricted to autosomes and re-applied quality filters. The  
314 filtered, merged dataset consisted of 422,224 variants and 1,057 individuals. The unsupervised  
315 clustering algorithm ADMIXTURE (Alexander et al., 2009) was run on this dataset to explore  
316 global patterns of population structure. As recommended by the ADMIXTURE manual, the input  
317 data was LD pruned using Plink (flag: --indep-pairwise 50 10 0.1). After LD pruning, 45,496  
318 variants remained for analysis. Ten ancestral clusters (K=2 through K=10) were tested and the  
319 best fit model was selected after examining cross-validation errors. To account for possible  
320 convergence variation, we performed 10 additional runs using different random seeds per run  
321 and estimated parameter standard errors using 200 bootstrap replicates per run. ADMIXTURE  
322 results were plotted with the R pophelper package (Francis, 2017). Principal components

323 analysis (PCA) was applied to the LD pruned dataset using EIGENSOFT v7.2.1 (Patterson et al.,  
324 2006) and plots were generated using the ggplot2 package in R v4.0.3 (R Core Team, 2018,  
325 Wickham, 2016).

326

### 327 Phasing and local ancestry estimation

328 We used RFMix v1.5.4 (Maples et al., 2013) to determine genome wide local ancestry  
329 proportions for the Puno cohort founders, assuming a model of K=3 ancestral populations. The  
330 choice of K=3 reference populations was informed by the ADMIXTURE results. The reference  
331 panel included 108 YRI and 94 CEU individuals from 1000 Genomes Phase 3, and 94 native  
332 individuals from Mexico (30 Mixe, 15 Zapotec, 49 Nahua) genotyped as part of the GALA II  
333 study (Galanter et al., 2014). These reference samples were used as proxies for African,  
334 European, and Native American ancestral source populations, respectively. After merging, the  
335 analysis ready dataset consisted of 420,105 overlapping variants and 899 individuals. The data  
336 were phased with SHAPEIT2 (O'Connell et al., 2014). RFMix was run with default parameters  
337 and EM=2 iterations. Ancestry call cutoffs were determined with a 0.9 posterior probability  
338 threshold as recommended in (Kidd et al., 2012).

339

### 340 Ancestry proportions analysis

341 We tested for significant differences in proportions of Native American, European, and African  
342 ancestry components between PRE cases, PUN and UNA controls. We applied the Wilcoxon  
343 signed ranks test in R v3.5.1 (pairwise.wilcox.test function) with Bonferoni correction for  
344 multiple testing. This non-parametric test assesses whether significant differences exist

345 between two distributions (Moore et al., 2009). Our null hypothesis was that the distribution of  
346 each ancestry proportion was identical between PRE cases, PUN and UNA controls.

347

#### 348 [Statistical analysis of clinical phenotypes](#)

349 We assessed batch bias of clinical phenotypes and correlation with each other by statistical  
350 analysis in R v3.4.0 (R Core Team, 2018). The following dichotomous phenotypes were tested  
351 for batch association with a chi squared test: severity of diagnosis (mild or severe), proteinuria  
352 (+/++ or +++), parity (nulliparous or more than one birth), sex of newborn and mode of delivery  
353 (vaginal or C-section). The following continuous phenotypes were tested for batch association  
354 by t-test: gestational time measured by the mother (date of last menstrual period, or LMP) and  
355 by the fetus (Capurro test), neonate weight, systolic and diastolic blood pressure  
356 measurements, and maternal age.

357

#### 358 [Transmission-disequilibrium test \(TDT\) and parent of origin \(POO\)](#)

359 Leveraging the trio family structure, we applied the transmission disequilibrium test (TDT) and  
360 parent-of-origin (TDT-POO) test on all 87 parent-offspring case trios (preeclamptic families with  
361 offspring) in Plink v1.9 using the --tdt flag, with and without the 'poo' modifier. Variants were  
362 then filtered by MAF > 0.05 within the analyzed cohort. The TDT test assumes Mendelian rules  
363 for transmission of alleles and tests if the queried allele is being transmitted/untransmitted  
364 disproportionately from parents to the affected offspring population (Purcell et al., 2007,  
365 Purcell et al., 2005). The POO analysis is part of TDT, and separately queries transmission from

366 each parent individually to assess paternal or maternal specific transmission. This test self-  
367 corrects for covariate effects by treating each trio as a separate unit.

368

#### 369 GWAS for case-control association

370 Puno cohort individuals were divided into offspring and mothers for two separate case-control  
371 GWAS analyses using logistic regression in Plink (flag: --logistic) with the first 3 PCs and  
372 sequencing batch as covariates. The analysis on the mothers includes 254 PRE and 70 PUN  
373 controls. The offspring analysis includes 225 PRE cases and 60 PUN controls. These analyses  
374 included individuals in trios, duos, and singletons. Variants were filtered by MAF > 0.05 within  
375 the analyzed cohort.

376

#### 377 GWAS in additional phenotypes

378 Multiple phenotypes measured and captured in the recruited patient's medical history allow for  
379 testing of additional genetic associations. We performed additional genome-wide association  
380 analyses of endophenotypes in the PRE mothers (N=254) and offspring (N=225), separately.  
381 These analyses included individuals in trios, duos, and singletons. The endophenotypes tested  
382 for each were: (1) gestational age, maternal measurement; (2) gestational age, fetal  
383 measurement; (3) diastolic blood pressure at diagnosis of preeclampsia; (4) systolic blood  
384 pressure at diagnosis of preeclampsia; (5) proteinuria at diagnosis and (6) severity of diagnosis.  
385 The first four were treated as continuous variables and analyzed by linear regression in Plink  
386 (flag: --linear). Proteinuria and severity of diagnosis were dichotomous variables analyzed in  
387 Plink by logistic regression (flag: --logistic), with proteinuria reduced to + and ++ vs. +++.



388 Genotyping batch was included as a discrete covariate and the first 3 PCs as continuous  
389 covariates. Several of these analyses included less individuals due to missing data. Specifically,  
390 GWAS with systolic and diastolic blood pressure included 253 PRE mothers and 224 PRE  
391 offspring, and GWAS with maternal measurement of gestational age included 252 PRE mothers  
392 and 223 PRE offspring.

393

#### 394 GWAS data visualization

395 All genome-wide analyses were filtered by  $MAF \geq 0.05$  within the analyzed cohorts and  
396 visualized by Manhattan plots using the qqman R package v0.1.4 (Turner, 2017). QQ plots were  
397 generated with the same package to confirm no effects from population structure or other  
398 confounders. Regions of interest were selected if they met two criteria: (1) p-value ( $p < 10E-4$  in  
399 most cases—unless specified in the results section) and (2) the presence of nearby associated  
400 SNPs forming a skyscraper-like structure in the Manhattan plot. Top SNPs in these regions were  
401 selected, and their genomic regions plotted using LocusZoom (Pruim et al., 2010). Maps  
402 displaying the geographic distribution of candidate associated variants were produced using the  
403 Geography of Genetic Variants (GGV) browser (Marcus and Novembre, 2017).

404

#### 405 Capture sequencing

406 We conducted fine mapping of potential causal variants in a subset of families genotyped in  
407 Batch 1 previous to Batch 2 genotyping. Preliminary data obtained from Batch 1 genotypes  
408 were analyzed using standard family-based TDT on Plink for preeclampsia associations (as  
409 above), and regression analysis on secondary phenotypes was conducted using linear mixed

410 models in GTCA (Yang et al., 2011) (flag: `-mlma-loco`). Based on these preliminary results, we  
411 designed a target capture assay including windows around top hits for preeclampsia and  
412 secondary phenotypes, as well as several genes previously suggested to be associated with  
413 preeclampsia in the GWAS catalog (release 2.0.5) (Buniello et al., 2019). The total capture size  
414 was approximately 10Mb (Supplementary File 1).

415

416 We next selected families from Batch 1 with the strongest associations on the preliminary TDT  
417 analysis (n=86 individuals, Supplementary Table 1). Genomic DNA from 86 individuals  
418 (Supplementary Figure 3) was fragmented by mechanical shearing (Covaris) and prepared using  
419 the KAPA Hyperprep library preparation kit (Kapa Biosystems, now part of Roche, Switzerland).  
420 DNA capture was performed on the libraries using the Agilent (USA) SureSelect platform  
421 following manufacturer's instructions. Paired-end sequencing of captured libraries was  
422 performed on the Illumina NextSeq. Sequence data were analyzed through a standard FASTQC-  
423 BWA-GATK pipeline following guidelines as described in (Koboldt, 2020). We then performed  
424 the same GWAS analyses listed above (TDT test for the preeclampsia phenotype and linear  
425 regressions for continuous phenotypes) in the captured regions in a limited set of individuals:  
426 25 trios, 4 duos (3 mother-offspring, 1 father-offspring) and 3 singletons (1 offspring and 2  
427 mothers). Candidate loci identified in these analyses were individually merged and annotated  
428 with ANNOVAR (Yang and Wang, 2015) and overlapped with GTEx single-tissue cis-eQTL data  
429 (version V6p) from the online database (<https://gtexportal.org/home/datasets>) to find relevant  
430 GTEx annotations in our data set (Carithers et al., 2015, Carithers and Moore, 2015).

431

432 **ProZ ELISA**

433 ProZ levels in post-partum maternal and cord blood plasma were assayed using the human-  
434 ProZ ELISA kit from MyBioSource (USA, Cat. No. MBS765710), following manufacturer  
435 instructions. Maternal and fetal plasma samples were diluted at 1:400 in sample diluent and all  
436 washes were performed manually with a multichannel pipet. Final optical density absorbance at  
437 450nm was read using the Bio Rad (USA) iMark™ Microplate Absorbance reader. A 4-  
438 Parameter curve fit was applied to the standards, and the resulting equation was used to  
439 calculate concentration in the experimental samples. Boxplots and t-tests were done in R v3.4.0  
440 (R Core Team, 2018).

441

442

443 **Results**

444

445 We obtained blood samples and maternal clinical records from consented families at the  
446 *Hospital Regional Manuel Nuñez Butrón*, and blood alone from individuals recruited at the  
447 *Universidad Nacional del Altiplano*. At the time of recruitment, mothers from case families  
448 (labeled PRE throughout this study) were at hospital experiencing pregnancy with a  
449 preeclampsia diagnosis, defined as hypertension and proteinuria after 20 weeks of gestation. It  
450 is important to note that basal blood pressure in this population is lower than in the U.S., and  
451 hypertensive levels can be as low as 110/65 mmHg, compared to 140/90mmHg in U.S.  
452 guidelines. Rather than based on a cutoff, hypertension was defined as a systolic measurement  
453 30 mmHg higher than basal and diastolic at least 15 mmHg higher than basal for each individual

454 (see Materials & Methods for more details). For consistency, and to control for other  
455 hypertensive complications of pregnancy, we included proteinuria in the diagnosis, despite this  
456 factor not being currently required in many diagnostic guidelines (American College of  
457 Obstetricians and Gynecologists, 2020).

