The Effects of Genetic Variants on Protein Structure and their Associations with Preeclampsia

Elizabeth Geena Woo

Undergraduate Honors Thesis
April 20, 2020

Thesis Advisor: Alper Uzun, PhD
Second Reader: James Padbury, MD

A thesis submitted in partial fulfillment for the degree of Bachelor of Arts With Honors in Computational Biology

Table of Contents

Introduction

Preeclampsia is a complex pregnancy-specific disorder characterized by the onset of maternal hypertension and proteinuria.[1,2] This multifactorial disorder complicates 2-8% of US deliveries and is a major cause of maternal and fetal morbidity and mortality.[3] Preeclamptic pregnancies are associated with long-term outcomes for both the mother and offspring. Stroke, cardiovascular disease, diabetes, and premature mortality are linked to preeclampsia in affected mothers later in life, as well as higher blood pressure and increased risk of stroke in offspring.[4-8] The heritability of preeclampsia is estimated at about 55%, with contributions from both maternal and fetal genes.[9] While the etiology of preeclampsia remains poorly understood, there are ongoing efforts to better understand the disease mechanisms and its genetic underpinnings with the eventual goal of prevention and early detection.[10] Given that preeclampsia is a complex disease influenced by multiple genes and environmental factors, this remains a challenge. Understanding the effects of disease-associated variants on protein function can provide insight into the functional relationships between genotype and disease-associated phenotypes.

By investigating the potential structural effects and protein-protein interaction networks, we can identify candidates for further functional study and experimental validation. Here, we examine the structural effects of genetic variants and their associations with preeclampsia, identifying preeclampsia-associated genes and their protein products, and mapping pathogenic variants on these protein structures. Variants were analyzed on both individual structural and protein-protein interaction levels to identify whether the resulting effects are associated with pathogenesis in preeclampsia. Existing literature was searched to find previously known information on genes in relation to preeclampsia as well as how their respective protein

structures function on a molecular level, which can be informative in predicting how a specific variant will affect function.

Methods and Materials

Variant filtration

Whole-exome sequencing data came from a case-control study for preeclampsia at the Women & Infants Hospital in Providence, Rhode Island. This patient cohort includes 143 carefully phenotyped patients: 61 preeclamptic mothers with severe features and 82 controls. 528,631 variants were identified with a depth of coverage greater than 10x. Among these, 747 variants in 679 genes were considered pathogenic and included for further investigation (Fig. 1). Variants were considered pathogenic if they were rare with EXACT minor allele frequencies below 0.05, considered deleterious both by SIFT and PolyPhen, and had CADD-Phred scores greater than or equal to 30 (indicating the variant is amongst the top 0.1% of deleterious variants).[11-13]

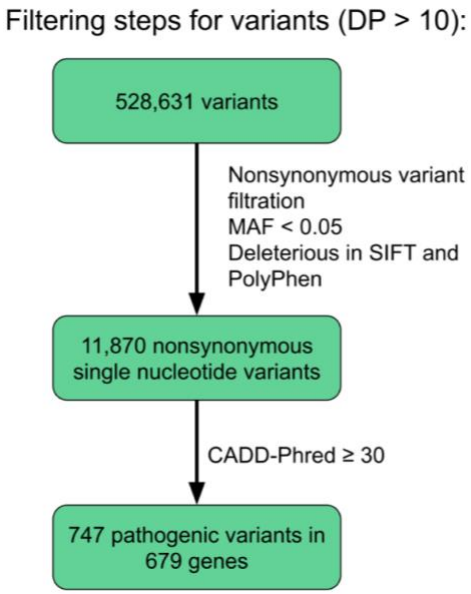

Figure 1: Filtration of variants (depth of coverage > 10)

Accession number annotation

NCBI protein database accession numbers were annotated for each variant using reference SNP RSIDs. In order to find RSIDs for all variants, chromosomal positions were converted from GRCh37 to GRCh38 using UCSC Genome Browser Liftover.[14] RSIDs were identified based on the GRCh38 position using dbSNP advanced search.[15] Some of the RS numbers (found from the chromosomal position) were associated with multiple distinct amino acid changes. VCF files were consulted to find the allele change (exact variant) and the corresponding amino acid change was identified using dbSNP. Using RSID lookup in RefSNP from dbSNP variation services (https://api.ncbi.nlm.nih.gov/variation/v0/), associated NP accession numbers for each variant were identified. In most cases, there was only one associated NP number returned by RefSNP for each variant. For some of the variants, however, RefSNP returned multiple NP numbers. In order to decide which NP number to use, we cross-referenced with the HUGO Gene Nomenclature Committee (HGNC) database and GenBank.[16,17] For a variant associated with multiple NP numbers, the corresponding NM accession number was located in the HUGO Gene Nomenclature Committee (HGNC) database. The NM number found in HGNC was then queried in GenBank to find the associated NP number.

Protein mapping and domains

Each variant's associated NP number was queried in the NCBI Protein database and the protein domain whose "Region" included the amino acid substitution position was included as the domain. UniProt and Interpro were also consulted for domain information.[18,19]

Identifying genes associated with preeclampsia

679 genes containing 747 pathogenic variants from the patient cohort were compared to genes found in Database for Preeclampsia (dbPEC) in order to identify variants most likely to be relevant to preeclampsia.[20] Among these patient cohort genes, 27 were found to be preeclampsia-related in dbPEC (Fig. 2). FASTA files and sequences were obtained for each of these 27 filtered genes and then run in Protein BLAST using the Protein Data Bank proteins database (without limiting results to Homo sapiens) in order to build structural models.[21] Of the 27 genes, 15 returned an alignment that covered the variant and had at least 50% percent identity. The PDB IDs of suitable protein structures were noted for use in subsequent modeling. If multiple results met this criterion, the alignment with the highest percent identity was chosen.

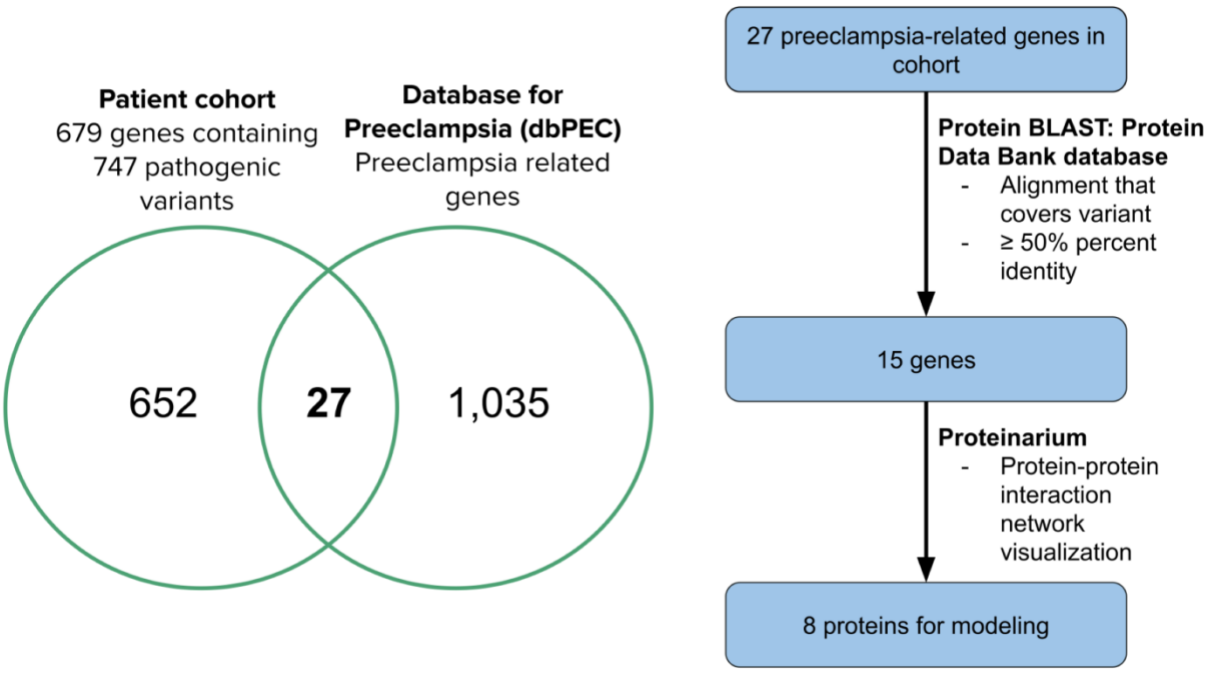

Figure 2: Identification of genes associated with preeclampsia

*dbPEC Obstet Gynecol 123(6):1155-61, 2014.

