# Variants in the fetal genome near *FLT1* are associated with risk of preeclampsia

Ralph McGinnis[1*†], Valgerdur Steinthorsdottir[2*†], Nicholas O. Williams[1*], Gudmar Thorleifsson[2], Scott Shooter[1], Sigrun Hjartardottir[3], Suzannah Bumpstead[1], Lilja Stefansdottir[2], Lucy Hildyard[1], Jon K. Sigurdsson[2], John P. Kemp[4,5], Gabriela B. Silva[6,7], Liv Cecilie V. Thomsen[6,8], Tiina Jääskeläinen[9], Eero Kajantie[10,11,12], Sally Chappell[13], Noor Kalsheker[13], Ashley Moffett[14], Susan Hiby[14], Wai Kwong Lee[15], Sandosh Padmanabhan[15], Nigel A. B. Simpson[16], Vivien A. Dolby[16], Eleonora Staines-Urias[17,18], Stephanie M. Engel[19], Anita Haugan[20], Lill Trogstad[20], Gulnara Svyatova[21], Nodira Zakhidova[22], Dilbar Najmutdinova[23], The FINNPEC Consortium[24], The GOPEC Consortium[24], Anna F. Dominiczak[15], Håkon K. Gjessing[20,25], Juan P. Casas[26], Frank Dudbridge[17], James J. Walker[16], Fiona Broughton Pipkin[27], Unnur Thorsteinsdottir[2,28], Reynir T. Geirsson[3], Debbie A. Lawlor[4,29], Ann-Charlotte Iversen[6], Per Magnus[20], Hannele Laivuori[9,30,31], Kari Stefansson[2,28], Linda Morgan[13,†]

[1]Wellcome Trust Sanger Institute, Cambridge, UK. [2]deCODE Genetics/Amgen, Reykjavik, Iceland. [3]Department of Obstetrics and Gynecology, Landspitali University Hospital, Reykjavik, Iceland. [4]MRC Integrative Epidemiology Unit at the University of Bristol, UK. [5]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Australia. [6]Centre of Molecular Inflammation Research (CEMIR) and Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway. [7]St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [8]Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway. [9]Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Finland. [10]National Institute for Health and Welfare, Helsinki, Finland. [11]Children's Hospital, University of Helsinki and Helsinki University Hospital, Finland. [12]PEDEGO Research Unit, Oulu University Hospital and University of Oulu, Finland. [13]School of Life Sciences, University of Nottingham, UK. [14]Department of Pathology, University of Cambridge, UK. [15]BHF Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, University of Glasgow, UK. [16]Leeds Institute of Biomedical and Clinical Sciences, University of Leeds, UK. [17]Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, UK. [18]Nuffield Department of Obstetrics & Gynaecology, University of Oxford, UK. [19]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, USA. [20]Norwegian Institute of Public Health, Oslo, Norway. [21]Scientific Center of Obstetrics, Gynecology and Perinatology, Almaty, Kazakhstan. [22]Institute of Immunology, Uzbek Academy of Sciences, Tashkent, Uzbekistan. [23]Republic Specialized Scientific Practical Medical Centre of Obstetrics and Gynecology, Tashkent, Uzbekistan. [24]A full list of members and affiliations appears at the end of the paper. [25]Department of Global Public Health and Primary Care, University of Bergen, Norway [26]Farr Institute of Health Informatics, University College London, UK. [27]Medical School, University of Nottingham, UK. [28]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [29]School of Social and Community Medicine, University of Bristol, UK. [30]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland. [31]Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Finland.

*R.M, V.S. and N.O.W. contributed equally to this work

† Correspondence should be addressed to R.M. (rm2@sanger.ac.uk), V.S. (Valgerdur.Steinthorsdottir@decode.is) or L.M. (linda.morgan@nottingham.ac.uk)

**Preeclampsia, which affects approximately 5% of pregnancies, is a leading cause of maternal and perinatal death[1]. The causes of preeclampsia remain unclear, but there is evidence for inherited susceptibility[2]. Genome-wide association studies (GWAS) have not identified maternal sequence variants of genome-wide significance which replicate in independent datasets[3,4]. We report the first GWAS of offspring of preeclamptic pregnancies and discovery of the first genome-wide significant susceptibility locus (rs4769613; $P = 5.4x10^{-11}$) in 4380 cases and 310,238 controls. The locus is near the gene encoding Fms-like tyrosine kinase 1 (*FLT1*), providing biological support since an isoform (sFlt-1) of placental origin is implicated in the pathology of preeclampsia[5]. The strongest association is in pregnancies where preeclampsia developed in late gestation and offspring birthweights exceeded the 10th centile. An additional nearby variant, rs12050029, associates with preeclampsia independent of rs4769613. The newly discovered locus may enhance understanding of the pathophysiology of preeclampsia and its subtypes.**

Our initial GWAS meta-analysis tested 7,476,169 sequence variants in 2,658 offspring of preeclamptic pregnancies and 308,292 controls of European descent from Iceland (deCODE cohort) and the UK (GOPEC and ALSPAC cohorts). We observed a single genome-wide significant association ($P$=3.2×10$^{-8}$, rs4769613) located on chromosome 13 near the *FLT1* gene (Fig. 1a). We genotyped rs4769613 and a correlated surrogate in 1722 independent cases and 1946 controls from Norway and Finland along with 26 variants marking GWAS meta-analysis signals elsewhere in the genome whose *P* values showed suggestive evidence of association (Supplementary Table 1). rs4769613 was significantly associated with preeclampsia in the replication datasets ($P$=3.6×10$^{-4}$) and joint analysis of GWAS and replication data yielded robust genome-wide association ($P$=5.4×10$^{-11}$) with an allelic odds ratio (OR) of 1.21 for allele C (Table 1). Forest plots show the frequency of allele C is consistently elevated in cases in all GWAS and replication datasets with no evidence of heterogeneity ($P_{het}$=0.678; Supplementary Fig. 1). None of the other genotyped loci achieved genome-wide significance ($P$<5x10$^{-8}$) in joint analysis of GWAS and replication data.

