



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

The genetic component of preeclampsia

A whole-exome sequencing study

Hansen, Anette Tarp; Bernth Jensen, Jens Magnus; Hvas, Anne-Mette; Christiansen, Mette

Published in:
PLOS ONE

DOI (link to publication from Publisher):
[10.1371/journal.pone.0197217](https://doi.org/10.1371/journal.pone.0197217)

Creative Commons License
CC BY 4.0

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Hansen, A. T., Bernth Jensen, J. M., Hvas, A-M., & Christiansen, M. (2018). The genetic component of preeclampsia: A whole-exome sequencing study. *PLOS ONE*, 13(5), 1-16. [e0197217].
<https://doi.org/10.1371/journal.pone.0197217>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- ? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- ? You may not further distribute the material or use it for any profit-making activity or commercial gain
- ? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

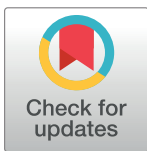
RESEARCH ARTICLE

The genetic component of preeclampsia: A whole-exome sequencing study

Anette Tarp Hansen^{1,2,3*}, Jens Magnus Bernth Jensen^{3,4}, Anne-Mette Hvas^{1,3}, Mette Christiansen⁴

1 Department of Clinical Biochemistry, Aarhus University Hospital, Aarhus, Denmark, **2** Department of Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark, **3** Department of Clinical Medicine, Health, Aarhus University, Aarhus, Denmark, **4** Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark

* anette.tarp.hansen@dadlnet.dk



Abstract

Preeclampsia is a major cause of maternal and perinatal deaths. The aetiology of preeclampsia is largely unknown but a polygenetic component is assumed. To explore this hypothesis, we performed an in-depth whole-exome sequencing study in women with (cases, N = 50) and without (controls, N = 50) preeclampsia. The women were identified in an unselected cohort of 2,545 pregnant women based on data from the Danish National Patient Registry and the Medical Birth Registry. Matching DNA was obtained from a biobank containing excess blood from routine antenatal care visits. Novogene performed the whole-exome sequencing blinded to preeclampsia status. Variants for comparison between cases and controls were filtered in the Ingenuity Variant Analysis software. We applied two different strategies; a *disease association panel* approach, which included variants in single genes associated with established clinical risk factors for preeclampsia, and a *gene panel* approach, which included biological pathways harbouring genes previously reported to be associated with preeclampsia. Variant variability was compared in cases and controls at the level of biological processes, signalling pathways, and in single genes. Regardless of the applied strategy and the level of variability examined, we consistently found positive correlations between variant numbers in cases and controls (all $R^2 > 0.88$). Contrary to what was expected, cases carried fewer variants in biological processes and signalling pathways than controls (all p -values ≤ 0.02). In conclusion, our findings challenge the hypothesis of a polygenetic aetiology for preeclampsia with a common network of susceptibility genes. The greater genetic diversity among controls may suggest a protective role of genetic diversity against the development of preeclampsia.

OPEN ACCESS

Citation: Hansen AT, Bernth Jensen JM, Hvas A-M, Christiansen M (2018) The genetic component of preeclampsia: A whole-exome sequencing study. PLoS ONE 13(5): e0197217. <https://doi.org/10.1371/journal.pone.0197217>

Editor: Obul Reddy Bandapalli, German Cancer Research Center (DKFZ), GERMANY

Received: January 19, 2018

Accepted: April 27, 2018

Published: May 14, 2018

Copyright: © 2018 Hansen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: According to the Danish Act on Processing of Personal Data (§10), we are not allowed to pass on our exome data to any public repository. Qualified researchers will be able to access our exome data, but this will require approval by the Danish Data Protection Agency for the specified project. We refer to their main website: <https://www.datatilsynet.dk/forside/>, on which all contact details and instructions for approval appear. The authors will assist with this process if necessary.

Introduction

Preeclampsia is a leading cause of maternal and perinatal deaths with estimated 343,000 women worldwide dying from preeclampsia in the last decennium [1]. In survivors, preeclampsia is associated with an increased risk of premature death from any cause, cardiovascular disease, and with adverse pregnancy outcomes in future pregnancies, including preeclampsia and impaired foetal growth requiring premature induction of delivery [2–6]. Preeclampsia

Funding: The trial was funded by The Danish Council for Independent Research (Grant no: 0602-02173B FSS). The funder had no role in the design or conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, critical review, and final approval of the paper.

Competing interests: The authors have declared that no competing interests exist.

affects 3–8% of pregnancies with an increasing incidence, probably due to an increased burden of maternal obesity and diabetes, and the trend of postponing pregnancy to higher maternal ages [7–10]. Identifying women at risk of developing preeclampsia enables early preventive treatment with low-dose aspirin [11].

The aetiology of preeclampsia is poorly understood, however, defect placentation with impaired utero-placental flow plays an essential role [9,12]. Disturbances in placental growth factors and regulators of angiogenesis, and reduced immune tolerance to “non-self” tissue in the placenta and the foetus are additional suggested mechanisms for preeclampsia [7,13]. Clustering of preeclampsia cases within families suggests a genetic etiological component from maternal, foetal, and/or paternal genes [9,14–16]. Women with an affected first relative are at three to five-fold increased risk of developing preeclampsia themselves [9,16,17] and within some families, preeclampsia seems to follow Mendelian patterns for disease inheritance of rare deleterious genetic variants [15,18]. Family studies based on large cohorts of affected women and relatives have suggested that variation in activin A receptor type 2A (*ACVR2*), rho associated coiled-coil containing protein kinase 2 (*ROCK2*), endoplasmic reticulum aminopeptidase 1 (*ERAP1*), and endoplasmic reticulum aminopeptidase 2 (*ERAP2*) genes are associated with preeclampsia [19–21], reviewed in [14]. However, for the majority of preeclampsia cases, the genetic contribution seems more complex and likely polygenetic [14,17]. This was supported by twin studies showing discordance in preeclampsia phenotype in monozygotic twin pairs, suggesting only minor genetic contribution [22,23]. Therefore, the findings from family studies may not be generalizable to the overall preeclampsia population [14,22,23].

Previous candidate gene studies and genome-wide association studies have chosen candidate genes based on the existent knowledge. More than 50 candidate genes for preeclampsia within various pathophysiological paths have been suggested, but no universally accepted susceptibility genes for preeclampsia have yet been identified [14]. Since preeclampsia probably has a polygenetic aetiology of rare genetic variants, a high-resolution systematic investigation of the whole exome is needed [24,25]. Kaartokallio et al. used pooled blood samples for an exome sequencing study, thus comparing the pooled frequency of gene variants to reference data. The authors concluded that no genetic variants reached statistical significance for preeclampsia [24]. However, this design rendered the origin of genetic variants blurred, *i.e.*, the reported variants could be clustered within few individuals or the other extreme be spread across individuals. Thus, transparent whole-exome investigations on single women are warranted to further unravel the genetic contribution in preeclampsia.

Yet, there is no published whole exome sequencing study for preeclampsia in the literature. Here, we report the first whole-exome sequencing study on blood samples from preeclampsia cases and controls, allowing direct comparison of the genetic variability in cases and controls.

Materials and methods

Setting

In Denmark, nearly all pregnant women attend a routine antenatal care visit at their general practitioner during early first trimester. Upon this first trimester visit, the general practitioner obtains a blood sample for maternal blood typing and screening for human immunodeficiency virus, hepatitis B virus, and syphilis.