458  
459 Mothers from control families (labelled PUN) were experiencing a pregnancy without  
460 complications at time of hospital recruitment. 88 PRE families and two PUN families were  
461 collected as complete trios—including both biological parents and offspring; the rest are duos  
462 (one parent and offspring) and single individuals (mothers) (Table I). Overall, the Puno cohort  
463 collected for this study includes 815 individuals from the PRE group, 204 from the hospital  
464 control group (PUN), and 110 from the university (UNA) as `population controls. We extracted  
465 DNA from blood and genotyped PRE, PUN and UNA individuals in two batches on the Affymetrix  
466 Axiom LAT array. Our final dataset after quality filtering included 439,314 genome wide SNPs  
467 and 883 individuals (see Table I and Supplementary Table 1 for breakdown of PRE, PUN and  
468 UNA).

469  
470 **Puno individuals have high proportions of Native American ancestry**

471 We sought to understand the demographic history of our test population by characterizing  
472 patterns of genetic diversity and population structure in the Puno study cohort. To this end we  
473 intersected the entirety of the Puno cohort dataset (883 individuals) with a reference panel  
474 including five continental populations from the 1000 Genomes (1KG) Project Phase 3 panel:  
475 Yorubans (YRI), Europeans (CEU), Mexicans (MXL), Han Chinese (CHB) and Peruvians from Lima

476 (PEL). Using principal component (PC) analysis, we find that individuals from Puno (either PRE,  
477 PUN, UNA) cluster together in PC space, and are distributed in a clinal pattern alongside  
478 Peruvians from Lima who have high proportions of Native American ancestry (Figure 1B,  
479 Supplementary Figure 4).

480

481 We next investigated admixture patterns in the Puno population with the goal of characterizing  
482 proportions of Native versus non-Native genomic ancestry. Using the clustering algorithm  
483 ADMIXTURE (Alexander, et al., 2009), we explored unsupervised models assuming K=2 through  
484 K=10 ancestral clusters (Supplementary Figure 5). Cross-validation errors for each K cluster are  
485 shown in Supplementary Figure 6. At K=4, we observe a clear separation of continental-scale  
486 ancestry components. We find that Puno individuals have large proportions of Native American  
487 ancestry and small proportions of European ancestry, represented by blue and red in Figure 1C,  
488 respectively. At the best fit model of K=6, ADMIXTURE analysis finds substructure within the  
489 Native American ancestry component of the Puno cohort. Specifically, we observe a Puno-  
490 specific ancestry component (shown in light blue in Figure 1C) which is not present within the  
491 Native American ancestry components of 1KG Mexican and Peruvian individuals. This  
492 substructure may derive from an Andean specific ancestry component that has been previously  
493 identified among Indigenous and mestizo communities from the Andean Highlands (Barbieri et  
494 al., 2019, Harris et al., 2018). Overall, we find that individuals in the Puno cohort are  
495 predominantly of Native American ancestry (95.7% on average) and have low levels of non-  
496 Native American admixture (approximately 4.2% on average; Supplementary Table 3). We  
497 further find that the Puno population carries a Highland-specific Native American sub-

498 continental ancestry component, as noted in previous work (Barbieri, et al., 2019, Harris, et al.,  
499 2018).

500

501 Finally, we tested for significant differences in ancestry proportions between cases (PRE) and  
502 controls (PUN, UNA) in the Puno cohort. Guided by the findings of the ADMIXTURE analysis, we  
503 used RFMix to determine local ancestry proportions in the Puno cohort assuming a model of  
504 K=3 ancestral components. We next extrapolated average ancestry proportions per individual  
505 from the RFMix local ancestry calls (Supplementary Tables 4-5). The results of this estimation  
506 further confirm the predominantly Native American ancestry background and highlight the  
507 small proportion of European admixture present in our sample. We next performed a Wilcoxon  
508 rank test to contrast ancestry proportions between PRE, PUN and UNA. This test identified a  
509 small but significant difference in European ancestry proportions between PRE and UNA but  
510 found no significant differences in Native American or African ancestry proportions  
511 (Supplementary Figure 7, Supplementary Table 6). Overall, UNA individuals have slightly higher  
512 proportions of European ancestry than PRE and PUN individuals. However, proportions of  
513 Native American ancestry are not significantly different between cases (PRE) and controls (PUN,  
514 UNA). These findings support the results of the ADMIXTURE analysis and further underscore the  
515 primarily Native American ancestry background of the Puno cohort.

516

517

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

520 Next, we sought to identify genetic loci associated with the risk of preeclampsia in this highly  
521 susceptible population adapted to the hypoxic conditions of the Andean Highlands. As decades  
522 of genetic research have shown a role for maternal, paternal and offspring genomes on  
523 preeclampsia risk (Galaviz-Hernandez et al., 2018, Gray et al., 2018, Phipps et al., 2019), we  
524 collected family trios from 88 cases, as well as duos when trio sampling was not possible (either  
525 for lack of consent or due to samples failing genotyping QC), enabling all three genomes to be  
526 evaluated. Since preeclampsia is a complex disease with wide ranging phenotypes, we provide  
527 summaries of relevant phenotypic data for all case pregnancies organized by batch and in trio  
528 cases only (Table II). By statistical comparison, we find that there is moderate batch bias in  
529 approximately half of the measured phenotypes (e.g., Batch 2 had significantly more vaginal  
530 deliveries than C-sections, when compared to Batch 1,  $p < 0.04$ ), but none likely to influence the  
531 analysis when supported by batch correction. In addition to the data shown in Table II, most  
532 mothers (>98%) had no history of chronic hypertension or diabetes and all were non-smokers.

533  
534 To find genetic linkage between genomic loci and preeclampsia, we first performed a parent-  
535 offspring trio GWAS analysis, or transmission-disequilibrium test (TDT), in the 88 affected (PRE)  
536 trios. The TDT offers a robust association test of genotype to phenotype in affected families by  
537 measuring over-transmission of alleles from heterozygous parents to the offspring. With this  
538 analysis, we identified a group of SNPs in linkage disequilibrium (LD) over a cluster of blood  
539 clotting factor genes with a high odds ratio for preeclampsia (Figure 2; Table III; Supplementary

540 Figure 8). The most significant SNP in this cluster, rs5960 (OR 3.05, 95% CI 1.841-5.054,  $p < 6 \times 10^{-6}$ ;  
541  $1000G$  MAF 0.623), is a synonymous variant in the clotting factor *F10*. Two other members of  
542 the coagulation cascade, *F7* and *PROZ*, are also in this region. Another top hit in the TDT, SNP  
543 rs553316 (OR 0.339, 95% CI 0.2041-0.5629,  $p = 1.15 \times 10^{-5}$ ;  $1000G$  MAF 0.408), is in high LD with  
544 rs5960 in 1KG Peruvian populations ( $R^2 = 0.7476$ ) (Machiela and Chanock, 2015). Additionally,  
545 rs553316 is annotated in GTEx as an eQTL for *PROZ* on mammary tissue (note that, as of our  
546 analysis, no placental or pregnancy blood data were available on GTEx). The global distribution  
547 of allele frequencies for rs5960 and rs553316 in 1KG reference populations are shown in  
548 Supplementary Figure 9 and noted in Supplementary Table 7.

549

550 Given the importance of clotting genes in pregnancy, we sought to complement the genotype  
551 analysis by performing deep sequencing of targeted genomic regions surrounding rs5960 in a  
552 subset of cohort participants (Supplementary Table 8, Supplementary Figure 3). To fine-map  
553 potential causal variants, we repeated the same TDT analysis described above in the fine-  
554 mapped individuals and cross-referenced with the GTEx database for expression phenotypes in  
555 relevant tissues. This analysis found a strong association of preeclampsia with several eQTLs for  
556 *PROZ* (Supplementary Table 9). Other top hits from the genotype TDT that were recapitulated  
557 in this analysis include variants in the *SLC46A3* and *CUL4A* genes, also located on chromosome  
558 13 (Supplementary Table 9). Both genes have been previously associated with preeclampsia risk  
559 in clinical studies (McGinnis, et al., 2017, Tan et al. 2017). These data suggest that clotting  
560 factors on chromosome 13 may play an important role in preeclamptic pregnancies.

561



562 Finally, we asked whether this *PROZ* eQTL resulted in differential *PROZ* protein expression  
563 between PRE cases and PUN controls. Since the TDT identifies associated variants in the  
564 offspring, we analyzed the umbilical cord plasma of 8 PUN controls and 16 PRE cases by ELISA.  
565 In this limited sample, we detected no difference of *PROZ* levels in umbilical cord plasma  
566 (difference in means = 41.550 ug/mL, 95% CI -342.758 to 425.858,  $p = 0.85$ ) collected after  
567 delivery (Supplementary Table 10, Supplementary Figure 10). However, future testing could  
568 evaluate *PROZ* levels in the placenta, where interaction with the maternal environment is more  
569 significant to the preeclampsia phenotype than in umbilical cord blood.

570

#### 571 Clotting factor locus shows paternal inheritance

572 We next examined whether there were loci associated with preeclampsia that were  
573 disproportionately inherited either maternally or paternally. To this end, we performed parent-  
574 of-origin TDT GWAS in the same 88 trios tested above. This test investigates whether any of the  
575 associated SNPs are disproportionately inherited from fathers versus mothers, and vice versa.  
576 The most significant SNP from the TDT analysis, rs5960 in *F10*, is suggested to be paternally  
577 inherited more often than expected by chance ( $p=10^{-4}$ , Figure 2, Table IV, Supplementary Figure  
578 11). Other loci show evidence of paternal inheritance, such as rs79278805 ( $p = 1.77E-04$ ),  
579 located within *SPAG6* on chromosome 10, and rs9399401 ( $p=2.76E-04$ ) in *ADGRG6/GPR126* on  
580 chromosome 6. Similarly, we find several SNPs that show maternal origin bias. The most  
581 significant is rs130121 ( $p=1.91E-04$ ) on chromosome 22 in the *FAM19A5/TAF5* gene, followed  
582 by rs10282765 ( $p=2.39E-04$ ) on chromosome 8 within a ncRNA (Table IV, Supplementary  
583 Figures 12-13). Several genes in the vicinity of these SNPs have been implicated in

584 reproduction. SPAG6 is recognized by anti-sperm antibodies and might be involved in infertility  
585 (Cooley et al., 2016, Neilson et al., 1999). ADGRG6/GPR126 is a G-coupled protein receptor  
586 involved in angiogenesis. It is upregulated in umbilical vein endothelial cells and was found  
587 previously to be upregulated in preeclamptic placentas (Cui et al., 2014, Sitras et al., 2009).  
588 Overall, these parent-of-origin effects support the hypothesis that maternal and/or paternal  
589 bias might contribute to preeclampsia disease.