We then examined genomic features of the *FLT1* locus in detail (Fig. 1b) and found that some of the association signal remained after conditioning out the effect of rs4769613 (Fig 1b, bottom panel). This suggested that other variants near *FLT1* might associate with preeclampsia independent of rs4769613. We therefore genotyped the replication datasets for 21 additional variants at the *FLT1* locus, representing 9 linkage disequilibrium (LD) blocks (Supplementary Table 2; Supplementary Fig. 2). rs12050029 and surrogates in the same LD block were significant in combined analysis of GWAS and replication data after *FLT1* region-wide correction for testing all common variants within 1 Mb of rs4769613 ($P$=3.9×10$^{-6}$, Table 1). Table 1 shows that rs149427560 also achieved *FLT1* region-wide significance but with an association signal weaker than rs12050029. In summary, our results imply that in addition to rs4769613, other independent variants near *FLT1* may modulate preeclampsia susceptibility.

As expected if risk allele rs4769613[C] increases susceptibility by acting through the fetal genome, Supplementary Table 3 shows that allele C frequency in preeclampsia mothers is midway between control frequency and the significantly elevated frequency in preeclampsia offspring. Preeclampsia offspring also preferentially inherited rs4769613[C] from heterozygous parents in the only dataset with DNA available for both parents, again implying that rs4769613[C] increases susceptibility by acting on the fetal genome (Supplementary Table 4). To examine if rs4769613 exerts effects on *both*

the fetal and maternal genomes, we applied the EMIM algorithm which simultaneously evaluates maternal cases, offspring and controls to calculate ORs for preeclampsia risk corresponding to one or two risk alleles carried in the fetus ($R_1$, $R_2$) or in the mother ($S_1$, $S_2$) [6]. Fig. 2a shows that fetal ORs are above 1.0 and that each fetal copy of rs4769613[C] increases these ORs. By contrast, maternal ORs are near 1.0 and are not significant. We conclude that rs4769613 exerts influence primarily through the fetal genome.

As there is evidence that genetic imprinting may operate in placental development[7] we examined rs4769613 allele transmissions from heterozygous parents, but found no parental gender difference in allele transmission in preeclampsia and hence no evidence for imprinting ($\chi^2$=0.046, *P*=0.83; Supplementary Table 5). We also applied EMIM[6] to meta-analyse cohorts with available DNA from one or both parents, but again found no evidence that maternal and paternal alleles at rs4769613 confer differential preeclampsia risk (*P*=0.90).

Sub-classifications of preeclampsia are based on clinical features, in particular gestation at diagnosis and evidence of fetal growth restriction (FGR)[8]. Early-onset preeclampsia (EO-preeclampsia), affecting 12-15% of all preeclamptic pregnancies and defined as onset before 34 weeks gestation, is associated with higher maternal and perinatal mortality than later onset preeclampsia (LO-preeclampsia). It has been proposed that LO-preeclampsia results predominantly from maternal maladaptation to the physiological stresses of pregnancy, whilst EO-preeclampsia is primarily the result of sub-optimal placental implantation into the uterine wall, leading to inadequate placental perfusion and the release of damaging placental factors into the maternal circulation[8]. In keeping with this, EO-preeclampsia is frequently associated with FGR, resulting in babies who are small for gestational age (SGA) at birth. SGA defined as birthweight <10th centile is widely used as a surrogate for FGR.

To assess the impact of rs4769613 on gestation at onset and fetal growth subtypes we noted that rs4769613 risk allele C had higher frequency in LO-preeclampsia than EO-preeclampsia cases (case-control OR 1.23 vs. 1.06) and found the difference was significant in case-case meta-analysis (*P*=0.017, Fig. 2b). Similarly, allele C had higher frequency in nonSGA-preeclampsia than SGA-preeclampsia cases (case-control OR 1.25 vs 1.10) and the difference was significant in case-case comparison (*P*=0.019, Fig. 2b). Further division of cases in Fig 2b into the four possible subcategories found that rs4769613 [C] confers greatest risk to LO+nonSGA cases (case-control OR=1.26, $P=1.2\times10^{-7}$) and least risk to EO+SGA cases (case-control OR=1.03, *P*=0.72) with case-case comparison of the two subcategories being significant ($P=5.8\times10^{-3}$). In summary, the results indicate that rs4769613 exerts its greatest influence in pregnancies where preeclampsia develops in late gestation and birthweights exceed the 10$^{th}$ centile. rs12050029 was also associated with LO-preeclampsia, but the strength of the association did not differ between SGA- and nonSGA-preeclampsia (Supplementary Fig. 3).

*FLT1* encodes a trans-membrane tyrosine kinase receptor Flt-1 that mediates angiogenesis promoted by binding vascular endothelial growth factor (VEGFA) and placental growth factor (PlGF)[9]. The alternatively spliced soluble isoform sFlt-1 antagonizes angiogenesis by also binding VEGFA and PlGF. During pregnancy, *FLT1* is mainly expressed in fetal trophoblasts which release sFlt-1 as the most abundant isoform into the maternal circulation. The excessive release of sFlt-1 in

preeclampsia appears to mediate widespread maternal endothelial dysfunction, manifesting as hypertension, proteinuria, and vascular compromise to major organs. High sFlt-1 and low PlGF concentrations are established markers of EO-preeclampsia[8], but our evidence that *FLT1* polymorphisms are strongly associated with LO-preeclampsia suggests trophoblast function is also important in this preeclampsia subgroup.

The signals around rs4769613 and rs12050029 are both located in placental enhancer regions (Fig. 1b) suggesting a mechanism by which variants could affect *FLT1* expression. We explored possible association between fetal *FLT1* genotype and protein expression by placental immunohistochemistry and intensity scanning in 37 preeclamptic and 44 control pregnancies. There was no detectable association between fetal rs4769613 genotype and trophoblast Flt-1 and sFlt-1 expression in cases ($P$=0.47) or controls ($P$=0.26). We also compared maternal serum sFlt-1 from the first or third trimester with fetal rs4769613 genotype in mother-baby pairs from 242 control and 276 preeclamptic pregnancies. Control pregnancies exhibited a trend towards increasing maternal serum sFlt-1 levels with each fetal copy of rs4769613[C] in the third trimester, which reached nominal significance ($P$ = 0.04), while in case pregnancies the levels were higher ($P$<0.001) but with no detectable difference between genotype groups ($P$=0.47) (Fig. 2c). We did not have suitable placental tissue for mRNA studies, but the Genotype-Tissue expression (GTEx) database (www.gtexportal.org) does not provide evidence for rs4769613 or rs12050029 allele-specific differences in *FLT1* expression in 42 tissues, although data for placental tissue are not recorded. The evidence that fetal rs4769613 genotype affects maternal serum levels of sFlt-1 is therefore modest. Subtle changes in sFlt-1 concentration driven by fetal *FLT1* genotype, as suggested by the data from control pregnancies, may be masked in preeclampsia, where the overall levels are already high. Also, this effect is minimal compared to the increase in serum sFlt-1 seen in preeclamptic pregnancies so it may not reflect the role of the preeclampsia associated variants in the pathophysiology of preeclampsia.