In the period from May 2014 to June 2015, we collected all first trimester blood samples received for analysis every second day at the Department of Clinical Immunology, Aarhus University Hospital, Denmark (N = 2,545 women). The present study was based on these samples. EDTA stabilized whole blood was centrifuged, aliquoted in plasma and cell pellet, and stored at -80°C until genomic analysis.

We obtained data on the course of the 2,545 pregnancies from the Danish National Patient Registry [26] and the Medical Birth Registry [27]. Those born in or immigrating to Denmark receives a unique Civil Personal Registration number, enabling accurate linkage among Danish registries at the individual level [28]. The Danish National Patient Registry has recorded data on all admissions and discharges from Danish non-psychiatric hospitals according to the *International Classification of Diseases, Eighth Revision (ICD-8)* from 1977 until the end of 1993 and *Tenth Revision (ICD-10)* thereafter [26]. Each hospital discharge or outpatient visit is coded in the Danish National Patient Registry. The Medical Birth Registry contains prospectively collected data on all deliveries in Denmark since 1 January 1973 [27].

Study population

From the Danish National Patient Registry and the Medical Birth Registry, we identified women registered with a preeclampsia diagnosis among the 2,545 pregnancies. Danish preeclampsia patients are generally diagnosed according to the criteria for preeclampsia stated in the American College of Obstetricians and Gynaecologists' task force report on hypertension during pregnancy: hypertension ($\geq 140/90$ mm Hg) debuting from gestational weeks 20 and proteinuria (www.acog.org). In total, we identified 58 women with a preeclampsia diagnosis during index pregnancy. We included the 50 women developing preeclampsia (cases) at the earliest gestational ages. Skjaerven and co-workers previously reported that especially severe cases of preeclampsia seemed to have a genetic component [29]. Early onset of preeclampsia is in clinical practice considered severe preeclampsia. For that reason, we selected the 50 women with the earliest onset of preeclampsia as cases in the present study. These women accounted for 86% of the total number of preeclampsia cases in the entire cohort. The distribution of non-genetic known risk factors for preeclampsia (maternal age, parity, and body mass index; data not shown) did not differ between the subgroup of women included in the present study and the total group of preeclampsia cases. Additional non-genetic risk factors for preeclampsia were un-likely to have played a role for our case-selection and overall conclusions in the study.

Fifty women were randomly included as controls, if they had no history of diabetes, arterial cardiovascular disease, venous thromboembolism, transient ischemic attack, cerebral ischemic stroke, hypertension, acute or chronic renal disease or gestational diabetes diagnosed within the index pregnancy.

Whole-exome sequencing

Genomic DNA was purified from the cell pellets using the Qiasymphony DNA Midi kit (Qiagen, The Netherlands). The DNA concentration was determined using Qubit Broad Range DNA kit (ThermoFisher). Mean concentration was 174 ng/ μ L (range 27–394 ng/ μ L). Novogene Bioinformatics Technology Co., Hong Kong, performed the whole-exome sequencing blinded to preeclampsia disease status.

Post-sequencing bioinformatics

The variant call files containing variant info were uploaded to Ingenuity Variant Analysis Software (Qiagen), hereafter denominated IVA Software, and filtered as illustrated in Fig 1.

Initial filtering

First, we filtered for confidence by excluding variants of call quality below 20 (equivalent to base call accuracy of 99%), read depth below 10, allele fraction below 25%, and genotype quality below 30 (equivalent to genotype call accuracy of 99.9%). Subsequently, we excluded

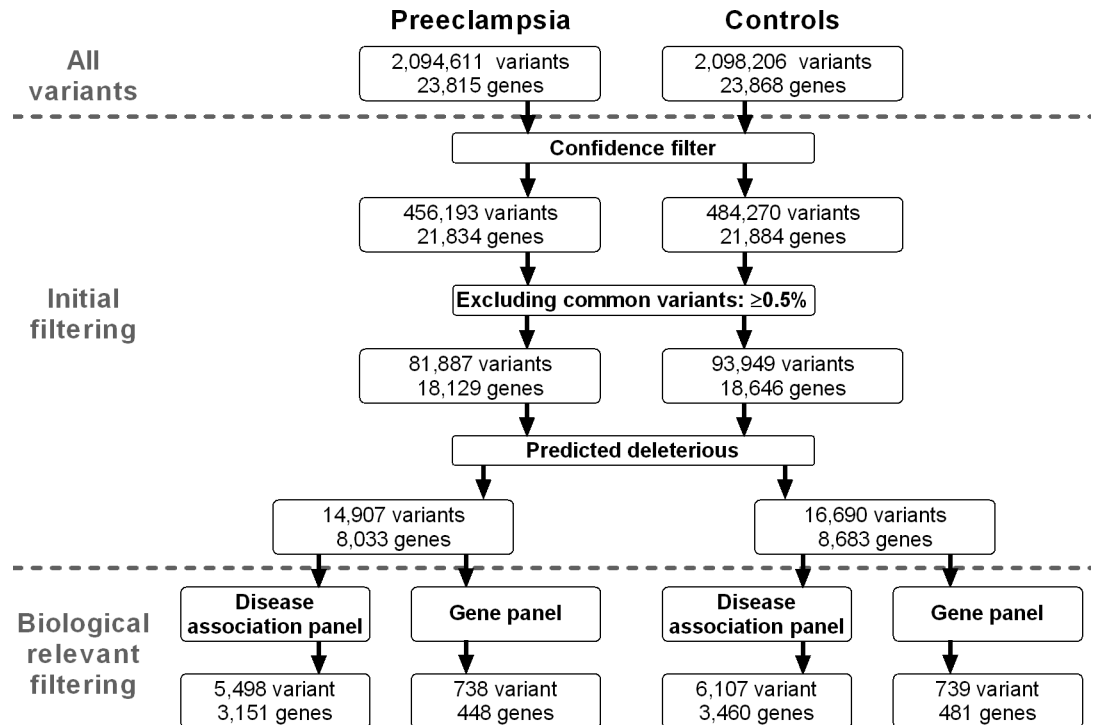


Fig 1. Filtering cascade. Number of variants and genes present after initial filtering and biological filtering.

<https://doi.org/10.1371/journal.pone.0197217.g001>

common variants. These were defined as variants with reported allele frequencies of 0.5% or greater in one of the following databases: 1000 genomes project [30], NHLBI ESP exomes [31], Allele frequency Community [32], Exome Aggregation Consortium [33] and gnomAD [34]. Variants predicted to be disease associated were included if the following criteria were met: variants classified as pathogenic or likely pathogenic according to computed American College of Medical Genetics and Genomics guidelines [35], disease-associated variants according to the Human Gene Mutation Database [36] or disease-associated variants according to the CLINVAR database [37].

Lastly, variants predicted to be deleterious were included according to the following criteria: frameshift, in-frame insertion and deletions, stop codon changes, missense unless predicted innocuous by SIFT [38] or polyphen-2 [39], CADD score > 15.0 [40] disrupt splice site up to 2 bases into intron, or predicted to disrupt splicing by MaxEntScan [41]. The initial filtering is further specified in [S1 Appendix](#).

Biological filtering

We employed two different approaches; a *disease association panel* approach and a *gene panel* approach for the filtering according to biological mechanisms (specified in [S1 Appendix](#)). In the *disease association panel* approach, we filtered for variants in genes associated with clinical risk factors for preeclampsia [7]. These included genes associated with abnormal immune tolerance, chronic kidney disease, coagulopathy, diabetes, hypertension, preeclampsia, systemic lupus erythematosus, and thrombophilia. In the *gene panel* approach, we searched for relevant biological pathways harbouring genes previously reported to be protective or disease causing for preeclampsia. These included the renin-angiotensin system pathway, antigen presentation folding, peptide loading of class I major histocompatibility complex, cell adhesion endothelial

cell contacts by non-junctional mechanisms, DNA double-strand break repair, complement and coagulation cascades, epoxide hydrolase pathway, myometrial relaxation and contraction pathways, transforming growth factor-beta signalling pathway. In the *gene panel*, we additionally included genes previously reported to be associated with preeclampsia in a large genome wide association study [21]. We exported genes present in these pathways from PathCards [42].