3D Protein Modeling

SWISS-MODEL was used to build structural protein models.[22] The original sequence (without variants) was queried to build a native protein model using the PDB that covers the variant position. An altered sequence including the variant was created and the resulting new protein structure with the variant. Protein structures were visualized using UCSF Chimera, which allows for visualization between specific residues.[23] The location of the variant in the protein structure was investigated for structural differences and potential effects on protein function.

Protein-protein interactions

Proteinarium was used for visualization and analysis of protein-protein networks and interactions for the 15 genes of interest.[24] Proteinarium is a protein-protein interaction network visualization and analysis tool that is built by mapping proteins onto the interactome provided by STRING database.[25] Stringent criteria using a maximum path length of two allows for a maximum of one imputed gene between interacting input genes. Of the 15 genes from the patient cohort, 8 were included in the resulting protein-protein interaction network, each connected to another by a single imputed gene (a gene that was not found in the patient cohort).

Literature search

Literature searches were performed to find existing information on genes in relation to preeclampsia as well as how their respective protein structures function on a molecular level. This can be informative in predicting how a specific variant will affect function.

Software and databases

Analyses were performed using R version 3.3.2 (2016-10-31) and R package tidyverse_1.2.1. The rentrez_1.2.1 package was used for dbSNP query (https://cran.r-project.org/web/packages/rentrez/vignettes/rentrez_tutorial.html). The jsonlite_1.5 package (https://cran.r-project.org/web/packages/jsonlite/index.html) was used to parse JSON files returned by RefSNP.

Results and Discussion

We concentrated our investigation on the 8 genes that were included in the resulting protein-protein interaction network. We considered both the functional effects of identified variants and their structural effects. As it was part of the filtering during selection, each of the variants we studied further had significant predicted pathogenicity scores (Table 1).

Protein-Protein Interaction Network

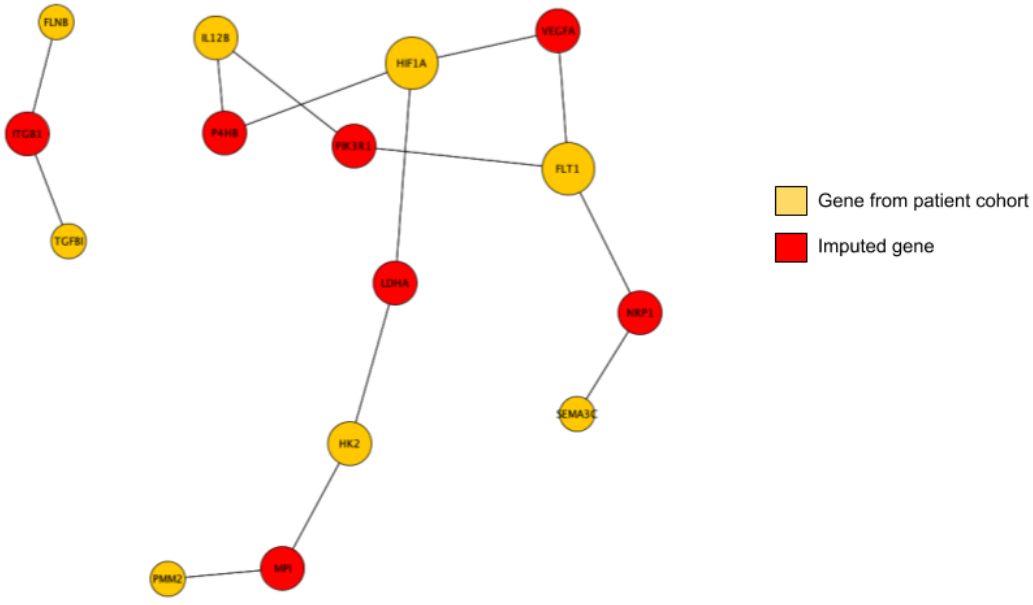

Figure 3: Protein-protein interaction network with maximum path length of two.

Proteinarium is a protein-protein interaction network visualization and analysis tool that is built by mapping proteins onto the interactome provided by STRING database.[24] Using a maximum path length of two, fifteen preeclampsia-related genes were input to Proteinarium, outputting two networks (Figure 3). Of the fifteen genes originally input (from the patient cohort, shown in red above), eight ended up in the resulting networks, interacting through seven imputed genes (not in the cohort, shown in yellow in Fig. 3). Network A is the smaller three-membered network, and Network B is the larger, 12-gene network. Network A includes two genes from the patient cohort, *FLNB* and *TGFBI*, interacting with imputed gene *ITGB1*. Network B includes six genes from the patient cohort (*IL12B, HIF1A, FLT1, SEMA3C, HK2, PMM2*). *P4HB, MPI, NRP1, LDHA, PIK3R1, and VEGFA* are imputed genes of Network B. *HK2* expression is regulated by HIF1-α.[26] *HK2*, *PMM2*, and *MPI* are involved in fructose and mannose metabolism pathways.[27-29]

Functional

FLNB

Two pathogenic variants in *FLNB* were found in the patient cohort. The first (rs143831841) is NP_001448.2:p.Phe1411Leu which results in a non-synonymous substitution of phenylalanine for leucine at position 1411. The second (rs111330368) NP_001448.2:p.Gly1262Ala was a non-synonymous substitution of glycine for alanine at position 1262. As can be seen in Table 1, both of these variants were associated with significant pathogenicity scores (Table 1).

FLNB encodes the filamin B protein, a cytoplasmic protein that helps to regulate cytoskeleton structure and activity. Human filamins are 280-kDa proteins involved in cross-linking actin into three-dimensional networks.[30,31] The filamin monomer contains an N-terminal actin binding

domain, which is followed by a series of 24 β-sheet repeats that have a role in protein-protein interaction and bind many cytoplasmic and transmembrane proteins.