We explored whether the preeclampsia associated variants affected other diseases or traits by using the deCODE database of common diseases and traits routinely measured at hospitals and clinical laboratories (see Methods). Given that rs4769613 and rs12050029 are not in LD (Supplementary Fig. 2), it is noteworthy that the only significant database association for both variants was red blood cell (RBC) count ($P$=5.0 x $10^{-4}$ and $P$=1.5 x $10^{-7}$ for rs4769613 and rs12050029 respectively), where the preeclampsia risk allele consistently associated with reduced RBC count (Supplementary Table 6). The RBC association with *FLT1* is intriguing since its VEGF ligand has previously been implicated in regulation of erythropoiesis, but the mode and sites of action are complex[10,11]. Our preeclampsia results suggest that in the fetus the two variants lead to an increase in sFlt-1, while the same alleles are associated with reduced RBC count in the general (non-pregnant) population. The effect on both preeclampsia and RBC is consistent with the variants acting through the neighbouring *FLT1*.

We note that SNP rs4769613 is located between *FLT1* and *POMP,* which encodes proteasome maturation protein, a ubiquitously expressed protein involved in proteasome assembly and MHC class I antigen presentation[12]. We cannot exclude the possibility that sequence variants at this locus affect expression of *POMP* or more distant genes, but the GTEx database does not provide any evidence to support this contention.

Evidence presented here implies that altered trophoblastic *FLT1* expression is not merely a secondary consequence of placental pathology in preeclampsia, but is central to its aetiology. A role for fetal sequence variants in susceptibility to preeclampsia is consistent with patterns of inheritance implicating both maternal and paternal factors[2]. The fetal *FLT1* gene has been indirectly implicated previously in pregnancies with fetal trisomy 13, which are associated with increased placental expression of sFlt-1, and an increased incidence of preeclampsia[13]. sFlt-1 is a marker of placental malfunction, a hallmark of EO-preeclampsia[8]; our observation that *FLT1* genotype is associated even more strongly with LO-preeclampsia implies that placental pathology is also a feature of late-onset disease. The variants we describe provide tools for experimental testing of whether, how, when and where they affect *FLT1* expression, and how this relates to the pathophysiology of preeclampsia and its subtypes. The discovery of sequence variants in the genome of the fetus that increase the risk of disease in the mother is an ultimate demonstration of the closeness of the remarkable symbiosis we call pregnancy.

**Acknowledgements**

Research leading to these results was conducted as part of the InterPregGen study, which received funding from the European Union Seventh Framework Programme under grant agreement no. 282540, and was supported by the Wellcome Trust grant 098051. The UK Medical Research Council, the Wellcome Trust (102215/2/13/2) and the University of Bristol provide core support for ALSPAC, with additional funds for this study from UK Medical Research Council (MC_UU_1201/5) and European Research Council (669545) (D.A.L; J.P.K.) The GOPEC collection was funded by the British Heart Foundation Programme Grant RG/99006. The Norwegian Mother and Child Cohort Study (MoBa) is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research, NIH/NIEHS (contract no N01-ES-75558), NIH/NINDS (grant no.1 UO1 NS 047537-01 and grant no.2 UO1 NS 047537-06A1) (P.M.) MoBa GWAS studies were supported in part by NICHD Grant R01HD058008 (P.M; S.M.E.) The FINNPEC study was supported by Jane and Aatos Erkko Foundation, Päivikki and Sakari Sohlberg Foundation, Academy of Finland, Research Funds of the University of Helsinki, government special state subsidy for health sciences (Erityisvaltionosuus funding) at the Hospital District of Helsinki and Uusimaa, Novo Nordisk Foundation, Finnish Foundation for Pediatric Research, Emil Aaltonen Foundation, and Sigrid Jusélius Foundation. The Preeclampsia Study is supported by the Research Council of Norway (205400/V50 and 223255/F50) and the Liaison Committee between the Norwegian University of Science and Technology and the Central Norway Regional Health Authority (A-C.I.)

This research makes use of data generated by the Wellcome Trust Case Control Consortium (WTCCC). A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for WTCCC, WTCCC2 and WTCCC3 was provided by the Wellcome Trust under awards 076113, 083948/Z/07/Z and 088841/Z/09/Z. This research also makes use of GWAS data from the ALSPAC study generated by G. Hemani and G. McMahon. The ALSPAC study website contains details of all the data that are available through a fully searchable data dictionary at the following webpage: www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/

We are grateful to all the families from Iceland, Norway, Finland and the UK who took part in this study, and the work of teams of volunteers, managers, midwives, nurses, doctors, computer and

laboratory technicians, clerical workers, research scientists and receptionists who made this study possible.

**Author contributions**

Manuscript preparation: R.M, V.S, N.O.W, L.C.V.T, A-C.I, L.M. All authors contributed to critical analysis and revision of the manuscript.
Study design: R.M, V.S, N.O.W, T. J, S.l, N.K, N.A.B.S, V.A.D, E.S-U, S.M.E, A.H, L.T, G.S, N.Z, D.N, A.F.D, H.K.G, J.P.C, F.D, J.J.W, F.B.P, U.T, D.A.L, A-C.I, P.M, H.L, K.S, L.M.
Phenotyping: S.H, G.B.S, L.C.V.T, T.J, E.K, FINNPEC Consortium, GOPEC Consortium, J.J.W, F.B.P, R.T.G, D.A.L, H.L, L.M.
Genotyping and quality control: S.S, S.B, L.H, W.K.L, S.P, U.T, H.L.
GWAS data analysis: R.M, V.S, N.O.W, G.T, S.S, L.S, J.K.S, J.P.K.
Immunohistochemistry: G.B.S, L.C.V.T, A-C.I.
Biomarker measurement: H.L.
Interpretation: R.M, V.S, N.O.W, G.T, J.P.K, L.C.V.T, T.J, E.K, S.C, N.K, N.S, V.D, E.S-U, L.T, H.K.G, F.D, J.J.W, F.B.P, D.A.L, A-C.I, P.M, H.L, K.S, L.M.