Biological processes and pathways

We furthermore investigated the genetic variability by comparing the variant frequency for cases and controls separately of the 100 most significant biological processes and pathways, as defined by the IVA software. The IVA software defines biological processes as the biological properties of particular molecules, or the effects that a given molecule has on a disease or function [43]. Based on classical models of signal transduction and information in the Qiagen Knowledge Base [43], the Qiagen's Content Curation team has defined pathways in the IVA software. The IVA software determined the significance of each particular biological process or pathway by the use of a right-tailed Fisher's exact test, testing if the frequency of genetic variants within these processes and pathways was higher than expected by random chance.

Statistics

The correlations of variants in the two groups were examined by linear regression, based on least squares residuals. Difference in variant prevalence was assessed using binomial distribution, as the probability of the observed prevalence given prevalence was equal in the groups. We defined the level of significance as 0.05. Raw *p*-values were corrected for multiple comparisons by the Bonferroni approach, unless otherwise stated. Ninety-five percent confidence intervals are presented in brackets []. Data analysis was made in Stata 11.0, StataCorp, USA.

Ethics

The Ethics Committee of Central Regional Denmark and the Danish Data Protection Agency approved the study (record number/date 1-16-02-294-13/ 20.06.2013 and record number 1-10-72-46-16).

Results

We collected blood samples from 2,545 unique women attending their first antenatal visit at their general practitioner. Of those, 58 (2.2%) subsequently had a preeclampsia diagnosis registered in the Danish Medical Birth Registry. We included the 50 women who developed preeclampsia at the earliest gestational ages. Table 1 shows demographic characteristics of women developing and not developing preeclampsia.

Women developing preeclampsia were younger, delivered infants with lower birth weights, and had a lower gestational age in comparison to controls. The raw sequencing data obtained in the case and control groups were of similar coverage and quality (Table 2). The *p*-values for all quality parameters were all > 0.1.

Comparison of the 100 most variable biological processes in cases and controls

Using the *disease association panel*, we identified the 100 most variable biological processes in each of the two groups. The two groups shared 94 of these processes (Fig 2A). All women had variants in each of these biological processes. For the shared variant processes, we found a

Table 1. Characteristics of women with (cases) and without (controls) preeclampsia.

	Cases (N = 50)		Controls (N = 50)	
	Median	10–90 percentile	Median	10–90 percentile
Gestational age at blood sampling, days	76	69–87	80	66–91
Gestational age at blood sampling, weeks ^{+days}	10 ⁺⁶	–	11 ⁺⁶	–
Maternal age at blood sampling, years	25	20–35	30	24–37
Maternal body mass index, kg/m ²	24	20–33	23	19–29
Gestational age at delivery, days	267	237–282	281	266–291
Birth weight, grams	3,088	1,605–3,770	3,505	2,804–4,140

<https://doi.org/10.1371/journal.pone.0197217.t001>

strong and highly significant correlation of total variant number (Fig 2B) and number of genes carrying variants (Fig 2C) across the case and control groups. Cases had 10% [9.4%–11%] fewer variants per biological process and 7.6% [6.8%–8.3%] fewer genes with variants per biological process compared to the control group ($p < 0.0001$). To explore this difference further, we identified the biological processes in which variant frequency differed between the two groups. We found 40 different biological processes with a lower number of variants in cases than in controls (Fig 2D).

We repeated these analyses using the *gene panel*. Even more biological processes, 97 of the 100, were shared for the case and control group (Fig 2E). Again, the two groups correlated strongly in number of variants in single processes (Fig 2F) and the number of genes harbouring variants (Fig 2G). In addition, cases had fewer variants per process (3.9% [1.2%–6.6%]) and fewer genes were affected per process (10% [8.5%–12%]) compared to controls ($p \leq 0.005$). However, the variant frequency did not differ for any single biological process.

Comparison of the 100 most variable pathways in cases and controls

We then addressed variability in pathways, as defined in the IVA Software. When using the *disease association panel*, we found that 82 pathways were shared for cases and controls (Fig 3A). Both the number of variants per pathway and the number of genes harbouring variants

Table 2. Whole-exome sequencing quality data.

	Cases Mean (SD)	Controls Mean (SD)
Raw data (Gb)	7.7 (0.6)	7.6 (0.7)
Q30 (%)	91.6 (1.1)	91.4 (1.0)
Effective sequences on or near target (Mb)	5,558 (415)	5,461 (443)
Mean sequencing depth on target	70 (5)	68 (6)
Fraction of target covered with at least 20x	0.95 (0.01)	0.95 (0.01)
Total number of SNP	218,254 (14,170)	220,074 (13,638)
Number of transitions	151,996 (9,940)	153,109 (9,500)
Number of transversions	66,257 (4,245)	66,965 (4,150)
Number of novel SNP	17,600 (4,735)	16,497 (3,613)
Number of SNP in CDS	22,883 (888)	23,001 (998)
Number of stop gains	86 (6)	88 (7)
Number of stop loss	12 (2)	12 (2)

SD: standard deviation; Gb: Giga-bases. Q30: the percent of bases with phred-scaled quality scores greater than 30.