Phe1411 lies on the Filamin 12 repeat and Gly1262 on the Filamin 11 repeat.[19] These are among the regions in *FLNB,* including repeats 10 through 13, that bind to FBLP-1 (filamin-binding LIM protein-1), a protein that has been identified as a key regulator of bone homeostasis.[32,33]

TGFBI (βig-h3)

Variant NP_000349.1:p.Pro501Thr (rs121909212) was found in our cohort which results in a non-synonymous substitution of proline for threonine at position 501 and was associated with significant pathogenicity scores (Table 1). The VCF file contained a chr5:135391459C>A substitution, corresponding to a p.Pro501Thr amino acid change. Pro501 lies on the FAS1-4 domain.[34] Variant P501T has been previously been implicated in LCD IIIA corneal dystrophy, with the wild-type residue functioning in inter-domain surface interactions.[35]

TGFBI (also known as beta ig h3) codes for the human protein transforming growth factor beta induced (TGFBI), which becomes part of the extracellular matrix.[36] TGFBI interacts with integrins and is involved in several physiopathological conditions, including wound healing, angiogenesis, cancer, and inflammatory diseases.[37,38] It is a major component of the human cornea and is associated with post-natal corneal development and maturation.[39,40] TGFBI is an RGD-containing protein that binds to type I, II, and IV collagens. The RGD motif serves as a ligand recognition sequence for several integrins. The protein is induced by transforming growth factor-beta and acts to inhibit cell adhesion. The TGFBI protein (TGFBIp) contains four

evolutionarily conserved fasciclin-1 (FAS1) domains and a carboxy-terminal Arg-Gly-Asp (RGD) sequence.[41,42] The FAS1 domains mediate cell adhesion through an interaction with alpha3/beta1 integrin.[43] The cooperative action of multiple FAS1 domains and the RGD motif in circulating TGFBI can inhibit vascular endothelial cell movement, tumor angiogenesis, and tumor growth through interactions with αvβ3 integrin.[44]

FLT1

Variant NP_002010.2:p.Arg1060Gln (rs753623232) was found in the patient cohort and associated with significant pathogenicity scores (Table 1). *FLT1*, also known as Fms Related Tyrosine Kinase 1 or VEGFR-1, encodes a member of the vascular endothelial growth factor receptor (VEGFR) family. VEGFR-1 is one of three members (VEGFR-1/Flt-1, VEGFR-2/KDR/Flk-1, VEGFR-3/Flt-4) in the VEGF receptor family, which bind to VEGFs, which are involved in angiogenesis, vasculogenesis, and de novo formation of vessels.[45-48] Members of this family are receptor tyrosine kinases (RTKs) that contain an extracellular ligand-binding region with seven immunoglobulin (Ig)-like domains, a transmembrane segment, and a tyrosine kinase (TK) domain within the cytoplasmic domain. The FLT1 protein binds to VEGFR-A, VEGFR-B and placental growth factor and plays an important role in angiogenesis and vasculogenesis. It may play an essential role as a negative regulator of embryonic angiogenesis by inhibiting excessive proliferation of endothelial cells. The VEGFR-1 protein contains seven Ig-like domains in the extracellular region and a tyrosine kinase domain with a long kinase insert.[45] It has been widely implicated in preeclampsia. Elevated circulating levels early in pregnancy have been associated with greater risk of subsequent development of preeclampsia.[49] Through its

antiangiogenic effects it likely interferes with placental angiogenesis resulting in the shallow invasion and local hypoxia associated with this disorder.[50]

Arg1060 is found on the catalytic protein tyrosine kinase domain.[18,19] The tyrosine kinase domain transduces signals for endothelial cells.[45]

SEMA3C

Variant NP_006370.1:p.Val321Ala (rs140244551) in *SEMA3C* was found in the patient cohort with significant pathogenicity scores (Table 1). *SEMA3C* encodes a secreted glycoprotein that belongs to the semaphorin class 3 family of neuronal guidance cues, containing an N-terminal sema domain, integrin and immunoglobulin-like domains, and a C-terminal basic domain.[36] SEMA3C is required for normal cardiovascular development during embryogenesis.[19] It has been shown to increase cell proliferation and migration, as well as promote integrin signaling and VEGF secretion in endothelial cells.[51] NRP1 and NRP2 are among the receptors to SEMA3C which activate VEGFR signaling.[52]

Val321 is in the conserved 500 amino acid Sema domain, a ligand-receptor interaction module located at the N terminus.[19,53] It contains 4 disulfide bonds formed by eight conserved cysteine residues, and serves as a receptor-recognition and -binding module.[54]

HIF1A

Variant NP_001521.1:p.Glu261Lys (rs540883086) was found in the patient cohort (Table 1). *HIF1A* codes for the alpha subunit of the hypoxia-inducible factor-1 (HIF-1) transcription factor. HIF-1α is one of three HIF-α proteins (HIF-1α, HIF-2α, HIF-3α) that mediate the adaptive transcriptional response to hypoxia and heterodimerize with the aryl hydrocarbon receptor nuclear translocator (ARNT) in order to function as productive transcription factors.[55-59] The HIF-α proteins function as the oxygen-sensitive regulatory subunits of HIF-α-ARNT heterodimers, and HIF-1α specifically is critical for the oxygen stress response.[55,60] It helps to regulate homeostatic response to hypoxia by activating transcription of other genes with roles in energy metabolism, angiogenesis, apoptosis, and others that aid in metabolic adaptation to hypoxia. It plays an important role in embryonic vascularization, tumor angiogenesis, and pathophysiology of ischemic disease.[36] It is a major regulator of trophoblast differentiation, which are the cells in placenta that facilitate the transport of nutrients, gases, and wastes between the mother and fetus.[61]

Glu261 is part of the PAS-B Cα domain and is aligned with Glu263 from HIF-2α.[62] The neighboring amino acid, L_{262} in HIF-1α, is a cancer (malignant melanoma) related mutation found in the same domain, with a possible role of internal stability.

IL12B

Variant NP_002178.2:p.Val298Phe (rs3213119) was found in the patient cohort with significant pathogenicity scores (Table 1). *IL12B* encodes the 40 kD receptor like subunit of interleukin 12, a cytokine that acts on T and natural killer cells and plays an important role in cell-mediated

immunity.[63] Interleukin 12 is a disulfide-linked heterodimer composed of the 40 kD cytokine receptor like subunit encoded by this gene, and a 35kD subunit encoded by IL12A.[36] The IL-12p40 subunit can also combine with a different subunit to form IL-23.[64] Alterations in pro-inflammatory cytokines have been widely implicated in the pathogenesis of preeclampsia.[65]

Val$_{298}$ is in the "FN3", or Fibronectin type-III, domain.[19] This is one of the three types of internal repeats found in the plasma protein fibronectin.[54]

HK2

Variant NP_000180.2p.Ala836Thr (rs147555773) in the *HK2* gene was found in the patient cohort with significant pathogenicity scores (Table 1). Hexokinase-2 (HK2) is an enzyme encoded by the *HK2* gene that catalyzes the phosphorylation of glucose as the first functionally irreversible reaction in the glycolytic pathway.[26] Ala$_{836}$ is in the hexokinase 2 C-terminal domain.[19]

PMM2

Variant NP_000294.1:p.Arg123Gln (rs141498002) in the *PMM2* gene was found in the patient cohort with significant pathogenicity scores (Table 1). PMM2 is a member of the haloalkanoate dehalogenase superfamily (HADSF) containing four conserved sequence motifs.[66]

Phosphomannomutase 2 (PMM2), is an enzyme encoded by the *PMM2* gene that catalyzes the mutase reaction providing mannose 1-phosphate (Man-1-P), essential for protein glycosylation.[67] Man-1-P is an important intermediate for the N-linked polysaccharide synthesis pathway as a

precursor to GDP-mannose.[68] In post-translational protein glycosylation, the mannose moiety is transferred from GDP-mannose to the growing oligosaccharide chain in the endoplasmic reticulum.[69] The PMM isozymes (PMM1 and PMM2) also have a "cap" domain between motifs 2 and 3 made of four alpha-helices and a four-stranded antiparallel beta-sheet.[66,70]

Arg123 lies on the cap domain of phosphomannomutase 2. The cap domain dissociates from the core domain, allowing the substrate to bind, and then associates to seal the active site from aqueous solvent.[71,72]

Table 1: Variants in 8 preeclampsia-associated genes from patient cohort. Chromosomal positions were based on the hg19 build.