**Competing financial interests statement**

Valgerdur Steinthorsdottir, Gudmar Thorleifsson, Lilja Stefansdottir, Jon K. Sigurdsson, Unnur Thorsteinsdottir and Kari Stefansson are employees of the biotechnology firm deCODE Genetics, a subsidiary of Amgen. Debbie A. Lawlor has received industry funding for biomarker research unrelated to this paper from Medtronic, Roche Diagnostics, and Ferring Pharmaceuticals.

**Consortia**

**GOPEC Consortium**: Linda Morgan[13], Fiona Broughton Pipkin[27], Noor Kalsheker[13], James. J. Walker[16], Sheila Macphail[32], Mark Kilby[33], Marwan Habiba[34], Catherine Williamson[35], Kevin O'Shaughnessy[36], Shaughn O'Brien[37], Alan Cameron[38], Lucilla Poston[35], Zofia Miedzybrodzka[39], Christopher W. G. Redman[18], Martin Farrall[40], Mark Caulfield[41], Anna F. Dominiczak[15]

**FINNPEC Consortium**: Hannele Laivuori[9,30,31], Seppo Heinonen[31], Eero Kajantie[10,11,12], Juha Kere[42,43], Katja Kivinen[44], Anneli Pouta [10,45]

[32]Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. [33]The Centre for Women's & Newborn Health, College of Medical and Dental Sciences, University of Birmingham, UK. [34]University of Leicester, Leicester, UK. [35]Division of Women's Health, Kings College London, UK. [36]Department of Medicine, University of Cambridge, UK. [37]Keele University School of Medicine, Stoke-on-Trent, UK. [38] Institute of Cardiovascular and Medical Sciences, University of Glasgow, UK. [39]College of Life Sciences and Medicine, University of Aberdeen, UK. [18]Nuffield Department of Obstetrics and Gynaecology, University of Oxford, UK. [40]Radcliffe Department of Medicine, University of Oxford, UK. [41]William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, London, UK. [42]Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden. [43]Molecular Neurology Research Program, Research Programs Unit, Helsinki, Finland. [44]Division of Cardiovascular Medicine, University of Cambridge, UK. [45]Department

of Obstetrics and Gynaecology and MRC Oulu, Oulu University Hospital and University of Oulu, Finland.

**References**

1. Souza, J. P. *et al.* Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): a cross-sectional study. *Lancet* **381**, 1747-1755 (2013).
2. Cnattingius, S. Reilly, M. Pawitan, Y. & Lichtenstein, P. Maternal and fetal genetic factors account for most of familial aggregation of preeclampsia: a population-based Swedish cohort study. *Am. J. Med. Genet. A.* **130A,** 365-371 (2004).
3. Johnson, M. P. *et al.* Genome-wide association scan identifies a risk locus for preeclampsia on 2q14, near the inhibin, beta B gene. *PLoS One* **7**, e33666. doi: 10.1371/journal.pone.0033666 (2012).
4. Zhao, l. *et al.* Genome-wide association study identifies a maternal copy-number deletion in PSG11 enriched among preeclampsia patients. *BMC Pregnancy Childbirth* **12**, 61 (2012).
5. Maynard, S. E. *et al*. Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. *J. Clin. Invest.* **111**, 649-658 (2003).
6. Ainsworth, H. F. *et al.* Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. *Genet. Epidemiol.* **35,** 19-45 (2011).
7. Monk, D. Genomic imprinting in the human placenta. *Am. J. Obstet. Gynecol.* **213,** S152-S162 (2015).
8. Staff A. C. *et al.* Redefining preeclampsia using placenta-derived biomarkers. *Hypertension* **61**, 932-942 (2013).
9. Cerdeira, A. S. & Karumanchi, S. A. Angiogenic factors in preeclampsia and related disorders. *Cold Spring Harb. Perspect. Med.* pii: a006585. doi: 10.1101/cshperspect.a006585 (2012).
10. Tam, B. Y. *et al.* VEGF modulates erythropoiesis through regulation of adult hepatic erythropoietin synthesis. *Nat. Med.* **12,** 793-800 (2006).
11. Rehn, M. *et al.* Hypoxic induction of vascular endothelial growth factor regulates murine hematopoietic stem cell function in the low-oxygenic niche. *Blood* **118,** 1534-1543 (2011).
12. Heink, S. Ludwig, D. Kloetzel, P. M. Krüger, E. IFN-gamma-induced immune adaptation of the proteasome system is an accelerated and transient response. *Proc. Natl. Acad. Sci. USA.* **102,** 9241-9246 (2005).
13. Bdolah, Y. *et al.* Circulating angiogenic proteins in trisomy 13. *Am. J. Obstet. Gynecol.* **194,** 239-245 (2006).

**Figure legends**

**Figure 1 | Manhattan Plots showing GWAS results across all autosomes and detailed view near *FLT1* on chromosome 13**.

a) Genome-wide Manhattan plot showing strength of association with PE in GWAS meta-analysis plotted as -log10(*P* value) on the y-axis and corresponding variant position on the x-axis. A single peak whose apex is the sentinel SNP rs4769613 near *FLT1* on chromosome 13 crosses the blue line denoting genome-wide significance ($P = 5\times10^{-8}$). Variants within 100 Kb of rs4769613 are coloured purple.

b) Detailed view near *FLT1* highlighting variants in Table 1. Panels from top to bottom show: unconditional -log10(*P* value) from GWAS meta-analysis and recombination rate shown as a blue line quantified by right hand y-axis; pattern of regional Linkage Disequilibrium (LD) shown by pairwise values of the LD metric D'; gene names with approximate length and position; the corresponding chromatin state annotations for selected Epigenome Roadmap tissues (http://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state); landscape of inferred chromatin interactions for 31 tissue types using an integrated method for predicting enhancer targets (IM-PET) (http://4dgenome.research.chop.edu/); conditional -log10(*P* value) for GWAS meta-analysis using logistic regressions with rs4769613 as a covariate.