Mb: Mega-bases

<https://doi.org/10.1371/journal.pone.0197217.t002>

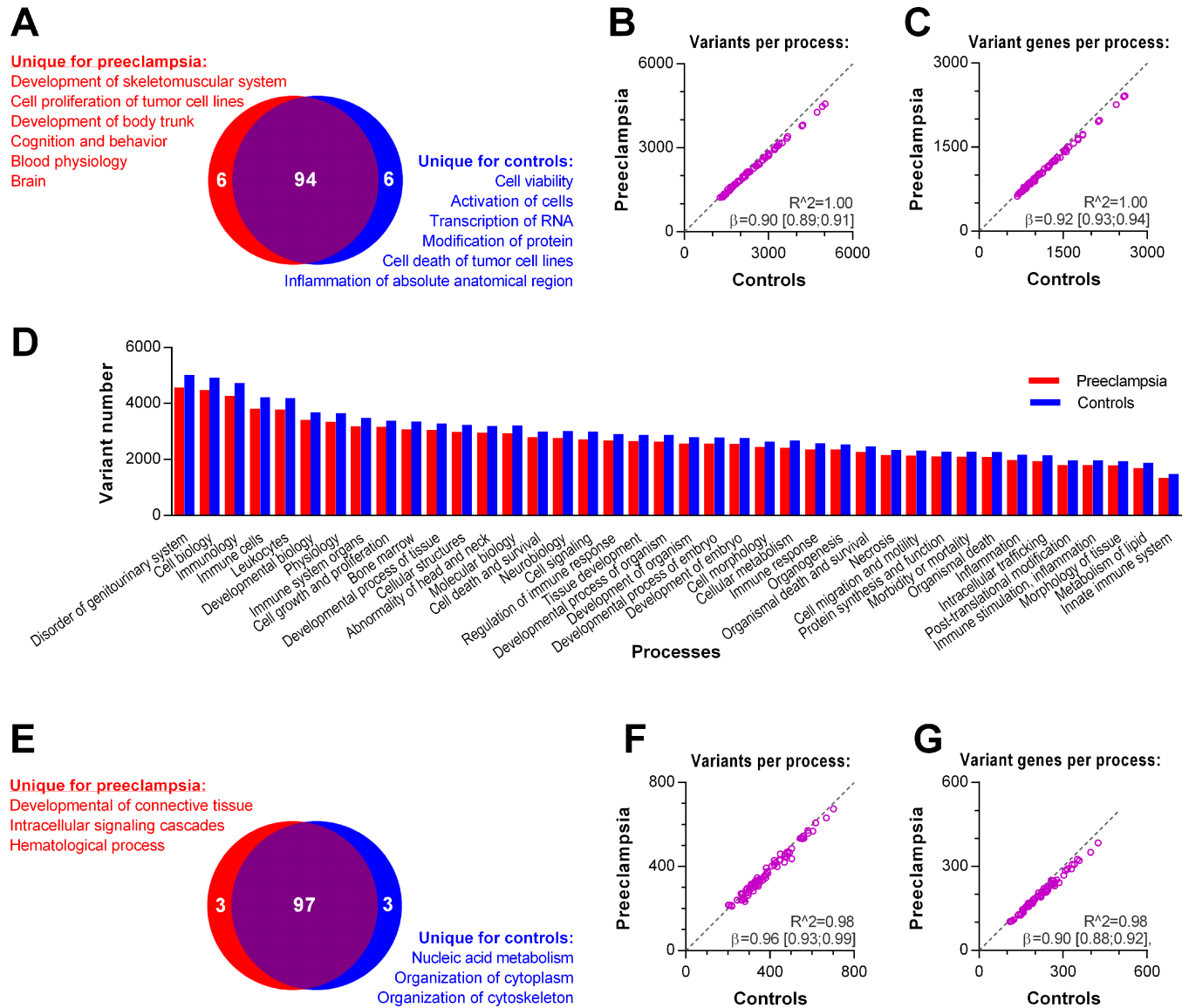


Fig 2. Comparison of the 100 most variable biological processes in cases and controls. Variant processes were identified with the *disease association panel* (A–D) or the *gene panel* (E–G). A: The number of unique and shared variant harbouring processes. For shared variant processes, correlation of total variant number (B) and number of genes carrying variants (C). Number of variants in 40 processes which less frequently carried variants in cases (D). E: Number of unique and shared variant harbouring processes. For shared variant harbouring processes, correlation of total variant number (F) and number of genes carrying variants (G).

<https://doi.org/10.1371/journal.pone.0197217.g002>

correlated strongly and significantly in the two groups (Fig 3B and 3C). On average, cases had 6.7% [1.1%–12%] fewer variants per pathway and 14% [7.8%–20%] fewer variant genes per pathway compared to controls ($p \leq 0.02$).

When applying the *gene panel* approach, we observed that 87 pathways were shared for cases and controls (Fig 3D). Both the number of variants per pathway and the number of genes harbouring variants correlated strongly in the two groups (Fig 3E and 3F). On average, cases had 18% [11%–26%] fewer variants per pathway and 21% [14%–27%] fewer variant genes per pathway than controls ($p < 0.0001$).

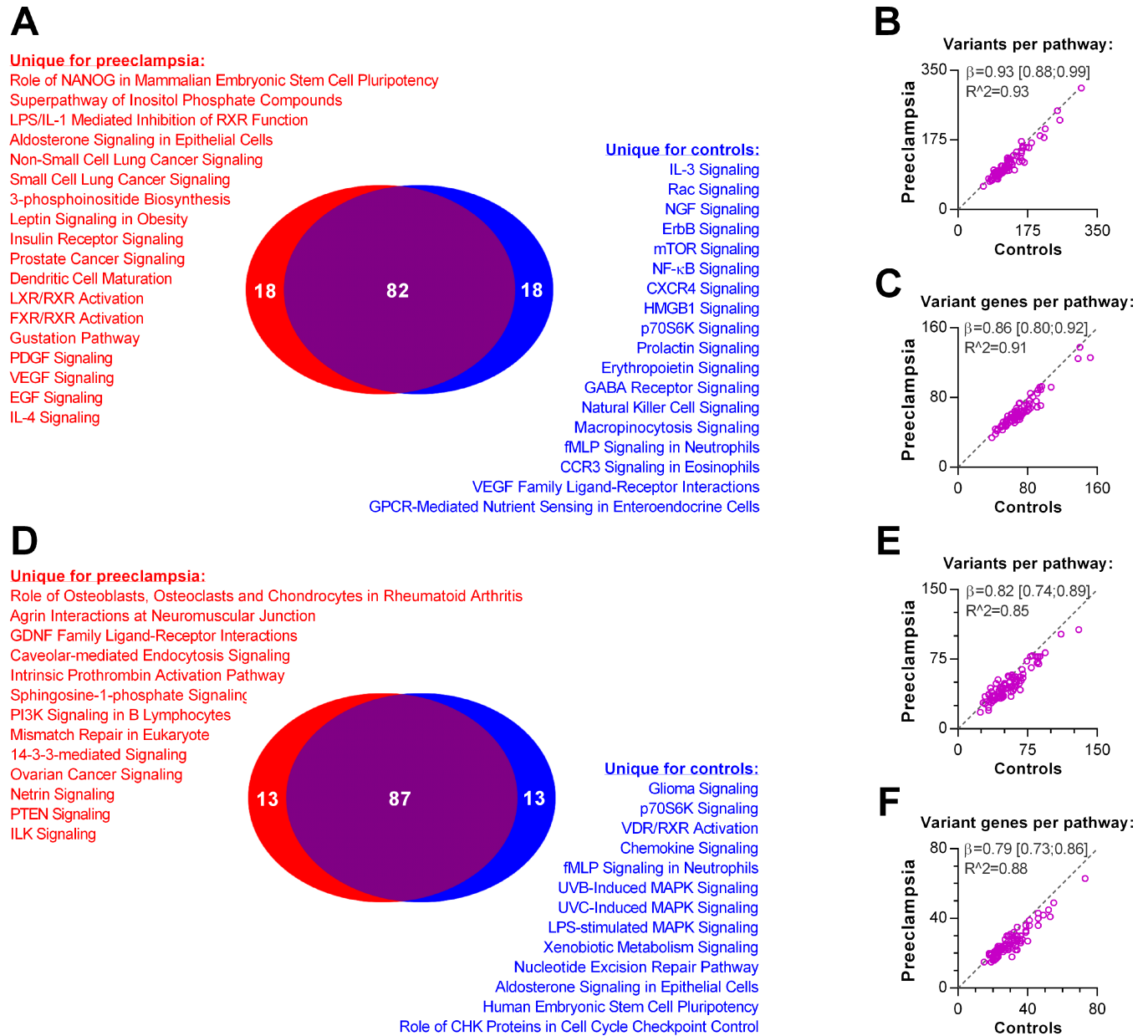


Fig 3. Comparison of the 100 most variable pathways in cases and controls. Variant pathways were identified with the *disease association panel* (A-C) or the *gene panel* (C-F). **A:** Number of unique and shared variant harbouring pathways. For shared variant pathways, correlations of total variant number and number of genes carrying variants are presented in **B** and **C**, respectively. **D:** Number of unique and shared variant harbouring pathways. For shared variant pathways, correlations of total variant number and number of genes carrying variants are presented in **E** and **F**.