590

591 *Case-control analysis, placental gene S100P is associated with preeclampsia in the offspring*  
592 While the TDT identifies preeclampsia risk variants from inheritance analysis, a more common  
593 way to test for disease risk variants is to compare cases and controls. The collection of control  
594 (PUN) mother-offspring duos allowed us to compare preeclamptic to healthy pregnancies in  
595 both the mothers and the offspring. To this end, we performed two case-control GWAS of  
596 preeclampsia using Plink (see Materials & Methods): (1) 268 PRE vs. 70 PUN mothers; and (2)  
597 230 PRE and 60 PUN offspring. Several genetic regions showed suggestive association with  
598 preeclampsia in both test groups (Supplementary table 11; Supplementary Figures 14-15). The  
599 most interesting association was the top SNP in the offspring, rs34360485 on chromosome 4 ( $p$   
600  $<2E-5$ , OR 3.615, 95% CI 2.003-6.524, MAF 0.36, (Table V), which contains the placental gene  
601 *S100P*. *S100P* is a calcium-binding protein strongly expressed in the placenta (Zhu et al., 2015)  
602 that promotes trophoblast proliferation in culture (Zhou et al., 2016). The global distribution of  
603 allele frequencies for rs34360485 in 1KG reference populations is shown in Supplementary  
604 Figure 16 and noted in Supplementary table 9.

605

606 Associations of secondary phenotypes reveal loci with roles in placental biology

607 Preeclampsia is a heterogeneous disease with varying potential markers of severity. For  
608 instance, the earlier in gestation preeclampsia occurs, the more severe it is considered to be  
609 (Gong et al., 2012, Wojtowicz et al., 2019). Likewise, all the characteristic clinical features  
610 associated with preeclampsia (such as proteinuria and elevated blood pressure) can present at  
611 varying levels of severity. Harnessing the availability of clinical records for all individuals in the  
612 PRE cohort, we next performed GWAS tests on six secondary phenotypes of preeclampsia  
613 measured at the time of diagnosis: (1) gestational age, maternal measurement; (2) gestational  
614 age, fetal measurement; (3) diastolic blood pressure; (4) systolic blood pressure; (5) proteinuria  
615 and (6) severity of diagnosis as stated by the clinician. It is worth clarifying that gestational age  
616 (the time of the fetus in the womb) was measured in two different ways throughout the study.  
617 The fetal measurement was done by the “Capurro” test, which combines five different  
618 measurements in the neonate, while the maternal measurement relies on the date of the  
619 mother’s last menstrual period before pregnancy.

620

621 To investigate possible genetic associations with secondary phenotypes of preeclampsia, we  
622 performed GWAS analyses by logistic and linear regression for each of the six phenotypes in  
623 254 mothers and 225 offspring, separately. In total, we ran 12 GWAS tests. Logistic regression  
624 was applied to binary phenotypes (proteinuria and severity of diagnosis), while linear  
625 regression was applied to continuous phenotypes (gestational age and blood pressure  
626 measurements). All analyses were corrected for batch and the first three principal components  
627 were included as continuous covariates. With this analysis we found several strong associations

628 of SNPs to secondary maternal phenotypes (Table V; Supplementary table 12). These findings  
629 point to several genetic regions containing relevant genes associated with pregnancy and the  
630 complex biology of preeclampsia, as detailed below.

631

### 632 Gestational Age

633 Gestational age was associated in mothers with one locus on chromosome 1 (rs952593, beta -  
634 1.66, 95% CI  $\pm$  0.61,  $p=3.12 \times 10^{-7}$ , MAF 0.13). This region is near *TBX15* (Table V; Supplementary  
635 Table 12; Supplementary Figure 17-20), a t-box transcription factor shown to be downregulated  
636 in intrauterine growth restricted placentas (Chelbi et al., 2011). The association held true with  
637 both measurements of gestational age (by maternal last period and neonate Capurro test). The  
638 maternal measurement, but not the fetal measurement, of gestational age was associated with  
639 a multigenic locus on chromosome 11 (top SNP rs2581927, beta -2.03, 95% CI  $\pm$  0.85,  $p =$   
640  $4.85 \times 10^{-6}$ ; MAF 0.06). A gene of interest in this locus is *APLNR*, the receptor to *ELABELA*, which  
641 causes preeclampsia symptoms in mice (Supplementary Figures 21-22) (Ho et al., 2017).

642

### 643 Diastolic and Systolic Blood Pressure

644 Diastolic blood pressure reached genome-wide significance for one association in the maternal  
645 genome on chromosome 4 (top SNP rs1874237,  $p < 5 \times 10^{-8}$ , beta -4.257, 95% CI -5.711 – -2.804,  
646 MAF 0.45; Table V; Figure 3). This SNP is within an uncharacterized non-coding RNA locus near  
647 *NKX6-1*, a gene involved in  $\beta$ -cell development and function (Taylor et al., 2013). In the  
648 offspring, both systolic and diastolic blood pressure were strongly associated with SNPs in  
649 *KCNS3/K(V)9.3* (top SNP rs4553827, beta 7.44, 95% CI  $\pm$  2.82,  $p = 5.26 \times 10^{-7}$ , MAF 0.25), a

650 voltage-gated potassium channel gene that is highly expressed in the human placenta, where it  
651 localizes to placental vascular tissues and syncytiotrophoblast cells (Fyfe et al., 2012)  
652 (Supplementary Table 13; Supplementary Figures 23-26).

653

#### 654 Proteinuria and Severity of Diagnosis

655 Proteinuria was most strongly associated in the mothers with rs2760751 on chromosome 17  
656 (OR  $2.83 \pm 1.02$ ,  $p = 5.65E-06$ , MAF 0.29). This SNP is intronic to SMG6, a telomerase binding  
657 protein. A second association with proteinuria in the maternal genome was found with SNP  
658 rs12276362 (OR  $0.41 \pm 0.14$ ,  $p = 1.19E-05$ , MAF 0.49) in chromosome 11, by the *PIWIL4* gene  
659 (Supplementary Figures 27-30). This region is also correlated with severity of diagnosis in the  
660 mothers (rs1940640, OR  $2.4 \pm 0.8$ ,  $p = 1.30E-05$ , MAF 0.43; Supplementary Figures 29, 31). It is  
661 not surprising that proteinuria and severity of diagnosis share a common association, since  
662 these two phenotypes are correlated—clinically severe cases generally have higher levels of  
663 protein in the urine. Aberrant PIWI proteins, which interact with pi-RNAs to drive post-  
664 transcriptional gene regulation, have been found in cancers (Wang et al., 2016), and theoretical  
665 evidence from piRNA evolution suggests a role in placentation, although this has yet to be  
666 proven empirically (Chirn et al., 2015). In the offspring genome, proteinuria showed an  
667 association with placental gene *RARB*, or retinoic acid (RA) receptor beta (rs4241542, OR  $0.26 \pm$   
668  $0.14$ ,  $p=7.04 \times 10^{-6}$ , MAF 0.21) (Comptour et al., 2016, Huebner et al., 2018), while the strongest  
669 association with proteinuria is on a different region, in a SNP intronic to *STK32B* (rs62297274  
670 (OR  $0.35 \pm 0.12$ ,  $p=3.5104 \times 10^{-6}$ ; Supplementary Table 13; Supplementary Figure 32-33).

671 Interestingly, the minor allele for SNP rs62297274 is found at high frequencies in Peruvians

672 compared to other global populations. In the Puno cohort MAF for this variant is 0.49, slightly  
673 higher than among Peruvians from Lima sampled in the 1KG (PEL MAF 0.41) (Supplementary  
674 Figure 34). In contrast, the minor allele is found at low frequencies in the rest of the Americas  
675 (1KG AMR MAF 0.19) and is rarely observed globally (1KG MAF <0.05) (Supplementary Table 9).

676

677

## 678 **Discussion**

679

680 In this analysis, we investigate the genetic diversity of a preeclampsia cohort of Andean families  
681 from Puno, Peru; a population with one of the highest incidences of this disease in the world  
682 (Bristol, 2009, Gil Cipirán, 2017). We harness the power of a trio study design to uncover  
683 maternal, paternal, and fetal genetic factors influencing the incidence and severity of  
684 preeclampsia in this cohort. In contrast to previous preeclampsia GWAS studies, which have  
685 been hampered by limited phenotyping and heterogeneous sampling (Williams and Broughton  
686 Pipkin, 2011), the present work includes a case-control cohort sampled from a single  
687 population, treated at the same hospital, and exposed to similar selective pressures due to  
688 long-term residence at high altitude. Thus, despite a small sample size, our family based GWAS  
689 design permits identification of novel significant and suggestive associations with preeclampsia  
690 that would remain otherwise undiscovered (Tishkoff, 2015).

691

692 Most genetic studies on preeclampsia have not investigated whole family units (Boyd, et al.,

693 2013, Cincotta and Brennecke, 1998, McGinnis, et al., 2017, Salonen Ros et al., 2000), despite

694 the evidence of a complex genetic risk involving factors from both parents and the fetus  
695 (Valenzuela, et al., 2012). This reinforces the strength of our approach, where the top  
696 association in the trio study was rs5960, an intronic variant in the clotting factor gene *PROZ*, in  
697 a locus with two other clotting factors: *F7* and *F10*. *PROZ*, a vitamin K-dependent factor, is an  
698 anticoagulant protein with a role in factor X inhibition (Almawi et al., 2013). Several previous  
699 studies have suggested a hypercoagulative state in preeclampsia (reviewed in Ismail and  
700 Higgins, 2011), as spiral arteries of preeclamptic pregnancies often present thrombosis and  
701 atherosclerosis (Haram et al., 2014). In fact, strong evidence supporting an effect of thrombotic  
702 processes on preeclampsia is based on the observation that aspirin, a known blood thinner,  
703 successfully delays preeclampsia onset (Wright and Nicolaides, 2019).

704

705 Low *PROZ* levels are associated with thrombotic disorders, and many adverse pregnancy  
706 outcomes have also been linked with maternal *PROZ* levels (Almawi et al., 2013). A small,  
707 prospective case-control study found low *PROZ* levels associated to intrauterine growth  
708 restriction (IUGR) and intrauterine fetal demise, but not preeclampsia (Bretelle et al., 2005). In  
709 contrast, a larger cross-sectional study found lower median levels of *PROZ* in preeclampsia  
710 outcomes but not IUGR or fetal demise (Erez et al., 2007). One study found a correlation  
711 between lower *PROZ* levels and severity of HELLP syndrome (a complication of preeclampsia  
712 that stands for haemolysis, elevated liver enzymes, and low platelets), which occurs in 10-20%  
713 of preeclamptic pregnancies (Haram, et al., 2014, Kaygusuz et al., 2011). However, no study on  
714 *PROZ* or other clotting factors in preeclampsia has been successfully replicated, likely due to the  
715 extreme heterogeneity of the disease and the mix of populations studied.

716 As most previous studies on PROZ have focused on the mother's genome (Erez, et al., 2007, Xu  
717 et al., 2018), ours is the first study to suggest a correlation between the fetal *PROZ/F7/F10*  
718 locus on chromosome 13 and preeclampsia. In a subset of our sample, we found no differences  
719 in protein plasma levels of PROZ between preeclamptic and healthy pregnancies in the mother  
720 or the offspring. However, this analysis was limited by small sample size and post-natal blood  
721 sampling. In other words, since samples were only collected immediately after birth, we were  
722 unable to monitor changes in PROZ protein levels throughout the pregnancy. Further  
723 longitudinal studies could analyze clotting factor levels and activity in this pregnant population  
724 to assess the impact of thrombosis in preeclampsia risk among Andean highlanders.