Gene	Chromosomal position	Variant	RefSNP accession	MAF	CADD Phred	PolyPhen2 HDIV	SIFT
FLNB	chr3:58116478	p.Phe1411Leu	rs143831841	0.0007	35	0.999	0
FLNB	chr3:58110119	p.Gly1262Ala	rs111330368	0.0055	33	1	0
TGFBI	chr5:135391459	p.Pro501Thr	rs121909212	0.0003	31	1	0.01
FLT1	chr13:28893667	p.Arg1060Gln	rs753623232	1.77E-05	33	1	0.02
SEMA3C	chr7:80430097	p.Val321Ala	rs140244551	0.0038	31	0.97	0
HIF1A	chr14:62199143	p.Glu261Lys	rs540883086	8.28E-06	34	0.999	0
IL12B	chr5:158743788	p.Val298Phe	rs3213119	0.0206	35	0.978	0.01
HK2	chr2:75116502	p.Ala836Thr	rs147555773	2.49E-05	33	1	0
PMM2	chr16:8904956	p.Arg123Gln	rs141498002	0.0001	36	0.998	0

Structural

FLNB

p.Phe1411Leu was modeled using Protein Data Bank accession code 2DI9_A. This position lies on a beta sheet, with relatively close proximity to Ala1333 (~3.959 Å) and Leu1338 (~3.863 Å). The wild-type amino acid, Phe, is an aromatic, hydrophobic amino acid. It is fairly non-reactive, and therefore rarely directly involved in protein function, although it can play a role in substrate recognition.[73] It has a molecular weight of 165.2 g/mol.[74] The variant amino acid, Leu, is ~3.703 Å from Ala1333 and ~3.826 and ~2.972 Å from Leu1338 (variant Leu pictured in red in Fig. 4B). The variant amino acid, leucine (MW: 131.2 g/mol) is an aliphatic, hydrophobic amino acid.[73]

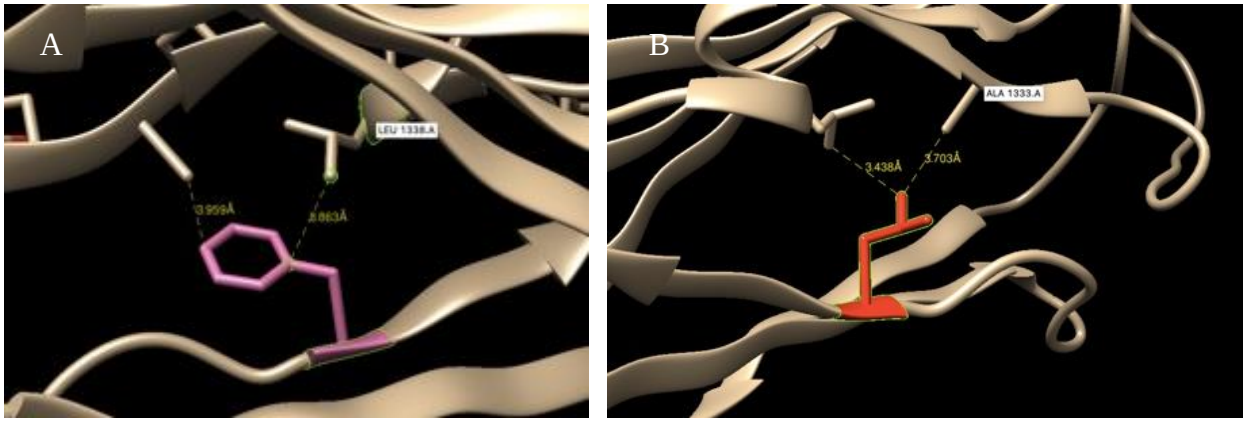

Figure 4: FLNB amino acid position 1411 and distances from nearby neighboring amino acids. (A) Wild-type associated amino acid, Phe. (B) Variant associated amino acid, Leu.

p.Gly1262Ala was modeled using PDB accession 2DIB_A. As a small residue (75.07 g/mol), it does not protrude far into space. Glycine has hydrogen as its side chain, lending itself to conformational flexibility that allows it to reside in tight turns in structures that other amino acids wouldn't. It can also play a distinct functional role, for example using its sidechain-less backbone to bind to phosphates.[73] The variant amino acid, alanine (MW: 89.09 g/mol) is non-polar and generally non-reactive, rarely directly involved in protein function. It is closer to

Asp1284 (~2.605 Å) than Glycine was. Gly1262 corresponds to Gly55 in Chain A on PDBsum, which has previously been indicated as a human variant residue.[75] The introduction of the variant, Alanine, rotates the structure such that Ala55 is 2.71 Å away from nearby Thr52.

TGFBI (βig-h3)

p.Pro501Thr was modeled using PDB accession 5NV6_1A (Fig. 5). p.Pro501Thr changes a nonpolar residue to polar residue. Proline (MW: 115.13 g/mol) has a 5-membered nitrogen-containing ring, and is less flexible than other amino acids. It is often found in very tight turns in protein structures and is important in producing bends in a peptide chain. It is aliphatic and hydrophobic, and rarely involved in protein active or binding sites.[73] Variant amino acid Threonine (MW: 119.12 g/mol) is slightly polar. While most amino acids only have one non-hydrogen substituent attached to their C-beta carbon, threonine contains two, leading to more bulkiness near the protein backbone and therefore restricting the conformations that the main chain can adopt. The FAS1-4 domain and the RGD motif are part of the carboxy-terminus of TGFBI.

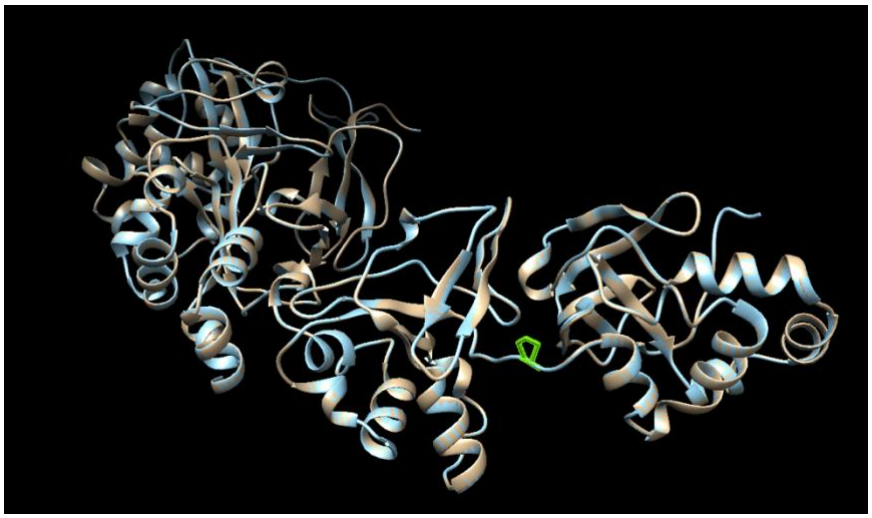

Figure 5: Pro501 in TGFBI (PDB accession 5NV6_1A) is highlighted in green.

FLT1 (VEGFR1)

p.Arg1060Leu was modeled using PDB accession 3HNG. Wild-type Arg1060 is relatively close to Arg1055 on the opposite side of the loop (~4.503 Å), and Tyr1048 (~4.414 Å). It is not on a beta sheet or alpha helix (Fig. 6). Arginine (MW: 174.2 g/mol) is basic, positively charged and polar, while the variant amino acid, leucine (MW: 131.2 g/mol) is a smaller aliphatic, hydrophobic amino acid. This change in charge and physical and chemical properties could affect molecular interactions and stability, for which further molecular dynamics simulations may prove useful.