**Figure 2 | Key observations about rs4769613 in relation to preeclampsia.**

a) Forest plot showing Odds Ratio (OR) and 95% confidence intervals (95% CI) calculated by the EMIM algorithm[6] for preeclampsia risk conferred by one or two copies of rs4769613 risk allele C carried in the fetus ($R_1$, $R_2$) or carried in the mother ($S_1$, $S_2$). Individual datasets (GOPEC, MoBa, FINNPEC) and meta-analysis across the datasets ([Meta]) show the OR is increased by each fetal copy of risk allele C where for $R_1$ $P$=3.7x10$^{-4}$ and for $R_2$ $P$=3.7x10$^{-9}$. By contrast, the OR for maternal copies ($S_1$, $S_2$) are not significantly different from 1 implying that, after accounting for fetal copies, maternal copies of allele C confer no additional increased risk of preeclampsia.

b) Forest Plot for preeclampsia subtypes defined by early and late onset (EO-PE, LO-PE) and by birthweight that is small-for-gestational age (SGA-PE) or not (nonSGA-PE). Case-case comparisons in blue font show risk allele C is more significantly associated with LO-PE and nonSGA-PE than with EO-PE and SGA-PE. Dividing cases into the four possible subcategories found that allele C confers greatest risk to LO+nonSGA cases and least risk to EO+SGA cases.

c) Box-and-whisker plots of first and third trimester maternal serum sFlt-1 concentration in preeclampsia cases and controls, showing the effect of fetal rs4769613 genotype. Boxes span the first to the third quartile of sFlt-1 concentration, with horizontal bars within the box denoting the median; whiskers extend to the 10$^{th}$ and 90$^{th}$ centiles. Maternal sFlt-1 is higher in cases than controls in the third trimester across all fetal genotypes (*t*-test ln(sFlt-1): t=7.79; 200 d.f.; 2-tailed $P$< 0.001). In third trimester controls, each copy of the rs4769613[C] allele carried by the fetus is associated with an increase in maternal sFlt-1 (linear regression of ln(sFlt) with SNP genotype (coded 0, 1 and 2) and gestational age as covariates: $P$=0.04).

**Table 1 Meta-analysis results at three independent variants near *FLT1* giving evidence for association with preeclampsia**

| | | | | | GWAS N=2,658 / 308,292 | | | Replication N=1,722 / 1,946 | | | GWAS+Replication N=4,380 / 310,238 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variant | Chr13 Position | Risk/Alt Allele | RAF | Covariate | OR | 95%CI | *P* | OR | 95%CI | *P* | OR | 95%CI | *P* |
| **rs4769613** | **29138609** | **C/T** | **0.53** | **None** | **1.22** | **1.14-1.31** | **$3.2\times10^{-8}$** | **1.18** | **1.08-1.30** | **$3.6\times10^{-4}$** | **1.21** | **1.14-1.28** | **$5.4\times10^{-11}$** |
| rs12050029 | 29227519 | G/A | 0.14 | None | 1.20 | 1.09-1.33 | $1.5\times10^{-4}$ | 1.18 | 1.05-1.32 | $5.9\times10^{-3}$ | 1.19 | 1.11-1.28 | $3.0\times10^{-6}$ |
| rs149427560 | 29105870 | G/GGT | 0.06 | None | 1.30 | 1.14-1.49 | $9.3\times10^{-5}$ | 1.16 | 0.96-1.40 | $1.1\times10^{-1}$ | 1.23 | 1.09-1.38 | $4.1\times10^{-5}$ |
| rs12050029 | 29227519 | G/A | 0.14 | rs4769613 | 1.19 | 1.09-1.32 | $2.9\times10^{-4}$ | 1.18 | 1.05-1.33 | $4.2\times10^{-3}$ | 1.19 | 1.11-1.28 | $3.9\times10^{-6}$ |
| rs149427560 | 29105870 | G/GGT | 0.06 | rs4769613 | 1.31 | 1.15-1.50 | $6.7\times10^{-5}$ | 1.17 | 0.97-1.41 | $9.1\times10^{-2}$ | 1.23 | 1.10-1.38 | $2.4\times10^{-5}$ |

Results are ordered by strength of association with preeclampsia in GWAS+Replication meta-analysis at three variants near *FLT1* not in linkage disequilibrium. Row for rs4769613 is bold because its GWAS+Replication *P* value is genome-wide significant ($p<5\times10^{-8}$). rs12050029 and rs149427560 are included in the table because their GWAS+Replication *P* values are below *FLT1* region-wide significance threshold ($6.01\times10^{-5}$) calculated by the method of Gao (see Main Text and Methods). "N", total cases / controls in meta-analysis; "Chr13 Position", NCBI Build 37 position on chromosome 13; "Risk/Alt Allele", allele with higher frequency in cases than controls and alternate allele; "RAF", risk allele frequency in UK GWAS controls; "Covariate", covariate in conditional logistic regression; "OR" and "95%CI", allelic odds ratio and 95% confidence interval; "*P*", *P* values of case-control association. Genotypes of rs149427560 for the FINNPEC cohort were proxied from rs11619261 (pairwise $r^2=0.88$ in Finland 1000Genomes phase 3).

## Methods

### Cohorts

Three European cohorts of offspring from pregnancies affected by preeclampsia provided cases for GWAS meta-analysis: the GOPEC and ALSPAC cohorts from the UK, and the Icelandic deCODE cohort. Two independent cohorts were used for replication genotyping: the Finnish FINNPEC collection, and the Norwegian Mother and Child Cohort Study (MoBa). Recruitment criteria were not identical in all cohorts, so subsets were selected for this study based on an internationally recognised definition of preeclampsia[14]: new-onset hypertension after the 20$^{th}$ week of gestation, with systolic blood pressure ≥140mmHg or diastolic blood pressure ≥90mmHg on at least two occasions; and new-onset proteinuria of 0.3g/24 hours or more, or ≥1+ on dipstick analysis of urine. All were singleton pregnancies in a white Western European woman; preeclamptic pregnancies in women with a previous history of essential hypertension, type 1 or type 2 diabetes mellitus, ischaemic heart disease, cerebrovascular accident or chronic renal disease were excluded. Phenotypic details are summarised in Supplementary Tables 7 and 8.

Written informed consent was obtained from participants, or from parents on behalf of minors, and all studies were approved by local Research Ethics Committees.

*GOPEC (Genetics of Pre-eclampsia)*

The UK GOPEC collection includes 1157 DNA samples from mother-baby pairs with preeclampsia recruited at diagnosis between 1992 and 2009 for genetic studies of preeclampsia[15]. Control data were derived from the WTCCC2 genome-wide analysis of 2930 samples from the 1958 Birth Cohort and 2737 samples from the National Blood Services, providing control data for 5297 individuals after QC[16].

*ALSPAC (Avon Longitudinal Study of Parents and Children)*

The ALSPAC prospective birth cohort study recruited pregnant women living in the South West of England between 1991 and 1992, and has been described elsewhere[17]. They included 13,678 singleton pregnancies resulting in a live birth. After exclusion of women with existing hypertension, diabetes and gestational diabetes, there were 7382 pregnancies for which blood pressure and proteinuria measurements and fetal GWAS data were available. Of these, 146 met the definition of preeclampsia and were included. The control group of 6130 subjects was derived from all other included pregnancies with fetal GWAS data in women not affected by essential hypertension or gestational hypertension.