725

726 Expanding the TDT to a parent of origin analysis (POO), we found several associations to genetic  
727 regions with suggested paternal inheritance. For instance, the top TDT hit on F10, rs5960, is  
728 also the locus with the strongest paternal origin effect in the TDT-POO. Although future  
729 research examining variation at the *PROZ/F7/F10* region in a larger population will be needed  
730 to confirm this finding, our results are of interest to studies investigating the role of paternal  
731 genetic factors, genomic imprinting and paternal-offspring conflict in preeclampsia and other  
732 pregnancy disorders (Christians et al., 2017, Galaviz-Hernandez, et al., 2018, Hollegaard et al.,  
733 2013, Pilvar et al., 2019, Wikstrom et al., 2012, Zadora et al., 2017).

734

735 Other top regions in the TDT-POO include biologically relevant genes SPAG6 and ADGRG6,  
736 previously described as being involved in infertility and the immune system (SPAG6) (Cooley, et  
737 al., 2016, Neilson, et al., 1999), or angiogenesis (ADGRG6/GPR126) (Cui, et al., 2014, Sitras, et



738 al., 2009). Of these, only ADGRG6 has been associated with preeclampsia in previous research  
739 that found it upregulated in preeclamptic placentas (Cui et al., 2014, Sitras et al., 2009). Future  
740 work could investigate potential roles of these candidate genes in the maternal-fetal interface  
741 and elucidate their involvement in the pathophysiology of preeclampsia.

742

743 We also found several placental genes associated with secondary phenotypes that underline  
744 the severity of preeclampsia, such as hypertension, gestational age, and proteinuria.

745 Differential expression of these genes may contribute to the insufficiency of placental  
746 development in early pregnancy that leads to hypertension and proteinuria in the third  
747 trimester. Some of our suggestive associations are near genes previously shown to have roles in  
748 pregnancy, vascular processes, and even preeclampsia. One such gene is APLNR, the receptor  
749 to ELABELA, which causes preeclampsia symptoms in mice (Ho, et al., 2017) and is lower in the  
750 serum and placentas of some women with late-onset, but not early-onset preeclampsia (Zhou  
751 et al., 2019). However, this gene is in a multigenic locus, and fine-mapping approaches with  
752 functional studies are required to discover the effect of this locus in our cohort.

753

754 Our study is one of only a few preeclampsia GWAS studies to include the offspring genome.  
755 One recent study with a large cohort found a gene, sFLT1, associated with late (but not early)  
756 preeclampsia (Gray, et al., 2018, McGinnis, et al., 2017), suggesting that dysregulation of genes  
757 in the fetal genome contribute to preeclampsia. In our study, we found novel fetal associations  
758 with preeclampsia and its severity phenotypes in the fetus. For instance, we found an  
759 association between severity of hypertension (systolic and diastolic pressure measurements)

760 and KCNS3/K(V)9.3 a gene that is highly expressed in the human placenta, where it localizes to  
761 placental vascular tissues and syncytiotrophoblast cells (Fyfe, et al., 2012). We also found an  
762 association of the retinoic acid (RA) signaling gene RARB and severity of the proteinuria in the  
763 preeclamptic fetal genome. RA signaling is essential for healthy placental and fetal  
764 development in animal models, with evidence of similar requirement in humans (reviewed in  
765 (Comptour, et al., 2016)). RARB is expressed in the extravillous part of the placenta and its  
766 activation induces RARRES, shown in one study to be overexpressed in preeclamptic placentas  
767 (Huebner, et al., 2018). Our study adds to this body of literature and highlights the role of RA in  
768 proper placentation. Lastly, the most interesting region in the offspring genome was identified  
769 in our case-control study; the S100P gene, a calcium-binding protein strongly expressed in the  
770 placenta (Zhu, et al., 2015) that promotes trophoblast proliferation in culture (Zhou, et al.,  
771 2016). This finding suggests that fetal biology, and specifically placental development driven by  
772 fetal genes, highly contributes to the pathology of preeclampsia.

773

774 We examined the global distribution of allele frequencies for each of the candidate associated  
775 SNPs detailed above. Most alleles were shared among several global populations (see global  
776 distribution plots in Supplementary Figures). A notable exception is SNP rs62297274, an  
777 intronic variant located in gene STK32B which is associated with proteinuria in the offspring  
778 genome. The minor allele reaches its highest global frequency in Peruvian populations  
779 (Supplementary Figure 34). As of this writing SNP rs62297274 has no reported clinical  
780 significance in dbSNP. However, intronic variants are known to have functional impacts on RNA  
781 splicing patterns (Cooper, 2010). To elucidate the functional significance of this variant, future

782 research could evaluate its pathogenic potential in Peruvian populations (Joynt et al., 2020, Lin  
783 et al., 2019).

784

785 As discussed, several genes found in our analyses are involved in placental function.

786 Interestingly, morphological studies comparing placentas from Andean-descent and European-

787 descent individuals in Bolivia, at both low and high altitudes, describe differences in placental

788 composition (Jackson et al., 1987, Jackson et al., 1988). Highland placentas from individuals of

789 both ancestries show more intervillous space but less villi, and the Andean highland placenta,

790 compared to the European, have more trophoblast and villous stroma on average. Differences

791 in placental morphology suggest an adaptive mechanism to the lower oxygen pressure at high

792 altitude, but one that does not lower the risk of preeclampsia.

793

794 In conclusion, this study investigates a cohort of preeclamptic Highland Andean families from

795 Puno, Peru to elucidate the genetic basis of this pregnancy disorder at high altitudes. We

796 generated high-density genotype data at over 400,000 positions across the genome and used

797 these data to determine ancestry patterns and map associations between genetic variants and

798 preeclampsia phenotypes. Our trio-based recruitment strategy, including genotype data from

799 mothers, fathers, and offspring, allowed us to identify novel genetic regions not previously

800 reported in preeclampsia genome-wide association studies. Specifically, we identified strong

801 associations with several variants near genes involved with placental and blood vessel function,

802 and therefore, of functional importance for human pregnancy biology. The strongest

803 association hit involves a cluster of clotting factor genes on chromosome 13 including *PROZ*, *F7*

804 and *10* in the fetal genome. This finding provides supporting evidence that coagulation plays an  
805 important role in the pathology of preeclampsia and potentially underlies other pregnancy  
806 disorders exacerbated at high altitude.

807

808 Studying diverse human groups with unique genetic adaptations enables identification of the  
809 primary genetic factors underlying complex phenotypes and gene function. This research  
810 examined Andean populations as a model to understand human pregnancy physiology in  
811 hypoxic conditions. This natural experimental setting provides a unique opportunity to  
812 understand the genetic factors influencing human reproductive fitness in challenging  
813 environments worldwide and to discover population-specific variants underlying biomedical  
814 traits. Our work also underscores the importance of including diverse populations in genome  
815 wide association studies and functional variant discovery efforts to better understand human  
816 physiology and disease globally.

817

818

### 819 **Authors' roles**

820 K.M.B.R. and M.A.N.C. wrote the article with input from G.L.W., A.M.E. and J.C.B. K.M.B.R.,  
821 M.A.N.C., J.W.C., E.T.Z., C.R.G. and G.L.W. performed data analyses. P.O.T. and K.S.M. designed  
822 and coordinated data collection. P.O.T., K.S.M, L.E.L., V.V.D., J.C.M.C., F.M.C. and G.P.Y.P.  
823 collected samples and medical records in Puno and handled fieldwork logistics. K.M.B.R.,  
824 M.A.N.C., A.S., E.R., G.M.H., R.C.S, R.C., C.E., S.H., E.G.B., E.T.Z., G.P. and C.G. performed  
825 laboratory work. C.D.B, J.C.B., C.R.G., A.M.E., C.G., and M.A.N.C. provided resources, funding,

826 and/or laboratory space. All authors revised the article and approved the final submitted  
827 version.

828

829

### 830 **Acknowledgements**

831 We extend our deepest gratitude to the people of Puno, Peru who participated in this study at  
832 *Hospital Regional Manuel Nuñez Butrón* and *Universidad del Altiplano*. We are tremendously  
833 grateful to Javier Mendoza Revilla who provided commentary on the final version of this  
834 manuscript, and to the Mendoza Revilla family, who provided lodging and logistics support in  
835 Lima during fieldwork seasons of the technical team.

836

837

### 838 **Funding sources:**

839 This work was supported in part by the National Science Foundation (NSF) Graduate Research  
840 Fellowship Program Grant No. DGE–1147470 awarded to K.M.B.R. (fellow no. 2014187481);  
841 NSF SBE Postdoctoral Research Fellowship Award No. 1711982 awarded to M.N.C.; an A.P.  
842 Giannini Foundation postdoctoral fellowship, a Stanford Child Health Research Institute  
843 postdoctoral award, and a Stanford Dean’s Postdoctoral Fellowship awarded to E.T.Z.; the Chan  
844 Zuckerberg Biohub Investigator Award to C.D.B.; a Burroughs Welcome Prematurity Initiative  
845 Award to J.C.B.; the George Rosenkranz Prize for Health Care Research in Developing Countries,  
846 and the International Center for Genetic Engineering and Biotechnology (ICGEB, Italy) grant  
847 CRP/ MEX15-04\_EC, and Mexico's CONACYT grant FONCICYT/50/2016, each awarded to A.M.E..

848 Further funding was provided by the Sandler Family Foundation, the American Asthma  
849 Foundation, the RWJF Amos Medical Faculty Development Program, Harry Wm. and Diana V.  
850 Hind Distinguished Professor in Pharmaceutical Sciences II, National Institutes of Health,  
851 National Heart, Lung, and Blood Institute Awards R01HL117004, R01HL128439, R01HL135156,  
852 R01HL141992, National Institute of Environmental Health Sciences Awards R01ES015794,  
853 R21ES24844, the National Institute on Minority Health and Health Disparities Awards  
854 R01MD010443, and R56MD013312, and the National Human Genome Research Institute Award  
855 U01HG009080, each awarded to E.G.B.

856

857

#### 858 **Conflicts of interest statement**

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

860 E.G.B. reports grants from the National Institute of Health, Lung, Blood Institute, the National

861 Institute of Health, General Medical Sciences, the National Institute on Minority Health and

862 Health Disparities, the Tobacco-Related Disease Research Program, the Food and Drug

863 Administration, and from the Sandler Family Foundation, during the conduct of the study.

864

865

#### 866 **References**

867 Abalos E, Cuesta C, Grosso AL, Chou D, Say L. Global and regional estimates of preeclampsia and  
868 eclampsia: a systematic review. *Eur J Obstet Gynecol Reprod Biol* 2013;170: 1-7.

869 Almawi WY, Al-Shaikh FS, Melemedjian OK, Almawi AW. Protein Z, an anticoagulant protein  
870 with expanding role in reproductive biology. *Reproduction* 2013; 146(2): R73-R80.