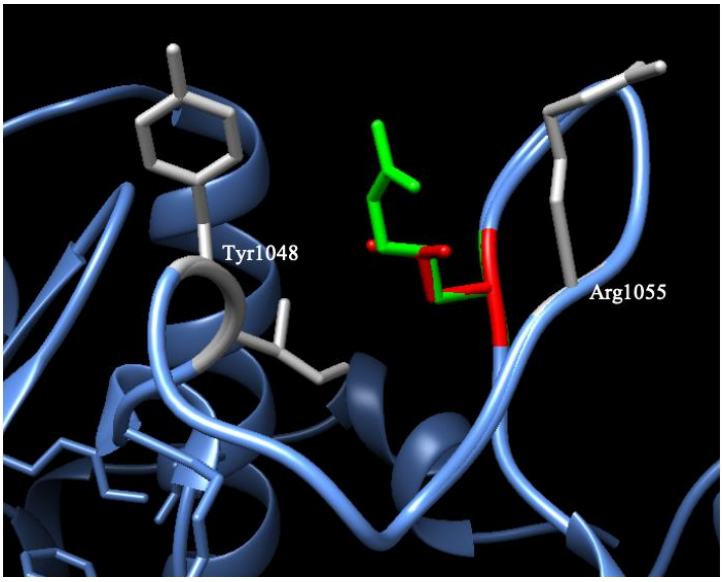

Figure 6: FLT1 (VEGFR1) position 1060 (PDB ID 3HNG). Amino acid without variant, Arg (green) and amino acid with variant, Leu (red). Closest neighboring amino acids in grey.

SEMA3C

Semaphorins have a 500–amino acid semaphorin domain that folds into a seven-bladed β-propeller. This domain is critical to protein-protein interactions between semaphorins and their receptors including neuropilins (NRPs) and plexins (PLXNs).[76] A SEMA3C homology model was constructed using the one of the monomers from the SEMA3A dimer complex, PDB ID 4GZ8_1A.[77] This model was used to visualize p.Val321Ala, which lies on a beta sheet (Fig. 7).

The wild-type amino acid Valine (MW: 117.15 g/mol) is an aliphatic, hydrophobic amino acid. It is C-beta branched, making it bulkier near the protein backbone.[73] Valine prefers to lie within beta-sheets, as seen in this case, rather than adopt an alpha-helical conformation. The variant amino acid, Alanine (MW: 89.09 g/mol) is non-polar and generally non-reactive. The protrusion of these amino acids into space does not appear to differ significantly.

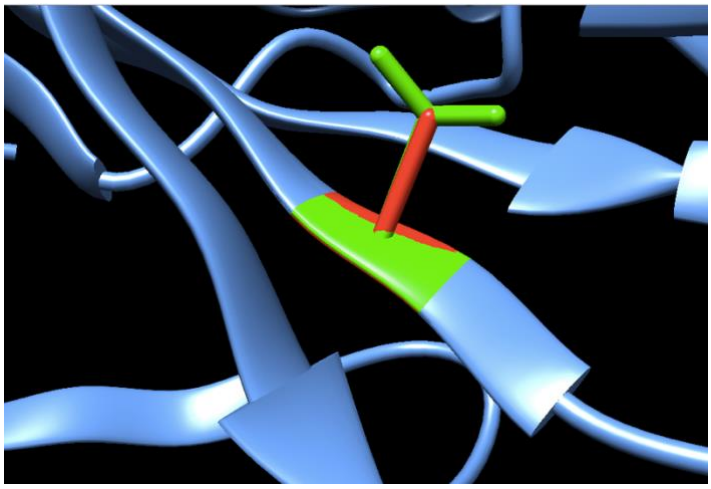

Figure 7: SEMA3C position 321 (PDB ID 4GZ8_1A). Amino acid without variant, Valine (green) and amino acid with variant, Alanine (red).

HIF1A

PDB ID 4ZPR_1B was used to model HIF1A and visualize p.Glu261Lys. Glu261 is found on the surface of the protein (Fig. 8A). The wild-type amino acid, Glutamate (MW: 146.12 g/mol), is a large, negatively charged, polar amino acid. The variant amino acid, Lysine (MW: 146.19 g/mol) is basic, positively charged, and polar. The change from negative to positive charge could have effects on protein stability or interactions.

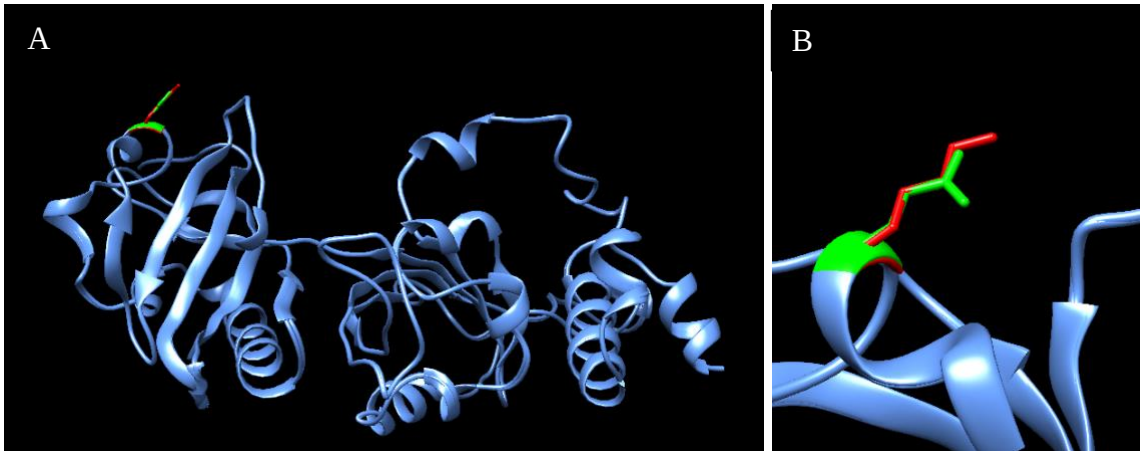

Figure 8: HIF1A position 261 (PDB ID 4ZPR_1B). (A) Glutamate at position 261 is on the surface. (B) Amino acid without variant, Glutamate (green) and amino acid with variant, Lysine (red).

IL12B

PDB ID 4GRW_1B was used for modeling. Val298 lies on a beta sheet. It lies relatively close to Val289 (~3.753 Å) and Val275 (~4.416 Å). The variant replaces Valine with Phenylalanine, which extends closer to these nearby amino acids, with ~2.283 Å to Val289 and ~2.714 Å to Val275 (Fig. 9). The wild-type amino acid Valine (MW: 117.15 g/mol) is an aliphatic, hydrophobic amino acid. The variant amino acid Phe (MW: 165.19 g/mol) is aromatic and hydrophobic.