<https://doi.org/10.1371/journal.pone.0197217.g003>

Comparison of variant genes in cases and controls

We then studied variants in single genes. Again, we applied the *disease association panel*. Cases had variants in 1,018 genes and controls had variants in 1,255 genes, whereas both groups had variants in 1,949 genes (Fig 4A). The total number of variants in the shared genes again showed a strong and highly significant correlation between the two groups (Fig 4B). Fig 4C depicts raw data on shared genes with the most skewed prevalence of variants (raw *p*-values below 0.05, but without significant difference after Bonferroni correction for multiple

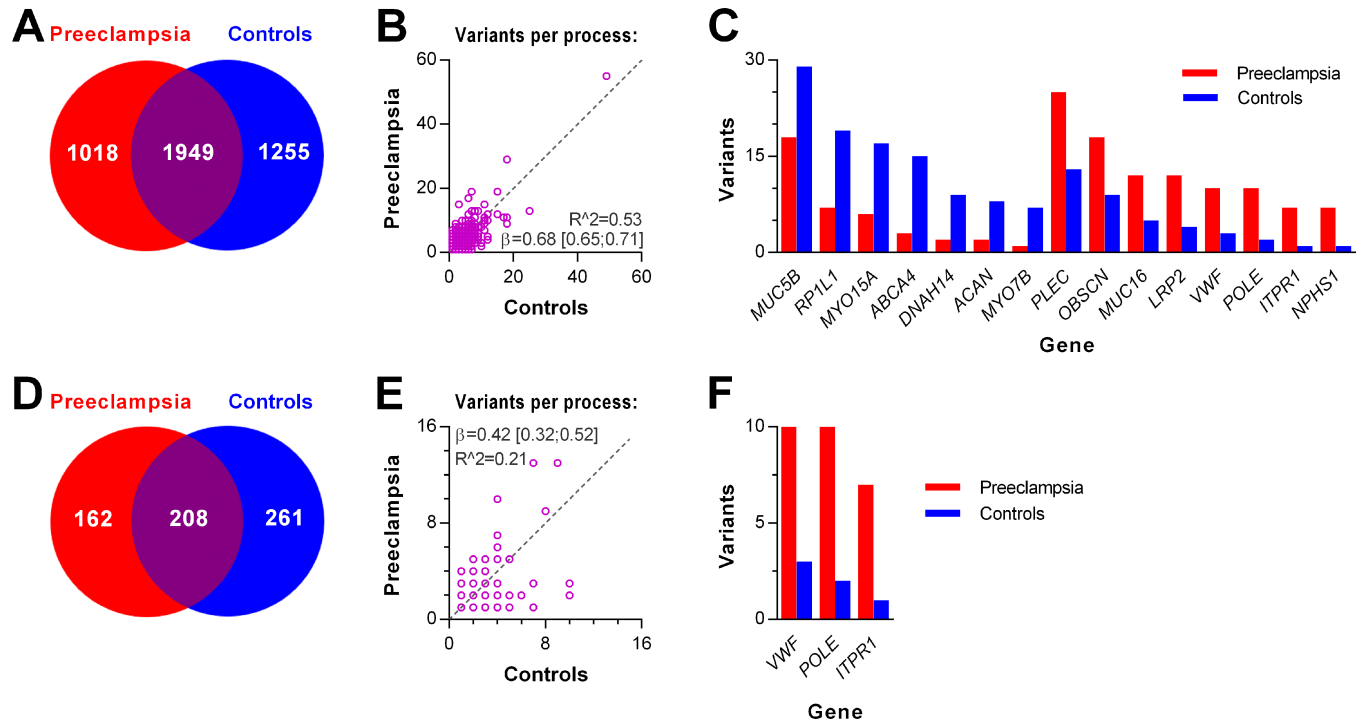


Fig 4. Comparison of variant genes in cases and controls. Genes containing variants were identified with the *disease association panel* (A-C) or the *gene panel* (C-F). A: Number of unique and shared variant harbouring genes. For the shared genes, correlation of total variant number is depicted in B. C: Genes with extreme distribution of variant numbers between groups (uncorrected *p*-values below 0.05). D: Number of unique and shared variant harbouring pathways. For the shared genes, correlation of total variant number is depicted in E. F: Genes with extreme distribution of variant numbers between groups (uncorrected *p*-values below 0.05).

<https://doi.org/10.1371/journal.pone.0197217.g004>

comparison). We then applied the *gene panel*. Cases had variants in 162 genes, controls had variants in 261 genes, whereas both groups shared 261 genes harbouring variants (Fig 4D). The total number of variants in the shared genes correlated strongly between the two groups (Fig 4E) and no difference was observed in variant numbers for any of these genes. Fig 4F depicts raw data on shared genes with the most skewed prevalence of variants (raw *p*-values below 0.05, but without significant difference after Bonferroni correction for multiple comparison).

Rare deleterious variants more frequent in cases

We performed a sensitivity analysis for uncovering the maximum genetic variability between cases and controls. We did so by comparing the frequency of rare deleterious variants for cases and controls directly. We selected for genes harbouring deleterious variants more frequent in preeclampsia cases than in controls (variants in single genes in $\geq 10\%$ of cases and $\leq 2\%$ of controls). Five genes were identified; methylenetetrahydrofolate reductase (*MTHFR*), inositol 1,4,5-trisphosphate receptor type 1 (*ITPR1*), discs large MAGUK scaffold protein 2 (*DLG2*), sucrose isomaltase (*SI*), and ataxin 1 (*ATXN1*) (Table 3).

Presence of previously reported genes associated with preeclampsia

Finally, we investigated the presence of variants in four selected genes previously reported in relation to preeclampsia; the *ROCK2*, *ACVR2A*, *ERAP1*, and *ERAP2* genes [17,19–21]. We chose to include rather frequent variants ($< 10\%$) to be able to assess potential accumulation of

Table 3. Rare predicted deleterious variants found in $\geq 10\%$ of women with preeclampsia (cases) and 2% of controls.

Gene Symbol Transcript ID	Transcript Variant	Protein Variant	gnomAD§	CADD Score*	Clinvar/ HGMD (Variant Class) α	Variants present	
						Cases	Controls
<i>MTHFR</i> NM_005957.4	c.*3046G>A		0.372	<10	RCV000344422.1 (VUS)	1	0
	c.*2594C>T		0.206	<10	RCV000293884.1 (VUS)	1	0
	c.*1294G>A	p.A1430A	0.054	11.4	RCV000370067.1 (VUS)	1	0
	c.1970G>C	p.*657S	0.001	<10	CM035841 (Low activity)	1	0
	c.1409A>T	p.E470V	0.168	23.3	RCV000319501.1 (VUS)	1	1
	c.870C>T	p.N290N	0.020	10.1	RCV000310090.1 (VUS)	1	0
<i>ITPR1</i> NM_002222.5	c.-2A>G		0.041	15.2	RCV000399171.1 (VUS)	1	0
	c.195C>G	p.R65R	0.172	13.0	RCV000336374.1 (VUS)	1	0
	c.971C>A	p.A324D	0.003	23.4	-	1	0
	c.2985A>G	p.S995S	0.092	<10	-	1	0
	c.5076C>T	p.N1692N	0.188	15.7	RCV000278808.1 (VUS)	1	0
	c.6007C>T	p.H2003Y	-	29.6	RCV000177981.1 RCV000359508.1 (VUS)	1	0
	c.6321+1G>T		-	23.5	-	1	0
<i>DLG2</i> NM_001142699.1	c.*137C>T		-	22.3	-	2	0
	c.1799C>T	p.T600M	0.012	33.0	-	1	0
	c.837G>T	p.G279G	0.005	18.0	-	1	0
	c.357+188529C>T		0.375	15.6	-	1	0
	c.-16T>C		0.004	16.4	-	1	0
<i>SI</i> NM_001041.3	c.5279G>A	p.G1760D	0.017	22.7	-	1	0
	c.2923T>C	p.Y975H	0.409	26.6	RCV000366023.1 (VUS)	2	0
	c.1958G>T	p.G653V	-	33.0	-	1	0
	c.878G>A	p.G293D	0.000	31.0	-	1	0
<i>ATXN1</i> NM_000332.3	c.2114G>C	p.S705T	-	16.8	-	1	0
	c.772G>T	p.G258C	0.172	24.2	-	1	0
	c.642G>T	p.Q214H	0.368	11.6	-	1	0
	c.302C>T	p.T101M	0.019	24.5	-	1	0
	c.-748G>C		0.002	15.9	-	1	0

§) gnomAD frequency in %

*) CADD score above 20 belongs to the 1% most deleterious variants.