- 871 Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated  
872 individuals. *Genome Res* 2009;19: 1655-1664.
- 873 American College of Obstetricians and Gynecologists. Hypertension in Pregnancy. Report of the  
874 American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy.  
875 *Obstetrics & Gynecology* 2013; 122(5): 1122-1131.
- 876 American College of Obstetricians and Gynecologists. Gestational Hypertension and  
877 Preeclampsia: ACOG Practice Bulletin, Number 222. *Obstetrics & Gynecology* 2020; 135(6):  
878 e237-260
- 879 Barbieri C, Barquera R, Arias L, Sandoval JR, Acosta O, Zurita C, Aguilar-Campos A, Tito-Alvarez  
880 AM, Serrano-Osuna R, Gray RD *et al.* The Current Genomic Landscape of Western South  
881 America: Andes, Amazonia, and Pacific Coast. *Mol Biol Evol* 2019;36: 2698-2713.
- 882 Beall CM. Adaptation to High Altitude: Phenotypes and Genotypes. *Annual Review of*  
883 *Anthropology* 2014;43: 251-272.
- 884 Bigham AW, Lee FS. Human high-altitude adaptation: forward genetics meets the HIF pathway.  
885 *Genes Dev* 2014;28: 2189-2204.
- 886 Bigham AW, Wilson MJ, Julian CG, Kiyamu M, Vargas E, Leon-Velarde F, Rivera-Chira M,  
887 Rodriguez C, Browne VA, Parra E *et al.* Andean and Tibetan patterns of adaptation to high  
888 altitude. *Am J Hum Biol* 2013;25: 190-197.
- 889 Boyd HA, Tahir H, Wohlfahrt J, Melbye M. Associations of personal and family preeclampsia  
890 history with the risk of early-, intermediate- and late-onset preeclampsia. *Am J Epidemiol*  
891 2013;178: 1611-1619.
- 892 Bretelle F, Arnoux D, Shojai R, D'Ercole C, Sampol J, Dignat F, Camoin-Jau L. Protein Z in patients  
893 with pregnancy complications. *Am J Obstet Gynecol* 2005;193: 1698-1702.
- 894 Bristol N. Dying to give birth: Fighting maternal mortality in Peru. *Health Affairs* 2009;28: 997-  
895 1002.
- 896 Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A,  
897 Morales J, Mountjoy E, Sollis E *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide  
898 association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47:  
899 D1005-D1012.
- 900 Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-  
901 Demchok J, Gelfand ET *et al.* A Novel Approach to High-Quality Postmortem Tissue  
902 Procurement: The GTEx Project. *Biopreserv Biobank* 2015;13: 311-319.
- 903 Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank*  
904 2015;13: 307-308.

- 905 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to  
906 the challenge of larger and richer datasets. *Gigascience* 2015;4: 7.
- 907 Chelbi ST, Doridot L, Mondon F, Dussour C, Rebourcet R, Busato F, Gascoin-Lachambre G,  
908 Barbaux S, Rigourd V, Mignot TM *et al.* Combination of promoter hypomethylation and PDX1  
909 overexpression leads to TBX15 decrease in vascular IUGR placentas. *Epigenetics* 2011;6: 247-  
910 255.
- 911 Chirn GW, Rahman R, Sytnikova YA, Matts JA, Zeng M, Gerlach D, Yu M, Berger B, Naramura M,  
912 Kile BT *et al.* Conserved piRNA Expression from a Distinct Set of piRNA Cluster Loci in Eutherian  
913 Mammals. *PLoS Genet* 2015;11: e1005652.
- 914 Christians JK, Leavey K, Cox BJ. Associations between imprinted gene expression in the  
915 placenta, human fetal growth and preeclampsia. *Biol Lett* 2017;13.
- 916 Cincotta RB, Brennecke SP. Family history of pre-eclampsia as a predictor for pre-eclampsia in  
917 primigravidas. *Int J Gynaecol Obstet* 1998;60: 23-27.
- 918 Comptour A, Rouzaire M, Belville C, Bouvier D, Gallot D, Blanchon L, Sapin V. Nuclear retinoid  
919 receptors and pregnancy: placental transfer, functions, and pharmacological aspects. *Cell Mol*  
920 *Life Sci* 2016;73: 3823-3837.
- 921 Cooley LF, El Shikh ME, Li W, Keim RC, Zhang Z, Strauss JF, Zhang Z, Conrad DH. Impaired  
922 immunological synapse in sperm associated antigen 6 (SPAG6) deficient mice. *Sci Rep* 2016;6:  
923 25840.
- 924 Cooper DN. Functional intronic polymorphisms: Buried treasure awaiting discovery within our  
925 genes. *Human Genomics* 2010;4: 284-288.
- 926 Cui H, Wang Y, Huang H, Yu W, Bai M, Zhang L, Bryan BA, Wang Y, Luo J, Li D *et al.* GPR126  
927 protein regulates developmental and pathological angiogenesis through modulation of VEGFR2  
928 receptor signaling. *J Biol Chem* 2014;289: 34871-34885.
- 929 Duley L. The global impact of pre-eclampsia and eclampsia. *Semin Perinatol* 2009;33: 130-137.
- 930 Erez O, Hoppensteadt D, Romero R, Espinoza J, Goncalves L, Nien JK, Kusanovic JP, Fareed J,  
931 Gotsch F, Pineles B *et al.* Preeclampsia is associated with low concentrations of protein Z. *J*  
932 *Matern Fetal Neonatal Med* 2007;20: 661-667.
- 933 Francis RM. Pophelper: An R package and web app to analyse and visualize population  
934 structure. *Mol Ecol Resour* 2017;17: 27-32.
- 935 Fyfe GK, Panicker S, Jones RL, Wareing M. Expression of an electrically silent voltage-gated  
936 potassium channel in the human placenta. *J Obstet Gynaecol* 2012;32: 624-629.



- 937 Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SS, Nguyen EA, Drake KA,  
938 Huntsman S, Hu D *et al.* Genome-wide association study and admixture mapping identify  
939 different asthma-associated loci in Latinos: The Genes-environments & Admixture in Latino  
940 Americans study. *J Allergy Clin Immunol* 2014;134: 295-305.
- 941 Galaviz-Hernandez C, Sosa-Macias M, Teran E, Garcia-Ortiz JE, Lazalde-Ramos BP. Paternal  
942 Determinants in Preeclampsia. *Front Physiol* 2018;9: 1870.
- 943 Gil Cipirán F. Situación epidemiológica de la mortalidad materna en el Perú *Boletín*  
944 *Epidemiológico del Perú*. 2017. Centro Nacional de Epidemiología, Prevención y Control de  
945 Enfermedades, Ministerio de Salud, Lima, pp. 1514-1516.
- 946 Gong YH, Jia J, Lu DH, Dai L, Bai Y, Zhou R. Outcome and risk factors of early onset severe  
947 preeclampsia. *Chin Med J (Engl)* 2012;125: 2623-2627.
- 948 Gray KJ, Saxena R, Karumanchi SA. Genetic predisposition to preeclampsia is conferred by fetal  
949 DNA variants near FLT1, a gene involved in the regulation of angiogenesis. *Am J Obstet Gynecol*  
950 2018;218: 211-218.
- 951 Guevara Ríos E, Meza Santibáñez L. Manejo de la preeclampsia/eclampsia en el Perú. *Revista*  
952 *Peruana de Ginecología y Obstetricia* 2014; October: 385-393.
- 953 Haram K, Mortensen JH, Nagy B. Genetic aspects of preeclampsia and the HELLP syndrome. *J*  
954 *Pregnancy* 2014;2014: 910751.
- 955 Harris DN, Song W, Shetty AC, Levano KS, Caceres O, Padilla C, Borda V, Tarazona D, Trujillo O,  
956 Sanchez C *et al.* Evolutionary genomic dynamics of Peruvians before, during, and after the Inca  
957 Empire. *Proc Natl Acad Sci U S A* 2018;115: E6526-E6535.
- 958 Ho L, van Dijk M, Chye STJ, Messerschmidt DM, Chng SC, Ong S, Yi LK, Boussata S, Goh GH, Afink  
959 GB *et al.* ELABELA deficiency promotes preeclampsia and cardiovascular malformations in mice.  
960 *Science* 2017;357: 707-713.
- 961 Hollegaard B, Byars SG, Lykke J, Boomsma JJ. Parent-offspring conflict and the persistence of  
962 pregnancy-induced hypertension in modern humans. *PLoS One* 2013;8: e56821.
- 963 Huebner H, Hartner A, Rascher W, Strick RR, Kehl S, Heindl F, Wachter DL, Beckmann Md MW,  
964 Fahlbusch FB, Ruebner M. Expression and Regulation of Retinoic Acid Receptor Responders in  
965 the Human Placenta. *Reprod Sci* 2018;25: 1357-1370.
- 966 Ismail SK, Higgins JR. Hemostasis in pre-eclampsia. *Semin Thromb Hemost* 2011;37: 111-117.
- 967 Jackson MR, Mayhew TM, Haas JD. The volumetric composition of human term placentae:  
968 altitudinal, ethnic and sex differences in Bolivia. *J Anat* 1987;152: 173-187.

- 969 Jackson MR, Mayhew TM, Haas JD. On the factors which contribute to thinning of the villous  
970 membrane in human placentae at high altitude. II. An increase in the degree of  
971 peripheralization of fetal capillaries. *Placenta* 1988;9: 9-18.
- 972 Joynt AT, Evans TA, Pellicore MJ, Davis-Marcisak EF, Aksit MA, Eastman AC, Patel SU, Paul KC,  
973 Osorio DL, Bowling AD *et al.* Evaluation of both exonic and intronic variants for effects on RNA  
974 splicing allows for accurate assessment of the effectiveness of precision therapies. *PLoS Genet*  
975 2020;16: e1009100.
- 976 Julian CG, Wilson MJ, Moore LG. Evolutionary adaptation to high altitude: a view from in utero.  
977 *Am J Hum Biol* 2009;21: 614-622.
- 978 Kaygusuz I, Firatli-Tuglular T, Toptas T, Ugurel V, Demir M. Low levels of protein Z are associated  
979 with HELLP syndrome and its severity. *Clin Appl Thromb Hemost* 2011;17: 214-219.
- 980 Keyes LE, Armaza JF, Niermeyer S, Vargas E, Young DA, Moore LG. Intrauterine growth  
981 restriction, preeclampsia, and intrauterine mortality at high altitude in Bolivia. *Pediatr Res*  
982 2003;54: 20-25.
- 983 Kidd JM, Gravel S, Byrnes J, Moreno-Estrada A, Musharoff S, Bryc K, Degenhardt JD, Brisbin A,  
984 Sheth V, Chen R *et al.* Population genetic inference from personal genome data: impact of  
985 ancestry and admixture on human genomic variation. *Am J Hum Genet* 2012;91: 660-671.
- 986 Koboldt DC. Best practices for variant calling in clinical sequencing. *Genome Med* 2020;12: 91.
- 987 Lin H, Hargreaves KA, Li R, Reiter JL, Wang Y, Mort M, Cooper DN, Zhou Y, Zhang C, Eadon MT *et*  
988 *al.* RegSNPs-intron: a computational framework for predicting pathogenic impact of intronic  
989 single nucleotide variants. *Genome Biol* 2019;20: 254.
- 990 Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific  
991 haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*  
992 2015;31: 3555-3557.
- 993 Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for  
994 rapid and robust local-ancestry inference. *Am J Hum Genet* 2013;93: 278-288.
- 995 Marcus JH, Novembre J. Visualizing the geography of genetic variants. *Bioinformatics* 2017;33:  
996 594-595.
- 997 McGinnis R, Steinhorsdottir V, Williams NO, Thorleifsson G, Shooter S, Hjartardottir S,  
998 Bumpstead S, Stefansdottir L, Hildyard L, Sigurdsson JK *et al.* Variants in the fetal genome near  
999 FLT1 are associated with risk of preeclampsia. *Nat Genet* 2017;49: 1255-1260.
- 1000 Michita RT, Kaminski VL, Chies JAB. Genetic Variants in Preeclampsia: Lessons From Studies in  
1001 Latin-American Populations. *Front Physiol* 2018;9: 1771.