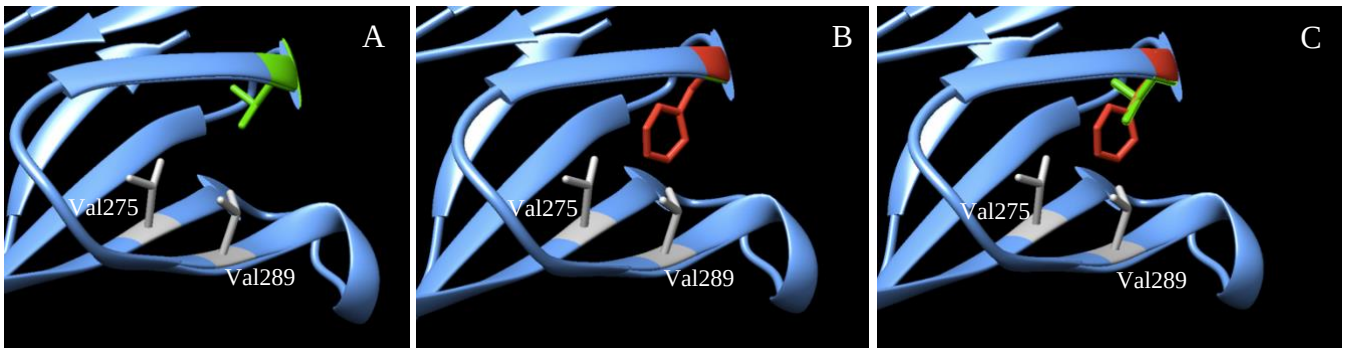

Figure 9: IL12B position 298 (PDB ID 4GRW_1B) and nearby amino acids, Val289 and Val275 (shown in gray). (A) Amino acid without variant, Val, in green. (B) Amino acid with variant, Phe. (C) Both amino acids.

HK2

Protein Data Bank accession code 5HEX was used for modeling. The amino acid with the wild type allele, alanine (MW: 89.09 g/mol), is hydrophobic, uncharged, and non-polar. The amino acid with the variant, threonine (MW: 119.12 g/mol), is also uncharged. The original amino acid, Ala836 lies on an alpha helix, but it is difficult for threonine to adopt an alpha-helical conformation (Fig. 10). Threonine is slightly polar, but has two non-hydrogen substituents, which contributes to more bulkiness near the protein backbone and restricts the conformations that the main chain can adopt. This change may have steric effects.

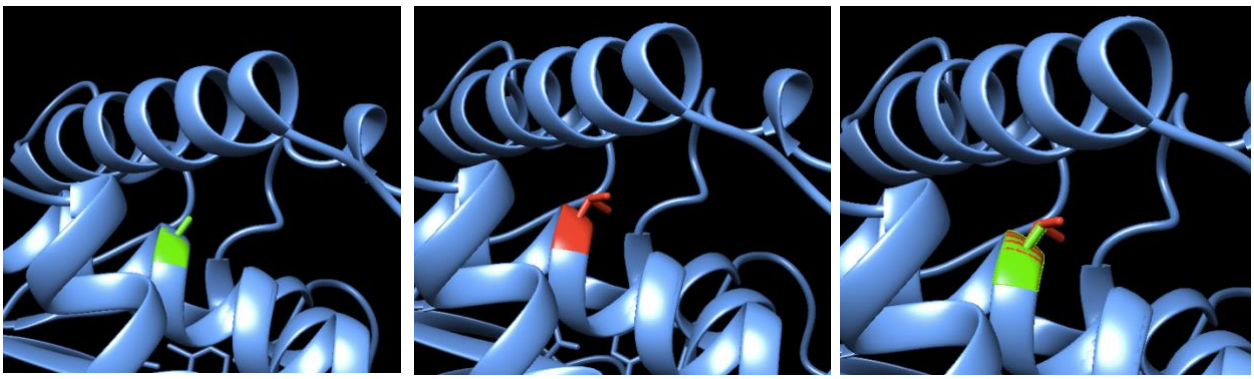

Figure 10: HK2 position 836 (PDB ID 5HEX). (A) Amino acid without variant, Ala, in green. (B) Amino acid with variant, Thr. (C) Both amino acids.

PMM2

Protein Data Bank accession code 6CFT was used for modeling. Structurally, Arg123 is oriented spatially in the same direction as the large neighboring amino acids, including Glu121, Met126, and Asn128 (Fig. 11). Arg123 in α-PMM2 is a conserved residue in a region of the cap domain that would contact the core domain and contains many positively charged residues.[78] In isozyme PMM1, the paralog amino acid Arg132 is involved in Man-1-P binding, which along with Asp190 holds the C-2 hydroxyl. Arg132 directly contacts the substrate. The Man-1-P-bound structure of paralog PMM1 shows that Arg132 of α-PMM1 (similar to Arg123 in α-PMM2) makes direct interactions with the C-2 hydroxyl of Man-1-P.[78] Mutations at this position would be likely to affect substrate binding. The wild-type amino acid, Arginine (MW: 174.2 g/mol) is a positively charged, polar amino acid. The variant amino acid, Glutamine (MW: 146.14 g/mol), while also polar, is uncharged. This change in charge could affect substrate binding.

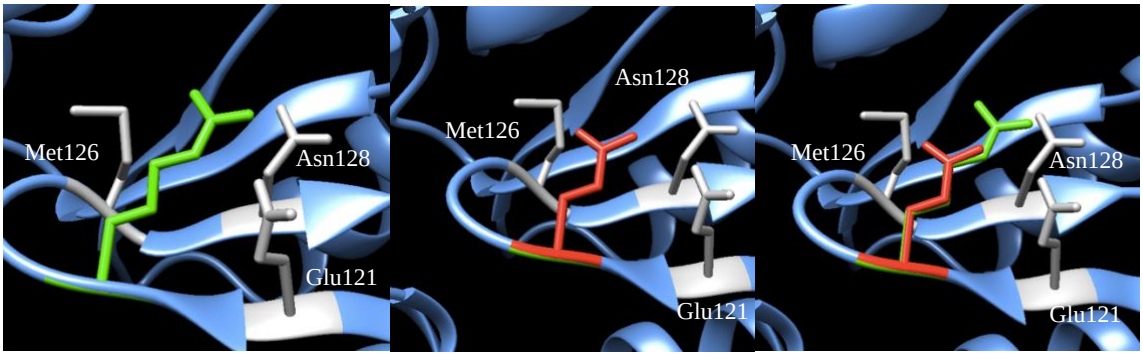

Figure 11: PMM2 position 123 (PDB ID 6CFT) with nearby amino acids Glu121, Met126, and Asn128 (in gray). (A) Amino acid without variant, Arg, in green. (B) Amino acid with variant, Gln. (C) Both amino acids.

Associations of identified variants with Preeclampsia

FLNB

FLNB was found to be upregulated in placentas obtained from women with severe preeclampsia.[79] Filamin B is concentrated mainly in endothelial cells.[80] It has a role in facilitating endothelial cell migration following VEGF stimulation, which is an essential process for angiogenesis, or the formation of new blood vessels from preexisting vasculature. Disruption of filamin B function is associated with loss of the proangiogenic effect of VEGF in 3D collagen gels and inhibition of cell migration.[80] In preeclampsia, placental anti-angiogenic factors are upregulated and released into the maternal circulation, disrupting the maternal endothelium and resulting in hypertension, proteinuria, and other features of preeclampsia.[81,82]

DNA methylation, the covalent addition of a methyl group to a cytosine, can regulate gene expression and has been studied in the placenta.[83-85] The placenta allows for the exchange of nutrients, gases, and wastes between the maternal and fetal circulations. In the placenta, trophoblasts are epithelial cells that are thought to originate from trophoblast stem cells and are essential for function. Impaired trophoblast differentiation is associated with preeclampsia.[86] CpG islands are short interspersed DNA sequences that are GC-rich, CpG-rich, predominantly nonmethylated and can serve as sites of transcription initiation.[87] CpG methylation is involved in trophoblast differentiation, but is likely not the only regulatory process involved in this process.[88] In a study of differentially methylated CpGs in early-onset pre-eclampsia and control placental samples, CpGs in *FLNB* showed negative correlation between DNA methylation and gene expression.[89] *FLNB* variants in our cohort are located in chromosome 3, positions 58116478 and 58110119. The CpG island nearest to these variants lies on chr3:58109101-58109369,

approximately 750 bp away from these variants. These variants are therefore within a CpG shore region. Of the variants analyzed in this study, these variants in *FLNB* are the only ones close to a CpG island and in a CpG island shore region. Within this nearby *FLNB* CpG island is the cg06313433 probe, which was found to be significantly differentially methylated between clusters of transcriptional subtypes of preeclampsia (a "canonical" PE cluster, including modifications in TGF-beta signaling, cell adhesion, oxidative phosphorylation, and a clinically healthier cluster).[14,90]