α) Clinical significance reported in ClinVar.—Not present in database. VUS: variant of unknown significance.

<https://doi.org/10.1371/journal.pone.0197217.t003>

more frequent variants among cases. For *ROCK2*, we identified one variant among cases and one among controls. For *ACVR2A*, we identified one variant among cases and none among controls. We identified four *ERAP1* variants among cases and none among controls. For *ERAP2*, we identified 30 variants among cases and 18 among controls. When we searched for more rare variants (<1%), no difference was detected (Table 4).

Discussion

We demonstrated a high degree of genetic concordance for cases and controls. Our findings therefore challenge the suggested hypothesis of a polygenetic aetiology for preeclampsia with a common network of susceptibility genes [13,20]. Our findings support previous twin studies that reported only minor genetic contribution for preeclampsia [22,23]. The occurrence of preeclampsia in our cohort was lower (2.2%) than reported by others [6–9], but similar to previous reports on preeclampsia registration in the Danish Medical Birth Registry [26]. This may

Table 4. Predicted deleterious variants in ROCK2, ACVR2A, ERAP1, and ERAP2 in women with (cases) and without (controls) preeclampsia.

Gene Symbol Transcript ID	Transcript Variant	Protein Variant	gnomAD§	CADD score*	Clinvar/ HGMD (Variant Class) ☞	Variants present	
						Cases	Controls
ROCK2 NM_004850.4	c.2833A>C	p.M945L	-	22.5	-	0	1
	c.407T>G	p.F136C	0.095	26.0	-	1	0
ACVR2A NM_001616.4	c.1077+36A>G	-	0.025	15.2	-	1	0
ERAP1 NM_016442.4	c.2101-89C>T	-	0.184	15.6	-	1	0
	c.1939G>A	p.V647I	0.297	27.8	CM0911242 (DM?)	1	0
	c.1398G>A	p.Q466Q	0.009	15.6	-	1	0
	c.663+145G>A	-	0.055	15.9	-	1	0
ERAP2 NM_022350.4	c.291C>T	p.I97I	7.912	18.3	-	12	8
	c.641C>T	p.P214L	1.824	32.0	-	2	0
	c.995T>C	p.V332A	0.010	<10	-	1	0
	c.1040C>T	p.T347M	2.149	26.8	-	5	2
	c.1232T>G	p.L411R	0.399	28.2	RCV000179970.1 (Benign)	0	1
	c.2006T>A	p.L669Q	3.937	26.9	CM0911418 (DP)	6	3
	c.2045A>T	p.D682V	0.020	24.0	-	0	1
	c.2726A>T	p.D909V	0.002	23.3	-	0	1
	c.2740-27G>A	-	0.375	16.4	-	4	1
	c.2855T>C	p.L952P	0.120	26.7	-	0	1

§) gnomAD frequency in %

*) CADD score above 20 belongs to the 1% most deleterious variants.

☞) Clinical Significance reported in ClinVar.—Not present in database. DM: disease-causing mutation, DP: disease-associated polymorphism.

<https://doi.org/10.1371/journal.pone.0197217.t004>

be explained by heterogeneity in the definition of preeclampsia used in different studies [13,16] and the fact that only preeclampsia requiring hospital admission are registered in the Danish Medical Birth Registry [30].

Cases did not appear more distinct from the reference genome than controls concerning variants in biological processes. Using the *disease association panel*, the controls carried a slightly higher frequency of variants in a higher number of genes compared to preeclampsia cases. This suggests that general variant accumulation, in the studied processes, does not contribute to the development of preeclampsia. Contrary, our finding that preeclampsia cases were less genetic diverse might be suggestive of a protective role of genetic diversity.

The genetic concordance was high for cases and controls for the signalling pathways. In the *gene panel* approach, cases carried slightly fewer variant genes per pathway than controls. The raw data from both the *disease association panel* and *gene panel* uncovered more variants in the von Willebrand factor gene, the polymerase DNA epsilon gene, and the *ITPR1* gene. Nevertheless, correction for multiple comparisons eliminated the statistical significant difference in variant numbers for any of the shared genes.

The overrepresentation of *ERAP2* variants in cases was explained by frequent variants. There was no difference when looking at the more rare variants (<1%). We found that deleterious variants in the *MTHFR*, *ITPR1*, *DLG2*, *SI*, and *ATXN1* genes were more frequent in cases compared to controls. The variants detected in *MTHFR* are reported in ClinVar as possibly associated with neural tube defects except p.*657S. p.*657S is reported in relation to *MTHFR* deficiency [44,45]. Variants in the *MTHFR* genes have, together with the most frequent inherited thrombophilia markers Factor V Leiden mutation and the prothrombin variant (G20210A), been reported associated with preeclampsia [17,46]. However, several reports have

also largely refuted this association [14,46]. The *IPTR1* gene has been reported to be involved in maintenance of normal blood pressure through IP3R1-mediated regulation of eNOS [47]. The *DLG2* gene is previously reported differentially expressed in transcripts of decidua basalis in preeclampsia [48]. Based on these findings, we cannot rule out a possible role of the *MTHFR*, *ITPR1*, and *DLG2* gene variants for risk of developing preeclampsia.

Variants in *SI* were reported in ClinVar as a variant of unknown significance. The *ATXN1* gene is a causative factor for spinocerebellar ataxia-1 [49] and has been suggested to participate in the highly conserved Notch signalling pathway with regulatory importance for embryonic development [50]. The variants in the *SI* and *ATXN1* genes demonstrated in the present study are of more dubious clinical relevance for preeclampsia.

Presence of previously reported genes associated with preeclampsia

The *AVCR2*, *ROCK2*, *ERAP1*, and *ERAP2* genes previously reported associated with preeclampsia [19–21], reviewed in [14] were not found to be more frequent in cases compared to controls in the present study. Thus, our findings do not support previous findings of their possible contribution to risk of developing preeclampsia.

Strengths and limitations

The major strength in the present study was the robust study design based on blood samples from an unselected cohort of pregnant women representative for the general Danish population of pregnant women. The Danish tax-paid health care system ensures free access for all inhabitants, including free antenatal care at midwives and general practitioners. Therefore, blood samples collected as a part of this antenatal care constituted a unique source for studying genetic variability according to risk of developing preeclampsia. The study was also strengthened by the fact that we based our case-identification on nationwide Danish registry data free from recall bias. The Danish Medical Birth Registry has complete coverage for all women giving birth in Denmark. The validity of the preeclampsia diagnosis is high in The Danish Medical Birth Registry, *i.e.*, a positive and negative predictive value of 88% and 97% [26,37]. To further validate our case-selection, we obtained maternal demographic data and data on birth outcomes for both cases and controls. These data showed lower gestational ages and birth weights in the case group, consistent with the known clinical course in preeclampsia. Finally, we performed an in-depth sequencing and bioinformatics analysis of the entire exome of both cases and controls. This enabled direct comparison of variant frequencies in the two groups with reference to the reference genome [43].