- 1002 Moore DS, McCabe GP, Craig BA. *Introduction to the Practice of Statistics*, 2009. W.H.  
1003 Freedman, New York.
- 1004 Moore LG, Charles SM, Julian CG. Humans at high altitude: hypoxia and fetal growth. *Respir*  
1005 *Physiol Neurobiol* 2011;178: 181-190.
- 1006 Moore LG, Hershey DW, Jahnigen D, Bowes W, Jr. The incidence of pregnancy-induced  
1007 hypertension is increased among Colorado residents at high altitude. *Am J Obstet Gynecol*  
1008 1982;144: 423-429.
- 1009 Moore LG, Shriver M, Bemis L, Hickler B, Wilson M, Brutsaert T, Parra E, Vargas E. Maternal  
1010 adaptation to high-altitude pregnancy: an experiment of nature--a review. *Placenta* 2004;25  
1011 Suppl A: S60-71.
- 1012 Neilson LI, Schneider PA, Van Deerlin PG, Kiriakidou M, Driscoll DA, Pellegrini MC, Millinder S,  
1013 Yamamoto KK, French CK, Strauss JF, 3rd. cDNA cloning and characterization of a human sperm  
1014 antigen (SPAG6) with homology to the product of the Chlamydomonas PF16 locus. *Genomics*  
1015 1999;60: 272-280.
- 1016 O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, Huang J, Huffman  
1017 JE, Rudan I *et al.* A general approach for haplotype phasing across the full spectrum of  
1018 relatedness. *PLoS Genet* 2014;10: e1004234.
- 1019 Osungbade KO, Ige OK. Public health perspectives on preeclampsia in developing countries:  
1020 Implications for health system strengthening. *J Pregnancy* 2011;2011: 481095.
- 1021 Palmer SK, Moore LG, Young D, Cregger B, Berman JC, Zamudio S. Altered blood pressure  
1022 course during normal pregnancy and increased preeclampsia at high altitude (3100 meters) in  
1023 Colorado. *Am J Obstet Gynecol* 1999;180: 1161-1168.
- 1024 Pappa KI, Roubelakis M, Vlachos G, Marinopoulos S, Zissou A, Anagnou NP, Antsaklis A. Variable  
1025 effects of maternal and paternal-fetal contribution to the risk for preeclampsia combining  
1026 GSTP1, eNOS, and LPL gene polymorphisms. *J Matern Fetal Neonatal Med* 2011;24: 628-635.
- 1027 Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2:  
1028 e190.
- 1029 Phipps EA, Thadhani R, Benzing T, Karumanchi SA. Pre-eclampsia: Pathogenesis, novel  
1030 diagnostics, and therapies. *Nat Rev Nephrol* 2019; 15(5): 275-289.
- 1031 Pilvar D, Reiman M, Pilvar A, Laan M. Parent-of-origin-specific allelic expression in the human  
1032 placenta is limited to established imprinted loci and it is stably maintained across pregnancy.  
1033 *Clin Epigenetics* 2019; 11: 94.

- 1034 Pruijm RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR,  
1035 Willer CJ. LocusZoom: regional visualization of genome-wide association scan results.  
1036 *Bioinformatics* 2010;26: 2336-2337.
- 1037 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker  
1038 PI, Daly MJ *et al.* PLINK: a tool set for whole-genome association and population-based linkage  
1039 analyses. *Am J Hum Genet* 2007;81: 559-575.
- 1040 Purcell S, Sham P, Daly MJ. Parental phenotypes in family-based association analysis. *Am J Hum*  
1041 *Genet* 2005;76: 249-259.
- 1042 Rana S, Lemoine E, Granger JP, Karumanchi SA. Preeclampsia: Pathophysiology, Challenges, and  
1043 Perspectives. *Circ Res* 2019;124: 1094-1112.
- 1044 Salonen Ros H, Lichtenstein P, Lipworth L, Cnattingius S. Genetic effects on the liability of  
1045 developing pre-eclampsia and gestational hypertension. *Am J Med Genet* 2000;91: 256-260.
- 1046 Segura-Vega L. New blood pressure levels in Peruvian high altitude populations and the new  
1047 North American high blood pressure guidelines. *Journal of Cardiology and Current Research*  
1048 2019;12: 84-87.
- 1049 Silva LM, Coolman M, Steegers EA, Jaddoe VW, Moll HA, Hofman A, Mackenbach JP, Raat H.  
1050 Low socioeconomic status is a risk factor for preeclampsia: the Generation R Study. *J Hypertens*  
1051 2008;26: 1200-1208.
- 1052 Sitras V, Paulssen RH, Gronaas H, Leirvik J, Hanssen TA, Vartun A, Acharya G. Differential  
1053 placental gene expression in severe preeclampsia. *Placenta* 2009;30: 424-433.
- 1054 Tan D, Liang H, Cao K, Zhang Q. CUL4A enhances human trophoblast migration and is associated  
1055 with pre-eclampsia. *Int J Clin Exp Pathol* 2017; 10: 1054—10551.
- 1056 Taylor BL, Liu FF, Sander M. Nkx6.1 is essential for maintaining the functional state of  
1057 pancreatic beta cells. *Cell Rep* 2013;4: 1262-1275.
- 1058 Team R Core. R: A language and environment for statistical computing. In *Computing RFFS* (ed).  
1059 2018, Vienna, Austria.
- 1060 Tissot van Patot MC, Murray AJ, Beckey V, Cindrova-Davies T, Johns J, Zwerdinger L, Jauiaux E,  
1061 Burton GJ, Serkova NJ. Human placental metabolic adaptation to chronic hypoxia, high altitude:  
1062 hypoxic preconditioning. *Am J Physiol Regul Integr Comp Physiol* 2009; 298: R166-R172.
- 1063 Tishkoff S. Strength in small numbers. *Science* 2015;349: 1282-1283.
- 1064 Turner SD. qqman: Q-Q and Manhattan Plots for GWAS Data. 2017.

- 1065 Valenzuela FJ, Perez-Sepulveda A, Torres MJ, Correa P, Repetto GM, Illanes SE. Pathogenesis of  
1066 preeclampsia: the genetic component. *J Pregnancy* 2012;2012: 632732.
- 1067 Wang Z, Liu N, Shi S, Liu S, Lin H. The Role of PIWIL4, an Argonaute Family Protein, in Breast  
1068 Cancer. *J Biol Chem* 2016;291: 10646-10658.
- 1069 Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 2 edn, 2016. Springer International  
1070 Publisher, New York.
- 1071 Wikstrom AK, Gunnarsdottir J, Cnattingius S. The paternal role in pre-eclampsia and giving birth  
1072 to a small for gestational age infant; a population-based cohort study. *BMJ Open* 2012;2.
- 1073 Williams PJ, Broughton Pipkin F. The genetics of pre-eclampsia and other hypertensive  
1074 disorders of pregnancy. *Best Pract Res Clin Obstet Gynaecol* 2011;25: 405-417.
- 1075 Wojtowicz A, Zembala-Szczerba M, Babczyk D, Kolodziejczyk-Pietruszka M, Lewaczynska O,  
1076 Huras H. Early- and Late-Onset Preeclampsia: A Comprehensive Cohort Study of Laboratory and  
1077 Clinical Findings according to the New ISHHP Criteria. *Int J Hypertens* 2019;2019: 4108271.
- 1078 Wright D, Nicolaidis KH. Aspirin delays the development of preeclampsia. *Am J Obstet Gynecol*  
1079 2019.
- 1080 Xu Z, Zhang Y, Liu W, Liu Y, Su Y, Xing Q, He X, Wei Z, Cao Y, Xiang H. Polymorphisms of F2,  
1081 PROC, PROZ, and F13A1 Genes are Associated With Recurrent Spontaneous Abortion in Chinese  
1082 Han Women. *Clin Appl Thromb Hemost* 2018;24: 894-900.
- 1083 Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and  
1084 wANNOVAR. *Nat Protoc* 2015;10: 1556-1566.
- 1085 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis.  
1086 *Am J Hum Genet* 2011;88: 76-82.
- 1087 Yong HEJ, Murthi P, Brennecke SP, Moses EK. Genetic Approaches in Preeclampsia. *Methods*  
1088 *Mol Biol* 2018;1710: 53-72.
- 1089 Zadora J, Singh M, Herse F, Przybyl L, Haase N, Golic M, Yung HW, Huppertz B, Cartwright JE,  
1090 Whitley G *et al*. Disturbed Placental Imprinting in Preeclampsia Leads to Altered Expression of  
1091 DLX5, a Human-Specific Early Trophoblast Marker. *Circulation* 2017;136: 1824-1839.
- 1092 Zamudio S. High-altitude hypoxia and preeclampsia. *Frontiers in Bioscience* 2007;12: 2967-2977.
- 1093 Zhou L, Sun H, Cheng R, Fan X, Lai S, Deng C. ELABELA, as a potential diagnostic biomarker of  
1094 preeclampsia, regulates abnormally shallow placentation via APJ. *Am J Physiol Endocrinol*  
1095 *Metab* 2019;316: E773-E781.

1096 Zhou T, Wang H, Zhang S, Jiang X, Wei X. S100P is a potential molecular target of cadmium-  
1097 induced inhibition of human placental trophoblast cell proliferation. *Exp Toxicol Pathol* 2016;68:  
1098 565-570.

1099 Zhu HY, Tong XM, Lin XN, Jiang LY, Wang JX, Zhang SY. Expression and Distribution of Calcium-  
1100 Binding Protein S100P in Human Placenta during Pregnancy. *Int J Fertil Steril* 2015;8: 445-452.

1101

## 1102 **Table legends**

1103

1104 **Table I. All individuals genotyped by group (case/control) and batch after QC filtering.**

1105 **Table II. Phenotypic characteristics of analyzed case families with preeclampsia (duos and**

1106 **trios).** The sum of batch 1 and 2 correspond to the total in the first column. "Trios only"

1107 identifies the subset from the total that are in whole trio units (the rest are mother-offspring

1108 duos). The last column represents chi-squared or t-test p-values for each phenotype between

1109 batches. Significant tests with  $p < 0.05$  are identified with an asterisk (\*).