TGFBI (βig-h3)

TGFBI has been found in urinary excretion of preeclamptic patients and may be involved in kidney pathology associated with preeclampsia.[91] TGFBI shows anti-angiogenic activity, which is 100-fold greater than that of the FAS1 domain alone, indicating that anti-angiogenic activity is enhanced by the cooperative conformation of FAS1 domains and the RGD motif.[44] TGFBI effectively inhibits vascular endothelial cell movement, tumor angiogenesis, and tumor growth through direct interactions of the FAS1 domains with αvβ3 integrin. The variant from the patient cohort, p.Pro501Thr, is located on the FAS1-4 domain and has previously been implicated in corneal dystrophy, with the wild-type amino acid at this position playing a role in inter-domain surface interactions. Similarly, it is possible the introduction of a different amino acid at this position could impact these inter-domain surface interactions in the cooperative conformation that influences angiogenic activity.

FLT1 (VEGFR1)

Flt1 (VEGFR-1) is a membrane spanning receptor for vascular endothelial growth factor and placenta growth factor, and is essential for normal placental vascular development. sFlt-1, a soluble form of VEGFR-1, is expressed in trophoblasts and negatively regulates angiogenesis as an antiangiogenic factor. Patients with preeclampsia have been reported to have abnormally high levels of sFlt-1 in serum and plasma, resulting from placental hypoxia.[49,81] Increased levels of sFlt-1 trap circulating vascular endothelial growth factor (VEGF) and decrease their free levels, which can lead to endothelial dysfunction and an anti-angiogenic state, leading to preeclamptic features. Failure of placental angiogenesis during early development has been linked to the pathogenesis of preeclampsia.[92]

SEMA3C

Sema3C has been reported to inhibit angiogenesis by signaling via Nrp1 and Plexin D1.[93] Neuropilin-1 (NP-1) was first identified as a semaphorin receptor involved in neuron guidance. Subsequent studies demonstrated that NP-1 also binds an isoform of vascular endothelial growth factor (VEGF) as well as several VEGF homologs, suggesting that NP-1 may also function in angiogenesis.[94] Sema3C inhibits VEGF-promoted endothelial cell survival, adhesion, and migration during angiogenesis by causing rearrangement of the actin cytoskeleton, breakdown of adherens junctions, and impairment of focal adhesion formation. Sema3C can act as an anti-angiogenic agent through inhibition of endothelial junction integrity and focal adhesion with the extracellular matrix.[93]

HIF1A

HIF-1α has been found to be overexpressed in women with preeclampsia, resulting in upregulation of antiangiogenic factors sFLT-1, soluble endoglin-1 (sENG-1), and endothelin-1 and leading to preeclamptic manifestations.[95] Upregulation of sENG and endothelin following placental hypoxia is likely mediated in part by HIF-1α. HIF-1α is a critical molecular mediator of placental development and regulates trophoblast differentiation.[96] Prolonged expression of HIF-1α in trophoblast cells of the placenta resulted in pregnancy-specific placental disorganization and inhibition of trophoblast differentiation with endothelial dysfunction.[61]

IL12B

Plasma IL-12 has been detected more frequently in preeclamptic patients, suggesting the involvement of this cytokine in the enhanced immune response observed in preeclampsia.[97]A methylation study done using neonatal umbilical cord blood from infants born at <34 weeks due to early onset preeclampsia found hyper-methylation of IL12B.[98] This corresponded with other findings that preeclampsia patients had significantly less IL-12 in villous trophoblasts and less IL-12 secretion from decidua.[99,100] Taken together, this suggests that IL12B is repressed at the gene expression level in cord blood cells.

HK2

HK2 is important for endometrial decidualization, a process of endometrial stromal cell proliferation and differentiation.[101] Endometrial decidualization plays a significant role in embryo implantation and maintenance of pregnancy, as it can regulate trophoblast invasion, uterine spiral artery remodeling, and placental formation.[102-104] Defective decidualization may

contribute to preeclampsia by impairment of early stages of placental invasion and development.[105] Downregulation of *HK2* was shown to inhibit human endometrial stromal cells proliferation and differentiation.[101] *HK2* stimulates glucose uptake and lactate production of human endometrial stromal cells. Reduced expression may contribute to the occurrence and development of preeclampsia by suppressing glycolysis and impairing decidualization. *HK2* was, however, found to be upregulated in placentas obtained from women with severe preeclampsia compared to controls.[79]

PMM2

PMM2 catalyzes the mutase reaction providing mannose 1-phosphate (Man-1-P). Man-1-P is an important intermediate for the N-linked polysaccharide synthesis pathway as a precursor to GDP-mannose and is essential for protein glycosylation. One study identified genes with differentially methylated CpG sites (between normal pregnant and preeclamptic women) and mapped these genes to gene ontology for molecular functions and biological processes.[106] GDP-mannose metabolic process and GDP-mannose 4,6-dehydratase activity were among the gene ontologies pertinent to the pathophysiology of preeclampsia and were found to be overrepresented in preeclamptic blood vessels. Deficiency of phosphomannomutase (PMM2) is the most common congenital disorder of glycosylation. PMM2 deficiency results in hypoglycosylation of various glycoproteins and other glycoconjugates, leading to congenital disorder of glycosylation type Ia (CDG-Ia).[107] In one case study, congenital disorder of glycosylation type Ia (CDG-Ia) was identified as the underlying fetal disease in mirror syndrome.[108] Mirror syndrome (fetal hydrops with subsequent edema in the pregnant woman) is considered to be a subset of preeclampsia and may progress into severe preeclampsia if left

untreated.[109] Severe preeclampsia is associated with inflammation and aberrant plasma glycosylation patterns, which promotes monocyte adhesion.[110] Further experimental investigation may help to elucidate the role of glycosylation and mannose metabolism in the pathogenesis of preeclampsia.