*deCODE Pre-eclampsia cohort*

The deCODE preeclampsia cohort is part of an ongoing sample collection including a large part of the Icelandic population. Preeclamptic pregnancies occurring between 1970 and 2009 were identified through scrutiny of hospital records at the Landspitali University Hospital, which provides secondary and tertiary services for the whole of Iceland. Initially, a group of women with hypertensive disease in pregnancy (ICD-9:642.0–9 and ICD-10: O10–16) in the years 1984-1999 were selected for further study based on familial relationships. All maternity records for these women were scrutinised and each affected pregnancy reclassified[18]. This identified 491 singleton preeclamptic pregnancies. Preeclamptic pregnancies from 2000-2009 were identified based on ICD-10 codes (O14-15), yielding 1,311 additional singleton preeclamptic pregnancies. Overall information on 1,802 singleton

preeclamptic pregnancies of 1,662 mothers is available. GWAS data was available from 1507 offspring of these pregnancies, identified retrospectively based on the national register. The control group comprises 296,865 individuals from the deCODE sample collection.

*MoBa (Norwegian Mother and Child Cohort Study)*

The Norwegian Mother and Child Cohort Study is a longitudinal study of over 110,000 pregnant women, their children and partners, recruited between 1999 and 2008 from maternity units throughout Norway[19]. 1200 pregnancies affected by preeclampsia were identified from Medical Birth Register of Norway records; the validity of the diagnosis has been assessed by retrieval and examination of antenatal records. 1200 non-hypertensive pregnancies provided the control group. Pregnancies were excluded from case and control groups if a maternal history of essential hypertension, chronic renal disease or diabetes mellitus was recorded in the Medical Birth Registry of Norway.

*FINNPEC (Finnish Genetic of Pre-eclampsia Consortium)*

The FINNPEC collection was assembled in Finland between 2008 and 2011 from two recruitment arms[20]. Samples were collected at the time of diagnosis of preeclampsia from 879 mothers, and during pregnancy from 922 non-pre-eclamptic mothers from antenatal and labour wards. Their children and partners were also enrolled. A further 525 pregnancies affected by preeclampsia were identified by examination of hospital records, and women and offspring were invited to participate by letter. After exclusion of pregnancies which did not meet the entry criteria for this study, offspring of 605 preeclamptic pregnancies were included as cases, and offspring of 800 non-hypertensive pregnancies provided the control group.

**Genotyping, quality control, genotype imputation and association analysis in GWAS datasets**

*GOPEC*

1157 offspring of preeclamptic pregnancies were assayed on the Illumina OmniExpress chip; maternal samples where available were similarly genotyped. A total of 730,525 variants were called with the GenCall algorithm. We carried out QC using PLINK (http://pngu.mgh.harvard.edu/~purcell/anal.shtml) and SMARTPCA[21]. A subset of the samples (186) were whole genome amplified (WGA). WGA genotype calls can be prone to calling artefacts. To address this we removed variants with either low call rate (95%) or Mendelian errors in the WGA samples and we then performed a pseudo-case control analysis of WGA vs. non-WGA and removed variants with significant genotypic association ($P$<0.001). We then applied standard QC to the combined WGA and non-WGA dataset on this reduced set of variants (670,435). Briefly, standard QC comprises the following subject level exclusion criteria: individual call-rate < 95%; heterozygosity >3 s.d. from the mean; any of the first 3 HapMap (based on CEU, YRI, CHB, JPT and GIH) principal axes of variation >4 s.d. from the mean and gender mismatch. Related individuals (IBD>0.1) with lowest call-rate were preferentially removed. The variant level exclusion criteria are: call-rate <95%, exact Hardy-Weinberg equilibrium $P<1 \times 10^{-6}$, minor allele frequency (MAF) <1% and non-random missingness of uncalled genotypes ("plink --test-mishap") with Bonferroni corrected $P$<0.05. These filters left 1005 samples (89 WGA) and 574,919 variants.

We used WTCCC2 population controls from the National Blood Donors Cohort and UK 1958 Birth Cohort[16]. These samples were genotyped on the Illumina 1.2M chip and called using GenCall. Strand ambiguous markers were removed and the standard QC described above was then applied to the two control datasets. The merged control datasets consisted of 5,297 samples and 438,912 variants. This control dataset was merged with the case dataset resulting in 429,754 post QC variants that were genotyped in both cases and controls.

Cases and controls were imputed together with IMPUTE2 (impute_v2.3.0)[22] and SHAPEIT[23] using the pre-phasing workflow against the 1000 Genomes Phase 1 reference panel (Dec. 2013) downloaded from the IMPUTE2 website. Imputation resulted in 10,404,388 bi-allelic variants with MAF > 0.25% that were either directly genotyped or imputed with IMPUTE2 INFO score >0.6.

Post imputation association analysis was carried out using SNPTEST (v2.4.1)[22] with the "expected" method with no ancestry principal components. We calculated the genomic control on variants with MAF>0.5% as $\lambda_{GC}$ =1.005.

*deCODE*

Details of GWAS genotyping, QC and imputation of the Icelandic dataset including the preeclampsia cases and controls used in this study have been described[24]. Briefly, samples were assayed with the Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni-1, Omni 2.5 or Omni Express bead chips at deCODE genetics. Following QC a final set of 676,913 autosomal SNPs were used for long range phasing of all chip-genotyped samples. Making use of the Icelandic genealogy untyped first and second degree relatives of chip-typed individuals were also included in the analysis to increase power[24]. In total 104,220 chip-typed individuals and 294,212 of their untyped relatives were imputed based on a panel of sequence variants identified through whole genome sequencing of 2,636 Icelanders to a mean depth of 20x.

GWAS analysis of Icelandic preeclampsia offspring included a total of 1,507 cases (380 chip typed) and 296,865 controls (91,326 chip-typed). The controls used in this study were Icelandic individuals from other ongoing GWAS studies at deCODE and their relatives. Logistic regression was used to test for association between sequence variants and disease, treating disease status as the response and genotype counts as covariates. Other characteristics also included in the model as nuisance variables were: sex, county of birth, current age or age at death (first and second order terms included), genotyping status and an indicator function for the overlap of the lifetime of the individual with the timespan of phenotype collection[25]. In order to account for relatedness and stratification within the case and control sample sets we applied the method of genomic control[25]. Based on a set of about 300,000 common variants distributed across the genome the inflation in the chi-squared statistic for preeclampsia offspring was estimated to be 1.115.