1110

1111 **Table III. GWAS statistics and genomic annotations of top hits ( $P < 5 \times 10^{-4}$ ) from the TDT.**

1112

1113 **Table IV. GWAS P values and genomic annotations of top hits ( $P < 5 \times 10^{-4}$ ) from the TDT-POO**

1114

1115 **Table V. Statistics and annotations of the top SNPs ( $p < 5 \times 10^{-4}$ ) with biological relevance for**

1116 **preeclampsia of secondary phenotype and case-control GWAS analyses.** All SNPs in this table

1117 are described in the text (for a complete list of regions at  $p < 5 \times 10^{-4}$ , see supplemental tables.

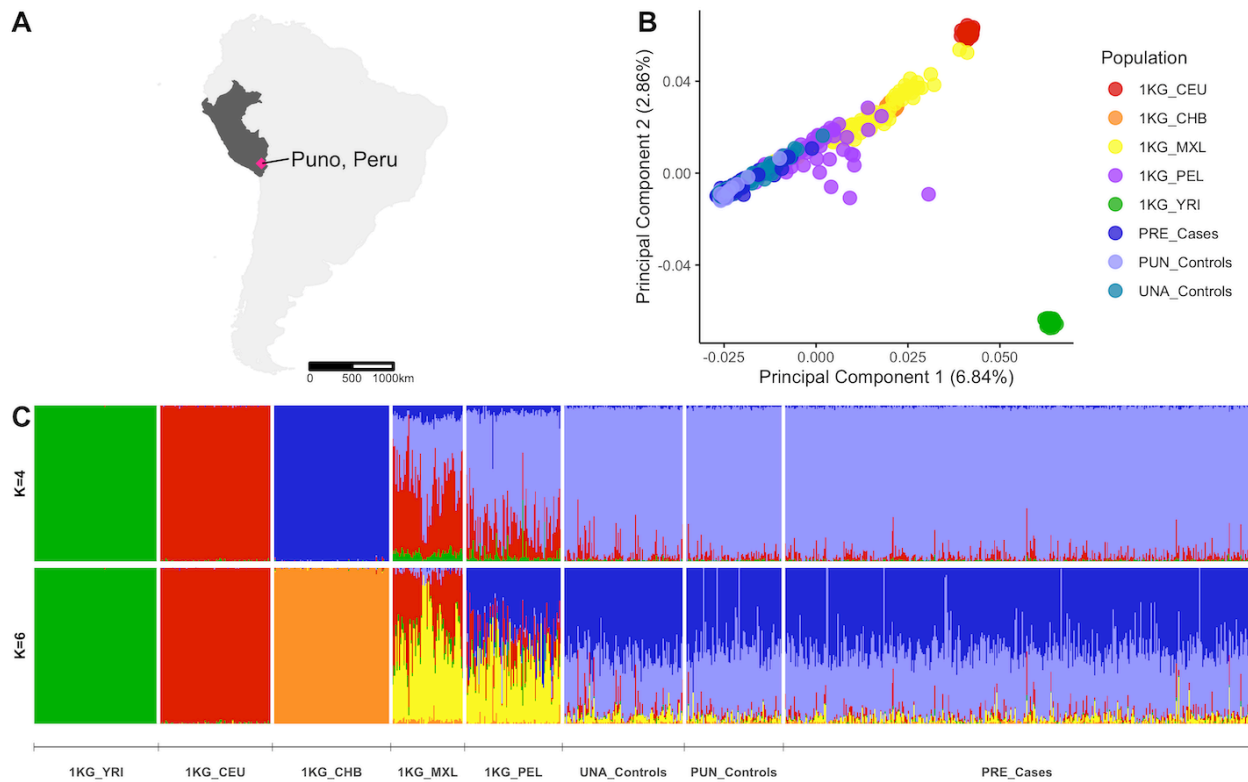
1118 Beta values are reported for linear regressions and odds ratio (OR) for logistic regressions. GA,

1119 gestational age; BP, blood pressure.

1120

1121 **Figures**

1122



1123

1124

1125 **Figure 1. Location and population structure of the Puno preeclampsia cohort.** A) Approximate

1126 location of Puno, Peru. B) Principal components analysis including PRE cases, PUN and UNA

1127 controls, and five continental reference populations from the 1000 Genomes. C) ADMIXTURE

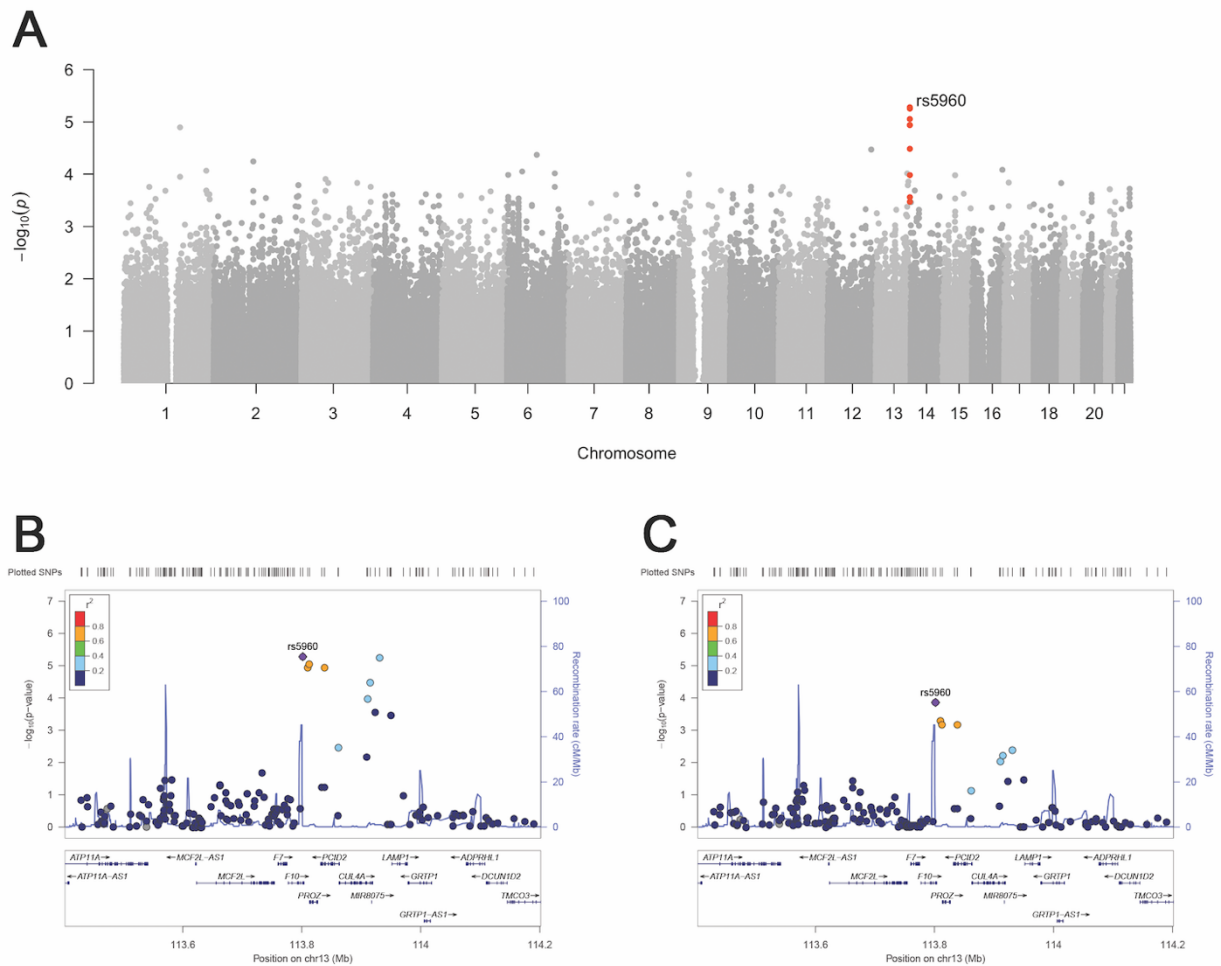
1128 analysis results showing unsupervised clustering models assuming K=4 and K=6. At K=6 a Puno-

1129 specific sub-continental ancestry component not shared with 1000 Genomes Peruvians from

1130 Lima appears in the Puno cohort (shown in light blue).

1131

1132



1133

1134

1135 **Figure 2. Top associations from trio analyses by TDT and TDT-POO.** A) Manhattan plot showing

1136 top association with preeclampsia in the offspring genome: SNP rs5960 on chromosome 13 at

1137  $p < 10e-5$  suggestive of significance (shown in red). B) Locus Zoom plot depicting the top

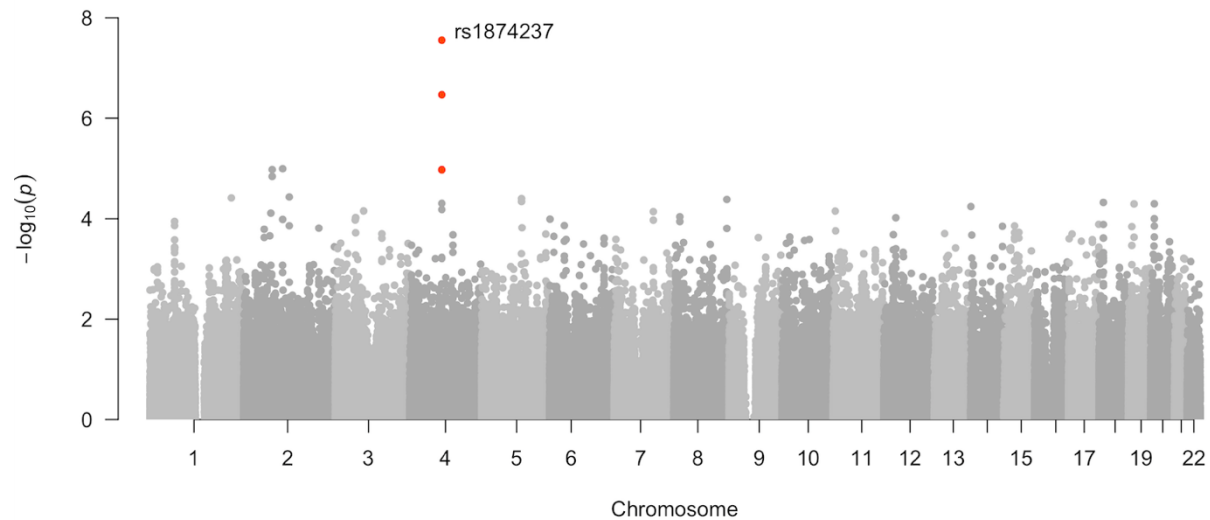
1138 associated SNP cluster from the TDT on chromosome 13. C) Locus Zoom plot depicting the top

1139 paternal region from TDT-POO analysis on chromosome 13.