We did not have access to data on family history of preeclampsia. Therefore, we cannot entirely rule out a family history of preeclampsia in the controls. The sample size may have limited the power of the present study, potentially leading to false-negative results [51]. However, we believe, that if any clinically relevant differences in the frequencies of protective or disease causing variants were present, it would have shown in the present study.

Conclusion

In this explorative whole-exome sequencing study, we found no evidence of a common network of genetic variability predisposing to preeclampsia. Preeclampsia affects the reproductive success, wherefore it is biologically plausible that preeclampsia susceptibility genes are under negative evolutionary control, thereby keeping the population frequencies of susceptibility variant low [23]. The present study indicates that genetic markers carry a minor potential for predicting preeclampsia. Cases even had a reduced genetic variability compared to controls.

Future studies should focus on the clinical impacts of this reduced variability in women suffering from preeclampsia.

Supporting information

S1 Appendix.

(DOCX)

Acknowledgments

We sincerely thank professor and director Eric Moses, Centre for Genetic Origins of Health & Disease, University of Western Australia, for his academic contributions into the bioinformatics approaches in the present study.

Author Contributions

Conceptualization: Anette Tarp Hansen, Jens Magnus Bernth Jensen, Anne-Mette Hvas, Mette Christiansen.

Data curation: Anette Tarp Hansen, Jens Magnus Bernth Jensen.

Formal analysis: Jens Magnus Bernth Jensen, Mette Christiansen.

Funding acquisition: Anette Tarp Hansen, Anne-Mette Hvas.

Investigation: Anette Tarp Hansen, Jens Magnus Bernth Jensen, Anne-Mette Hvas, Mette Christiansen.

Methodology: Anette Tarp Hansen, Jens Magnus Bernth Jensen, Anne-Mette Hvas, Mette Christiansen.

Project administration: Anette Tarp Hansen, Anne-Mette Hvas.

Writing – original draft: Anette Tarp Hansen.

Writing – review & editing: Anette Tarp Hansen, Jens Magnus Bernth Jensen, Anne-Mette Hvas, Mette Christiansen.

References

1. Say L, Chou D, Gemmill A, Tunçalp O, Moller AB, Daniels J, et al. Global causes of maternal death: a WHO systematic analysis. *Lancet Glob Health* 2014 Jun; 2(6):e323–33. [https://doi.org/10.1016/S2214-109X\(14\)70227-X](https://doi.org/10.1016/S2214-109X(14)70227-X) PMID: 25103301
2. Tooher J, Thornton C, Makris A, Ogle R, Korda A, Hennessy A. All Hypertensive Disorders of Pregnancy Increase the Risk of Future Cardiovascular Disease. *Hypertension* 2017 Oct; 70(4):798–803. <https://doi.org/10.1161/HYPERTENSIONAHA.117.09246> PMID: 28893895
3. Bellamy L, Casas JP, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *BMJ* 2007 Nov 10; 335(7627):974. <https://doi.org/10.1136/bmj.39335.385301.BE> PMID: 17975258
4. Loset M, Johnson MP, Melton PE, Ang W, Huang RC, Mori TA, et al. Preeclampsia and cardiovascular disease share genetic risk factors on chromosome 2q22. *Pregnancy Hypertens* 2014 Apr; 4(2):178–185. <https://doi.org/10.1016/j.preghy.2014.03.005> PMID: 26104425
5. Melamed N, Hadar E, Peled Y, Hod M, Wiznitzer A, Yogev Y. Risk for recurrence of preeclampsia and outcome of subsequent pregnancy in women with preeclampsia in their first pregnancy. *J Matern Fetal Neonatal Med* 2012 Nov; 25(11):2248–2251. <https://doi.org/10.3109/14767058.2012.684174> PMID: 22524456
6. Behrens I, Basit S, Melbye M, Lykke JA, Wohlfahrt J, Bundgaard H, et al. Risk of post-pregnancy hypertension in women with a history of hypertensive disorders of pregnancy: nationwide cohort study. *BMJ* 2017 Jul 12; 358:j3078. <https://doi.org/10.1136/bmj.j3078> PMID: 28701333

7. Bartsch E, Medcalf KE, Park AL, Ray JG, High Risk of Pre-eclampsia Identification Group. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. *BMJ* 2016 Apr 19; 353:i1753. <https://doi.org/10.1136/bmj.i1753> PMID: 27094586
8. Wallis AB, Saftlas AF, Hsia J, Atrash HK. Secular trends in the rates of preeclampsia, eclampsia, and gestational hypertension, United States, 1987–2004. *Am J Hypertens* 2008 May; 21(5):521–526. <https://doi.org/10.1038/ajh.2008.20> PMID: 18437143
9. Valenzuela FJ, Perez-Sepulveda A, Torres MJ, Correa P, Repetto GM, Illanes SE. Pathogenesis of pre-eclampsia: the genetic component. *J Pregnancy* 2012; 2012:632732. <https://doi.org/10.1155/2012/632732> PMID: 22175024
10. MacDorman MF, Declercq E, Thoma ME. Trends in Maternal Mortality by Sociodemographic Characteristics and Cause of Death in 27 States and the District of Columbia. *Obstet Gynecol* 2017 May; 129(5):811–818. <https://doi.org/10.1097/AOG.0000000000001968> PMID: 28383383
11. Roberge S, Bujold E, Nicolaides KH. Aspirin for the prevention of preterm and term preeclampsia: systematic review and metaanalysis. *Am J Obstet Gynecol* 2017 Nov 11.
12. Ho L, van Dijk M, Chye STJ, Messerschmidt DM, Chng SC, Ong S, et al. ELABELA deficiency promotes preeclampsia and cardiovascular malformations in mice. *Science* 2017 Aug 18; 357(6352):707–713. <https://doi.org/10.1126/science.aam6607> PMID: 28663440
13. Triche EW, Harland KK, Field EH, Rubenstein LM, Saftlas AF. Maternal-fetal HLA sharing and pre-eclampsia: variation in effects by seminal fluid exposure in a case-control study of nulliparous women in Iowa. *J Reprod Immunol* 2014 Mar; 101–102:111–119. <https://doi.org/10.1016/j.jri.2013.06.004> PMID: 23998333
14. Williams PJ, Broughton Pipkin F. The genetics of pre-eclampsia and other hypertensive disorders of pregnancy. *Best Pract Res Clin Obstet Gynaecol* 2011 Aug; 25(4):405–417. <https://doi.org/10.1016/j.bpobgyn.2011.02.007> PMID: 21429808
15. Chesley LC, Cooper DW. Genetics of hypertension in pregnancy: possible single gene control of pre-eclampsia and eclampsia in the descendants of eclamptic women. *Br J Obstet Gynaecol* 1986 Sep; 93(9):898–908. PMID: 3768285
16. Skjaerven R, Vatten LJ, Wilcox AJ, Ronning T, Irgens LM, Lie RT. Recurrence of pre-eclampsia across generations: exploring fetal and maternal genetic components in a population based cohort. *BMJ* 2005 Oct 15; 331(7521):877. <https://doi.org/10.1136/bmj.38555.462685.8F> PMID: 16169871
17. Fong FM, Sahemey MK, Hamed G, Eytayo R, Yates D, Kuan V, et al. Maternal genotype and severe preeclampsia: a HuGE review. *Am J Epidemiol* 2014 Aug 15; 180(4):335–345. <https://doi.org/10.1093/aje/kwu151> PMID: 25028703
18. Salonen Ros H, Lichtenstein P, Lipworth L, Cnattingius S. Genetic effects on the liability of developing pre-eclampsia and gestational hypertension. *Am J Med Genet* 2000 Apr 10; 91(4):256–260. PMID: 10766979
19. Ark M, Yilmaz N, Yazici G, Kubat H, Aktas S. Rho-associated protein kinase II (rock II) expression in normal and preeclamptic human placentas. *Placenta* 2005 Jan; 26(1):81–84. <https://doi.org/10.1016/j.placenta.2004.03.012> PMID: 15664415
20. Johnson MP, Roten LT, Dyer TD, East CE, Forsmo S, Blangero J, et al. The ERAP2 gene is associated with preeclampsia in Australian and Norwegian populations. *Hum Genet* 2009 Nov; 126(5):655–666. <https://doi.org/10.1007/s00439-009-0714-x> PMID: 19578876
21. Yong HE, Melton PE, Johnson MP, Freed KA, Kalionis B, Murthi P, et al. Genome-wide transcriptome directed pathway analysis of maternal pre-eclampsia susceptibility genes. *PLoS One* 2015 May 26; 10(5):e0128230. <https://doi.org/10.1371/journal.pone.0128230> PMID: 26010865
22. Thornton JG, Onwude JL. Pre-eclampsia: discordance among identical twins. *BMJ* 1991 Nov 16; 303(6812):1241–1242. PMID: 1747646
23. Treloar SA, Cooper DW, Brennecke SP, Grehan MM, Martin NG. An Australian twin study of the genetic basis of preeclampsia and eclampsia. *Am J Obstet Gynecol* 2001 Feb; 184(3):374–381. <https://doi.org/10.1067/mob.2001.109400> PMID: 11228490
24. Kaartokallio T, Wang J, Heinonen S, Kajantie E, Kivinen K, Pouta A, et al. Exome sequencing in pooled DNA samples to identify maternal pre-eclampsia risk variants. *Sci Rep* 2016 Jul 7; 6:29085. <https://doi.org/10.1038/srep29085> PMID: 27384325
25. Hazelett DJ, Conti DV, Han Y, Al Olama AA, Easton D, Eeles RA, et al. Reducing GWAS Complexity. *Cell Cycle* 2016; 15(1):22–24. <https://doi.org/10.1080/15384101.2015.1120928> PMID: 26771711
26. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015 Nov 17; 7:449–490. <https://doi.org/10.2147/CLEP.S91125> PMID: 26604824