*ALSPAC*

A total of 9,912 ALSPAC children were genotyped using the Illumina HumanHap550 quad genome-wide SNP genotyping platform (Illumina Inc., San Diego, CA, USA) by Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and Laboratory Corporation of America (LabCorp Holdings., Burlington, NC, USA). PLINK software (v1.07) was used to carry out quality control measures. Individuals were excluded from further analysis on the basis of having incorrect gender

assignments, minimal or excessive heterozygosity (< 0.320 and > 0.345 for the Sanger data and < 0.310 and > 0.330 for the LabCorp data), disproportionate levels of individual missingness (> 3%) and being of non-European ancestry (as detected by a multidimensional scaling analysis seeded with HapMap 2 individuals). EIGENSTRAT analysis revealed no additional obvious population stratification and genome-wide analyses with other phenotypes in the same cohort indicate a low lambda. SNPs with a minor allele frequency of < 1% and call rate of < 95% were removed. Furthermore, only SNPs that passed an exact test of Hardy–Weinberg equilibrium ($P > 5\times10^{-7}$) were considered for analysis. Related subjects (> 10% IBD) that passed all other quality control thresholds were retained during subsequent phasing and imputation. 9,115 subjects and 500,527 SNPs passed these quality control filters.

We combined 477,482 SNP genotypes in common between the sample of children and mothers. Genotyping and QC of the ALSPAC mothers can be found elsewhere[26]. SNPs with genotype missingness > 1% and those that failed the exact test of HWE were removed. A further 321 participants were removed due to potential ID mismatches (IBD < 1). The resultant dataset comprised 17,842 subjects of which 6,305 were mother-offspring pairs. An additional 112 SNPs were removed after a liftover of the merged genotyped data from Hg 18 to Hg19. Haplotype phasing was performed using SHAPEIT (v2.r644)[23] and known autosomal variants were imputed with IMPUTE V2.2.2[22] using the 1000 genomes reference panel (Phase 1, Version 3) consisting of 2186 reference haplotypes (including non-Europeans).

Logistic regression, as implemented in SNPTEST v2.5-beta4[22], was used to test for association between imputed genotype probabilities and disease status. Based on a set of ~9 million SNPs (MAF>0.5% and IMPUTE2 INFO>0.6), no evidence of genomic inflation was observed ($\lambda_{GC}$=1.008).

**Follow-up genotyping**

Follow-up variants in the *FLT1* locus were chosen to test association with the rs4769613 peak and to test possible association in the other 8 *FLT1* LD blocks shown in Supplementary Fig. 2. Some genotyped variants were highly correlated surrogates of rs4769613 or of other follow-up SNPs in case the assay for the primary variant failed, and to ensure that assertion of a true-positive association did not rely on genotyping of a single variant. Follow-up variants in non-*FLT1* regions of the genome were chosen based on GWAS meta-analysis *P* value and were selected to further test suggestive evidence of association exhibited by the top GWAS discovery meta-analysis signals. Replication genotyping was performed at the Wellcome Trust Sanger Institute using Sequenom iPLEX assays, and at the British Heart Foundation Glasgow Cardiovascular Research Centre using TaqMan Open Array genotyping. Variants were excluded from analysis if they had call rates < 95%; subjects with call rates < 80%, and families in the MoBa cohort that exhibited more than 1 Mendelian error were also excluded. For four variants (rs7305125, rs149427560, rs12050029 and rs4769628) follow-up data for the MoBa samples was *in silico* data based on 1046 cases and 961 controls assayed on the Illumina HumanCoreExome-12 v1.1 chip and imputed based on the 1000 Genomes Phase 3 reference panel. Of those, 908 cases and 909 controls were also included in the directly genotyped MoBa replication set.

**Meta-analysis**

Prior to meta-analysis GWAS results were adjusted by a genomic control $\lambda_{GC}$ factor where appropriate as described above for each GWAS cohort. Study level variants with a MAF<0.5% or an imputation quality score <0.6 were excluded from the analysis. This left 7,476,169 autosomal variants for analysis. The GWAS and the GWAS+Replication meta-analyses were conducted using the fixed effect inverse variance weighting method implemented in MetaSoft[27]. No genomic control adjustment was applied to the GWAS meta-analysis results since the inflation factor was negligible ($\lambda_{GC}$ =1.0075).

**Conditional and *FLT1* region-wide analyses**

The association between disease status on a variant conditional on rs4769613 was assessed by inverse variance weighted meta-analysis of the per cohort conditional analyses. Individual cohorts were analysed by logistic regression of the disease status against expected genotype dose with the expected doses of conditioning variants included as covariates. This approach was implemented for each cohort as follows: MoBa and FINNPEC replication cohorts were analysed using "plink --condition"; GOPEC and MoBa GWAS were analysed using "snptest -condition_on"; the deCODE association analysis is described above; ALSPAC conditional associations were inferred from the summary association statistics with the use of the 1000 Genomes Phase 3 EUR samples to estimate the LD structure using the joint analysis method[28]. We assessed the region-wide effective number of tests using the method of Gao[29] on the imputed WTCCC2 UK control dataset for the 5405 common variants (MAF>5%) within 1Mb of rs4769613, yielding a total of 832 independent tests and hence a *FLT1* region-wide significance threshold of $0.05/832=6.01\times10^{-5}$.

**Maternal, fetal and parent-of-origin effect analysis**

The family genotype data was jointly analysed using the EMIM method[6].The subjects were first partitioned into maximal family groups within each cohort (Supplementary Table 9). We then fitted EMIM models assuming Hardy-Weinberg equilibrium and Exchangeable Parental Genotypes. We considered two parameter sets: maternal and fetal effects ($R_1$, $R_2$, $S_1$ and $S_2$) and maternal, fetal and parent of origin effect ($R_1$, $R_2$, $S_1$, $S_2$ and $I_m$) where $I_m$ is the odds-ratio associated with the maternal transmission of the risk allele. The per cohort results were combined using inverse variance weighted meta-analysis.

**Preeclampsia subtype analysis**

To analyse the relation between rs4769613 and preeclampsia subtypes we pooled genotype and clinical data of the GOPEC, FINNPEC and MoBa cohorts (Supplementary Table 8). The phenotype associations were calculated using logistic regression with cohort and FINNPEC recruitment region included as indicator variables.