1140

1141





1142

1143

1144 **Figure 3. Manhattan plot showing top association in the maternal genome with diastolic**

1145 **blood pressure. SNP rs1874237 on chromosome 4 at  $p < 5 \times 10^{-8}$ , genome-wide significance**

1146 (shown in red).

1147

#### 1148 **Data Availability Statement**

1149 The data underlying this article are available in the European Genome-Phenome Archive (EGA)

1150 at <https://ega-archive.org/> and can be accessed with Data Access Committee approval under

1151 Study EGAS00001004625.

1152

#### 1153 **Author notes**

1154 Keyla M. Badillo Rivera and Maria A. Nieves-Colón contributed equally as first authors.

1155 Christopher R. Gignoux, Genevieve L. Wojcik and Andrés Moreno-Estrada contributed equally

1156 as last authors.

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

Family Category <sup>1</sup>	PUN and UNA Controls				PRE Cases				Total No. of Individuals
	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	
<b>Trios <sup>2</sup> (M+F+UC)</b>	<b>0</b>	<b>6</b>	<b>6</b>	<b>2</b>	<b>241</b>	<b>21</b>	<b>262</b>	<b>88<sup>2</sup></b>	<b>268</b>
<b>Duos</b>	<b>10</b>	<b>106</b>	<b>116</b>	<b>58</b>	<b>110</b>	<b>192</b>	<b>302</b>	<b>151</b>	<b>418</b>
<i>M+UC</i>	8	106	114	57	66	188	254	127	368
<i>P+UC</i>	2	0	2	1	18	2	20	10	22
<i>M+F</i>	0	0	0	0	26	2	28	14	28
<b>Singletons <sup>3</sup></b>	<b>106</b>	<b>24</b>	<b>130</b>	<b>--</b>	<b>19</b>	<b>48</b>	<b>67</b>	<b>--</b>	<b>197</b>
<b>All individuals</b>	<b>116</b>	<b>136</b>	<b>252</b>	<b>60</b>	<b>370</b>	<b>261</b>	<b>631</b>	<b>239</b>	<b>883</b>

<sup>1</sup>M= Mother, F=Father, UC=Umbilical cord.

<sup>2</sup>Includes one trio with two offspring from family PRE061.

<sup>3</sup>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).

**Table II. Phenotypic characteristics of analyzed case families with preeclampsia (duos and trios).** The sum of batch 1 and 2 correspond to the total in the first column. “Trios only” identifies the subset from the total that are in whole trio units (the rest are mother-offspring duos). The last column represents chi-squared or t-test p-values for each phenotype between batches. Significant tests with  $p < 0.05$  are identified with an asterisk (\*).

	<b>Total</b>	<b>Batch 1</b>	<b>Batch 2</b>	<b>Trios only</b>	<b>Batch pval</b>
<b>Diagnosis, n</b>	Mild (n=119)	63	56	47	0.68
	Severe (n=106)	60	46	41	
<b>Mode of delivery, n</b>	C-section (n=92)	58	34	42	0.04*
	Vaginal (n=132)	64	68	45	
<b>Maternal age, years mean(median)</b>	26.64(26.00)	26.75(25.00)	26.51(26.00)	27.44(26.00)	0.76
<b>Gestational age - maternal, weeks mean(median)</b>	38.11(39.00)	37.80(38.00)	38.48(39.00)	37.84(38.00)	0.04*
<b>Gestational age - fetal, weeks mean(median)</b>	38.29(39.00)	38.20(39.00)	38.39(39.00)	38.17(39.00)	0.52
<b>Newborn weight, grams mean(median)</b>	601.3(600.0)	611.9(605.0)	569.6(565.0)	620.6(620.0)	0.07
<b>Systolic BP at admission, mmHg mean(median)</b>	133.0(130.0)	131.1(130.0)	135.3(130.0)	130.9(130.0)	0.02*
<b>Diastolic BP at admission, mmHg mean(median)</b>	89.82(90.00)	87.68(90.00)	92.43(90.00)	87.1(85.0)	2.12E-05*
<b>Parity, n</b>	nulliparous (n=128)	72	56	45	0.68
	1 or more (n=97)	51	46	43	
<b>Sex of newborn, n</b>	female (n=100)	46	54	35	0.03*
	male (n=125)	77	48	53	
<b>Proteinuria, n</b>	+ (n=124)	64	60	49	0.38
	++/+++ (n=101)	59	42	39	

**Table III. GWAS statistics and genomic annotations of top hits ( $P < 5 \times 10^{-4}$ ) from the TDT.**

Chr	BP	cytoBand	rsID	Ref	Alt	TDT GWAS stats			Function	Genes in region	Puno MAF
						OR	95% CI	P			
13	113801737	13q34	rs5960	C	T	3.05	1.841-5.054	5.23E-06	exonic	F10	0.4929
13	113930853	13q34	rs9549724	C	T	0.2963	0.1696-0.5176	5.58E-06	intergenic	CUL4A, LAMP1	0.4278
13	113812962	13q34	rs2273971	A	G	0.3276	0.1951-0.55	8.81E-06	upstream	PROZ	0.499
13	113838015	13q34	rs553316	G	A	0.339	0.2041-0.5629	1.15E-05	intronic	PCID2	0.4827
13	113810186	13q34	rs7335409	T	C	2.95	1.777-4.899	1.15E-05	intergenic	F10, PROZ	0.4939
13	113915303	13q34	rs3814260	G	A	0.3396	0.199-0.5797	3.27E-05	intronic	CUL4A	0.4236
2	109581319	2q12.3	rs260692	C	T	19	2.544-141.9	5.70E-05	intronic	EDAR	0.0551
1	228715705	1q42.13	rs11586639	G	A	0.4	0.2492-0.6422	8.57E-05	intergenic	BTNL10, MIR7641-2	0.4587
6	42252385	6p21.1	rs9471831	A	G	0.4242	0.2727-0.6601	8.88E-05	intronic	TRERF1	0.4562
13	113910926	13q34	rs3861723	A	G	0.36	0.2101-0.617	1.04E-04	intronic	CUL4A	0.4228
1	228805855	1q42.13	rs765070	C	T	0.4107	0.2528-0.6673	2.05E-04	intergenic	DUSP5P1, RHOU	0.4347
2	109557099	2q12.3	rs260711	T	C	9.5	2.213-40.78	2.08E-04	intronic	EDAR	0.05612
1	229076157	1q42.13	rs10916389	G	A	0.325	0.1738-0.6076	2.08E-04	intergenic	RHOU, MIR4454	0.2368
13	113923202	13q34	rs77626225	A	G	0.3878	0.2283-0.6586	2.75E-04	intergenic	CUL4A, LAMP1	0.2307
13	113949751	13q34	rs9549380	G	A	0.4	0.2382-0.6718	3.36E-04	intergenic	CUL4A, LAMP1	0.3201
2	109513601	2q12.3	rs3827760	A	G	9	2.088-38.79	3.47E-04	exonic	EDAR	0.05793

**Table IV. GWAS P values and genomic annotations of top hits ( $P < 5 \times 10^{-4}$ ) from the TDT-POO**

Chr	BP	cytoBand	rsID	Ref	Alt	Pval	Function	Genes in region	Puno MAF
<b>Paternal TDT-POO</b>									
<b>13</b>	113801737	13q34	rs5960	C	T	1.38E-04	exonic	F10	0.4929
<b>10</b>	22686205	10p12.2	rs79278805	G	A	1.77E-04	intronic	SPAG6	0.05499
<b>6</b>	142668901	6q24.1	rs9399401	C	T	2.76E-04	intronic	ADGRG6	0.4575
<b>11</b>	115757874	11q23.3	rs4938220	C	T	3.86E-04	intergenic	LINC00900, LOC101929011	0.3585
<b>11</b>	115761165	11q23.3	rs639053	C	T	4.99E-04	intergenic	LINC00900, LOC101929011	0.3344
<b>Maternal TDT-POO</b>									
<b>22</b>	49095071	22q13.32	rs130121	G	A	1.91E-04	intronic	FAM19A5	0.2912
<b>8</b>	98411402	8q22.1	rs10282765	C	T	2.39E-04	ncRNA_intronic	LOC101927066	0.1194
<b>8</b>	98428772	8q22.1	rs2331465	A	G	2.39E-04	ncRNA_intronic	LOC101927066	0.122
<b>22</b>	49099888	22q13.32	rs4925446	C	T	3.86E-04	intronic	FAM19A5	0.2827
<b>8</b>	98432618	8q22.1	rs4588816	C	T	3.93E-04	ncRNA_intronic	LOC101927066	0.122

**Table V. Statistics and annotations of the top SNPs ( $p < 5 \times 10^{-4}$ ) with biological relevance for preeclampsia of secondary phenotype and case-control GWAS analyses.** All SNPs in this table are described in the text (for a complete list of regions at  $p < 5 \times 10^{-4}$ , see supplemental tables. Beta values are reported for linear regressions and odds ratio (OR) for logistic regressions. GA, gestational age; BP, blood pressure.

GWAS	Chr	BP	cytoBand	rsID	Ref	Alt	GWAS stats			Function	Genes in region	Puno MAF
							BETA/OR	95% CI	P			
GA, fetal measurement, maternal genome	1	119404210	1p12	rs952593	T	C	-1.656	-2.273 – -1.039	3.12E-07	intergenic	SPAG17, TBX15	0.13
GA, maternal measurement, maternal genome	1	119425100	1p12	---	A	C	-1.439	-2.013 – -0.8641	1.69E-06	downstream	TBX15	0.14
	11	57203942	11q12.1	rs2581927	C	T	-2.034	-2.887 – -1.182	4.85E-06	intergenic	SLC43A3, RTN4RL2	0.06
Diastolic BP, maternal genome	4	85200613	4q21.23	rs1874237	G	A	-4.257	-5.711 – -2.804	2.79E-08	ncRNA intronic	LOC101928978	0.45
Severity, maternal genome	11	94360812	11q21	rs1940640	T	G	2.407	1.622 – 3.572	1.30E-05	ncRNA intronic	LOC105369438	0.43
Proteinuria, maternal genome	17	2028106	17p13.3	rs2760751	A	G	2.829	1.806 – 4.432	5.65E-06	intronic	SMG6	0.29
	11	94356914	11q21	rs12276362	C	T	0.4146	0.2795 – 0.6148	1.19E-05	ncRNA intronic	LOC105369438	0.49
Systolic BP, fetal genome	2	18099832	2p24.2	rs4553827	C	T	7.443	4.622 – 10.26	5.26E-07	intronic	KCNS3	0.25
Proteinuria, fetal genome	4	5341148	4p16.2	rs62297274	C	T	0.3512	0.2257 – 0.5465	3.51E-06	intronic	STK32B	0.49
	3	25052754	3p24.2	rs4241542	C	T	0.2598	0.1443 – 0.4678	7.04E-06	intronic	RARB	0.21
Case-Control, Offspring	4	6671568	4p16.1	rs34360485	A	G	3.615	2.003 – 6.524	1.99E-05	downstream	LINC02482, S100P	0.36