27. Knudsen LB, Olsen J. The Danish Medical Birth Registry. *Dan Med Bull* 1998 Jun; 45(3):320–323. PMID: [9675544](#)
28. Schmidt M, Pedersen L, Sorensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol* 2014 Aug; 29(8):541–549. <https://doi.org/10.1007/s10654-014-9930-3> PMID: [24965263](#)
29. Skjaerven R, Vatten LJ, Wilcox AJ, Ronning T, Irgens LM, Lie RT. Recurrence of pre-eclampsia across generations: exploring fetal and maternal genetic components in a population based cohort. *BMJ* 2005 Oct 15; 331(7521):877. <https://doi.org/10.1136/bmj.38555.462685.8F> PMID: [16169871](#)
30. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature* 2015 Oct 1; 526(7571):68–74. <https://doi.org/10.1038/nature15393> PMID: [26432245](#)
31. Exome Variant Server, NHLB GO Exome Sequencing Project (ESP), Seattle, WA. Available at: <http://evs.gs.washington.edu/EVS>. Accessed February/21, 2017.
32. Allele Frequency Community. Available at: <http://www.allelefrequencycommunity.org/>. Accessed February/21, 2017.
33. Exome Aggregation Consortium. Available at: <https://doi.org/10.1101/030338>. Accessed February/21, 2017.
34. gnomAD. Available at: <https://doi.org/10.1101/030338>. Accessed February/21, 2017.
35. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015 May; 17(5):405–424. <https://doi.org/10.1038/gim.2015.30> PMID: [25741868](#)
36. Human Gene Mutation Database. Available at: <http://www.hgmd.cf.ac.uk/ac/index.php>. Accessed February/21, 2017.
37. CLINVAR database. Available at: <https://www.ncbi.nlm.nih.gov/clinvar/>. Accessed February/21, 2017.
38. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; 4(7):1073–1081. <https://doi.org/10.1038/nprot.2009.86> PMID: [19561590](#)
39. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010 Apr; 7(4):248–249. <https://doi.org/10.1038/nmeth0410-248> PMID: [20354512](#)
40. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014 Mar; 46(3):310–315. <https://doi.org/10.1038/ng.2892> PMID: [24487276](#)
41. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; 11(2–3):377–394. <https://doi.org/10.1089/1066527041410418> PMID: [15285897](#)
42. PathCards. Available at: <http://pathcards.genecards.org/>. Accessed February/21, 2017.
43. QIAGEN Knowledge Base. Accessed on 21 February 2017. Available at: <https://www.qiagenbioinformatics.com/products/qiagen-knowledge-base>. Accessed February/21, 2017.
44. Burda P, Schafer A, Suormala T, Rummel T, Burer C, Heuberger D, et al. Insights into severe 5,10-methylenetetrahydrofolate reductase deficiency: molecular genetic and enzymatic characterization of 76 patients. *Hum Mutat* 2015 Jun; 36(6):611–621. <https://doi.org/10.1002/humu.22779> PMID: [25736335](#)
45. Tonetti C, Saudubray JM, Echenne B, Landrieu P, Giraudier S, Zittoun J. Relations between molecular and biological abnormalities in 11 families from siblings affected with methylenetetrahydrofolate reductase deficiency. *Eur J Pediatr* 2003 Jul; 162(7–8):466–475. <https://doi.org/10.1007/s00431-003-1196-9> PMID: [12733064](#)
46. Lin J, August P. Genetic thrombophilias and preeclampsia: a meta-analysis. *Obstet Gynecol* 2005 Jan; 105(1):182–192. <https://doi.org/10.1097/01.AOG.0000146250.85561.e9> PMID: [15625161](#)
47. Yuan Q, Yang J, Santulli G, Reiken SR, Wronska A, Kim MM, et al. Maintenance of normal blood pressure is dependent on IP3R1-mediated regulation of eNOS. *Proc Natl Acad Sci U S A* 2016 Jul 26; 113(30):8532–8537. <https://doi.org/10.1073/pnas.1608859113> PMID: [27402766](#)
48. Loset M, Mundal SB, Johnson MP, Fenstad MH, Freed KA, Lian IA, et al. A transcriptional profile of the decidua in preeclampsia. *Am J Obstet Gynecol* 2011 Jan; 204(1):84.e1–84.27.
49. Banfi S, Servadio A, Chung MY, Kwiatkowski TJ Jr, McCall AE, Duvick LA, et al. Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nat Genet* 1994 Aug; 7(4):513–520. <https://doi.org/10.1038/ng0894-513> PMID: [7951322](#)

50. Tong X, Gui H, Jin F, Heck BW, Lin P, Ma J, et al. Ataxin-1 and Brother of ataxin-1 are components of the Notch signalling pathway. *EMBO Rep* 2011 May; 12(5):428–435. <https://doi.org/10.1038/embor.2011.49> PMID: [21475249](https://pubmed.ncbi.nlm.nih.gov/21475249/)
51. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003 Mar 8; 361(9360):865–872. PMID: [12642066](https://pubmed.ncbi.nlm.nih.gov/12642066/)