**deCODE phenotype database**

The deCODE Genetics phenotype database contains medical information on diseases and traits obtained through collaboration with specialists in each field. This includes information on cardiovascular diseases (myocardial infarction, coronary arterial disease, peripheral arterial disease, atrial fibrillation, sick sinus syndrome and stroke), metabolic disorders (obesity, diabetes, and metabolic syndrome), psychiatric disorders (schizophrenia, bipolar disorder, anxiety and depression), addictions (nicotine, alcohol), inflammatory diseases (rheumatoid arthritis, lupus, and asthma), musculoskeletal disorders (osteoarthritis, osteoporosis), eye diseases (glaucoma), kidney diseases (kidney stones, kidney failure) and 29 types of cancer. Anthropometric measures have also been

collected through several of these projects. Routinely measured traits from patient workups (sodium, potassium, bicarbonate, calcium, phosphate, creatinine, blood cell counts, haemoglobin, haematocrit, 15 immunoglobulins, iron, vitamins, lipids, liver function tests and more) were obtained from the Landspitali University Hospital, Reykjavík, and the Icelandic Medical Center Laboratory in Mjodd (Laeknasetrid), Reykjavik. The number of independent and uncorrelated secondary traits tested for association amounts to 400.

**Placental expression of Flt1 and sFlt1**

Women with singleton pregnancies delivering by caesarean section were recruited to the Pre-eclampsia Study between 2002 and 2012 at St. Olavs Hospital, Trondheim University Hospital and Haukeland University Hospital, Bergen. Healthy and pre-eclamptic pregnancies were included as described previously[30]. A tangential section (100 mg) from the maternal central side of the placenta was collected directly after delivery, fixed in 10% neutral-buffered formalin and paraffin embedded. Tissue sections of 3 µm were pre-treated in Target Retrieval Solution (#K8004, Dako) and stained by Flt-1 antibody (1:175, # ab32152, Abcam) using EnVision (#K4011, Dako) according to the manufacturer's protocol. This Flt-1 antibody recognises membrane-bound Flt-1 and splice isoforms sFlt-1, sFlt1-14, and isoform 4 (61 kDa). Staining was performed using Autostainer Plus (#S3800, Dako) and images taken at two sites per placenta with an Eclipse E400 microscope and DS-Fi1 camera. Staining intensity in syncytiotrophoblast was analysed by NIS-Elements BR 4.0 software (Nikon), excluding immature villi, and blinded for pregnancy outcomes. Staining intensity data were analysed separately in a general linear model incorporating SNP genotype, gestational age and hospital of origin as cofactors.

**Maternal serum sFlt1 and fetal genotype**

We identified mother-baby pairs from the FINNPEC collection for whom offspring DNA and maternal serum samples from the first and/or third trimester of pregnancy were available for analysis. Maternal serum sFlt-1 concentration was measured using electrochemiluminescence immunoassays (ECLIA; Roche Diagnostics GmbH, Mannheim, Germany) on a Cobas e 601 analyzer (Hitachi High Technology Co, Tokyo, Japan). Offspring genotype at rs4769613 was determined by Sequenom MassArray iPLEX genotyping in the FiMM Technology Centre (University of Helsinki, Finland). Investigators were blinded for pregnancy outcome during sample analysis. Serum sFlt-1 data were normalized by logarithmic transformation. Case and control data were compared by unpaired t-testing; genotypic associations with serum sFlt-1 were examined separately in cases and controls in a linear model, with SNP genotype and gestational age as covariates.

**Data Availability**

Meta-analysed GWAS data used in this study, and individual-level GWAS data from the GOPEC cohort, are deposited in the European Genome-phenome Archive (www.ebi.ac.uk/ega) with accession numbers EGAD00010001211 and EGAD00010001212.

**Methods references**

14. Brown, M. A., Lindheimer, M. D., de Swiet, M., Van Assche, A. & Moutquin, J. M. The classification and diagnosis of the hypertensive disorders of pregnancy: statement from the

International Society for the Study of Hypertension in Pregnancy (ISSHP). *Hypertens. Pregnancy* **20**, IX–XIV (2001).
15. The GOPEC Consortium. Disentangling fetal and maternal susceptibility for pre-eclampsia: a British multicenter candidate-gene study. *Am. J. Hum. Genet.* **77**, 127-131 (2005).
16. Evans, D. M. *et al.* Interaction between *ERAP1* and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat. Genet.* **43,** 761–767 (2011).
17. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'; the index offspring of The Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42,** 111-127 (2013).
18. Hjartardottir, S., Leifsson, B. G., Geirsson, R. T. & Steinthorsdottir, V. Paternity change and the recurrence risk in familial hypertensive disorder in pregnancy. *Hypertens. Pregnancy*. **23**, 219-225 (2004).
19. Magnus, P. *et al.* Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *Int. J. Epidemiol.* **45**, 382-388 (2016).
20. Jääskeläinen, T. *et al.* Cohort profile: the Finnish Genetics of Pre-eclampsia Consortium (FINNPEC). *BMJ Open.* Nov 10;6(11):e013148. doi:10.1136/bmjopen-2016-013148 (2016).
21. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* Dec;2(12):e190. doi:10.1371/journal.pgen.0020190 (2006).
22. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499-511 (2010).
23. O'Connell, J. *et al.* (2014) A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet*. Apr 17:10(4):e1004234. doi: 10.1371/journal.pgen.1004234 (2014).
24. Gudbjartsson, D. F. *et al*. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435-444 (2015).
25. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997-1004 (1999).
26. Hellmich, C. *et al*. Genetics, sleep and memory: a recall-by-genotype study of ZNF804A variants and sleep neurophysiology. *BMC Med. Genet.* Oct 24; 16:96. doi: 10.1186/s12881-015-0244-4 (2015).
27. Han, B. & Eskin, E. Genetics, sleep and memory: a recall-by-genotype study of ZNF804A variants. *Am. J. Hum. Genet.* **88,** 586-598 (2011).
28. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44,** 369-375 (2012).
29. Hendricks, A. E. *et al.* Correction for multiple testing in a gene region. *Eur. J Hum. Genet.* **22,** 414-418 (2014).
30. Austdal, M. *et al*. Metabolic profiles of placenta in preeclampsia using HR-MAS MRS metabolomics. *Placenta* **36**, 1455-1462 (2015).