Conclusion

We identified genetic variants in a whole exome sequencing study of patients with severe preeclampsia. We applied strict filters for pathogenicity to select candidates for prediction of structural effects. We then mapped and modeled these pathogenic variants in eight proteins that were known to be coded for by preeclampsia-associated genes: filamin B (*FLNB*), transforming growth factor beta induced (*TGFBI/βig-h3*), vascular endothelial growth factor receptor 1/fms-like tyrosine kinase (*VEGFR1/FLT1*), semaphorin 3C (*SEMA3C*), hypoxia inducible factor 1 subunit alpha (*HIF1A*), interleukin 12B (*IL12B*), hexokinase 2 (*HK2*), and phosphomannomutase 2 (*PMM2*). These proteins are involved in a wide range of biological processes and pathways, including angiogenesis, TGFβ-, and VEGF-signaling pathways. Previous genetic studies of preeclampsia have identified genes implicated in angiogenesis, invasion, oxidative stress, and inflammation. These processes are implicated in the genes investigated in this study as well.[111] A previous case-control whole exome sequencing study for preeclampsia identified genes significantly enriched for rare predicted deleterious variants.[112] Individually, none of these genes found to be significant in the univariate analysis were included among the 27 preeclampsia-associated genes identified in our study. Out of 27 preeclampsia-related genes in our cohort, variants in *NEK11, FLNB, PMM2, APC, VEGFC, TTN, ABCA4* are genes associated with coagulopathy (blood coagulation disorder), a clinical risk factor for preeclampsia included in their disease association panel; *HK2, FLT1, UCP2, SLC22A3, KIAA1109, PNPLA2* are

associated with diabetes (diabetes mellitus); *SEMA3C, BMP6, IL12B, TGFBI, ABCC3, HIF1A, TAGLN, STOX1* are associated with hypertension; *PAPPA, EML2, ABCA12, SLC16A6, DDIT3, SH3PXD2A* are not included in the disease association panel.[112,113]

Of the genes reviewed in this study, *FLT1/VEGFR1*'s role in preeclampsia has been particularly well-studied in the literature. The variant in this gene, p.Arg1060Leu lies on the catalytic protein tyrosine kinase domain, which transduces signals for endothelial cells. This variant results in a change from a basic, polar, positively charged amino acid to a smaller aliphatic, hydrophobic amino acid. The change in physical and chemical properties may affect protein stability and interactions in this significant domain. *HIF1A*'s role in preeclampsia has also been previously investigated. It encodes hypoxia-inducible factor-1, which is involved in the hypoxia response and regulation of trophoblast differentiation. The variant from *HIF1A*, p.Glu261Lys, results in a change from an acidic to basic amino acid. *TGFBI* contained a variant in the FAS1-4 domain, p.Pro501Thr, that has previously been implicated in corneal dystrophy. The substitution of a cyclic, rigid proline to threonine may have a potential effect on inter-domain surface interactions in the cooperative conformation that influences angiogenic activity.

Other genes included in this study have been less well-studied in direct relation to preeclampsia and its pathophysiology. For example, while *FLNB* plays a role in endothelial cell migration and angiogenesis, its direct role in preeclampsia has been little explored. The variants from this patient cohort in *FLNB* neighbor a CpG island, lying within a CpG island shore region. This CpG island contains a probe that has been found to be differentially methylated between healthier and canonical preeclampsia transcriptional subtype clusters. As CpG methylation is

involved in trophoblast differentiation, further investigation may be required to see if these variants affect impaired trophoblast differentiation in preeclampsia. The variants in this gene, p.Phe1411Leu and p.Gly1262Ala, result in changes from an aromatic to aliphatic amino acid and a small hydrogen side chain to a larger side chain, respectively. *SEMA3C* encodes semaphorin 3C, which has been shown to suppress angiogenesis through inhibition of endothelial junction integrity and focal adhesion with the extracellular matrix via Nrp-1 and PlexinD1 receptors. The variant from the cohort, p.Val321Ala, is on the semaphorin domain, which is critical to protein-protein interactions between semaphorins and their receptors including neuropilins (NRPs) and plexins (PLXNs).[76] The variant in *PMM2*, p.Arg123Gln, has been implicated in PMM2-congenital disorder of glycosylation (PMM-CDG) and likely affects substrate binding. PMM2-CDG has been identified as the underlying fetal disease in mirror syndrome, a subset of preeclampsia that can develop into severe preeclampsia if left untreated. Further experimental investigation could look into the potential role of glycosylation and mannose metabolism in preeclampsia.

This computational study identified variants in genes previously demonstrated to be involved in the pathogenesis of preeclampsia. We also identified other preeclampsia-associated genes for which their exact role in the pathophysiology of preeclampsia is unclear. The variants we have identified in several of these genes provide plausible mechanistic insights into preeclampsia and identify opportunities for further experimental validation. Inclusion of genes and protein products in this study was limited by the availability of existing protein data bank structures. Molecular dynamics simulations may be helpful in assessing effects of the variant

amino acid on protein stability. Further experimental studies can help to elucidate the mechanisms and roles of these genes in preeclampsia-related pathways.

References

1. Valenzuela FJ, Perez-Sepulveda A, Torres MJ, Correa P, Repetto GM, Illanes SE. Pathogenesis of preeclampsia: the genetic component. *J Pregnancy.* 2012;2012:632732.
2. Chappell S, Morgan L. Searching for genetic clues to the causes of pre-eclampsia. *Clin Sci (Lond).* 2006;110(4):443-458.
3. Triche EW, Uzun A, DeWan AT, et al. Bioinformatic approach to the genetics of preeclampsia. *Obstet Gynecol.* 2014;123(6):1155-1161.
4. 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;335(7627):974.
5. McDonald SD, Malinowski A, Zhou Q, Yusuf S, Devereaux PJ. Cardiovascular sequelae of preeclampsia/eclampsia: a systematic review and meta-analyses. *Am Heart J.* 2008;156(5):918-930.
6. Tenhola S, Rahiala E, Halonen P, Vanninen E, Voutilainen R. Maternal preeclampsia predicts elevated blood pressure in 12-year-old children: evaluation by ambulatory blood pressure monitoring. *Pediatr Res.* 2006;59(2):320-324.
7. Kajantie E, Eriksson JG, Osmond C, Thornburg K, Barker DJ. Pre-eclampsia is associated with increased risk of stroke in the adult offspring: the Helsinki birth cohort study. *Stroke.* 2009;40(4):1176-1180.
8. Wu CS, Nohr EA, Bech BH, Vestergaard M, Catov JM, Olsen J. Health of children born to mothers who had preeclampsia: a population-based cohort study. *Am J Obstet Gynecol.* 2009;201(3):269 e261-269 e210.
9. Williams PJ, Broughton Pipkin F. The genetics of pre-eclampsia and other hypertensive disorders of pregnancy. *Best Pract Res Clin Obstet Gynaecol.* 2011;25(4):405-417.
10. Agius A, Sultana R, Camenzuli C, Calleja-Agius J, Balzan R. An update on the genetics of pre-eclampsia. *Minerva Ginecol.* 2018;70(4):465-479.
11. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47(D1):D886-D894.
12. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.
13. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 2012;40(Web Server issue):W452-457.
14. Haeussler M, Zweig AS, Tyner C, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 2019;47(D1):D853-D858.
15. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308-311.
16. Yates B, Braschi B, Gray KA, Seal RL, Tweedie S, Bruford EA. Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res.* 2017;45(D1):D619-D625.
17. Benson DA, Cavanaugh M, Clark K, et al. GenBank. *Nucleic Acids Res.* 2018;46(D1):D41-D47.
18. Hunter S, Apweiler R, Attwood TK, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res.* 2009;37(Database issue):D211-215.

19. UniProt C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019;47(D1):D506-D515.
20. Uzun A, Triche EW, Schuster J, Dewan AT, Padbury JF. dbPEC: a comprehensive literature-based database for preeclampsia related genes and phenotypes. *Database (Oxford).* 2016;2016.
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403-410.
22. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296-W303.
23. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(13):1605-1612.
24. Armanious D, Schuster J, Istrail S, Padbury J, Uzun A. Proteinarium: Multi-Sample Protein-Protein Interaction Visualization and Analysis Tool. *bioRxiv.* 2019:589085.
25. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-D613.
26. Lin H, Zeng J, Xie R, et al. Discovery of a Novel 2,6-Disubstituted Glucosamine Series of Potent and Selective Hexokinase 2 Inhibitors. *ACS Med Chem Lett.* 2016;7(3):217-222.
27. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 2019;28(11):1947-1951.
28. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27-30.
29. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 2019;47(D1):D590-D595.
30. Stossel TP, Condeelis J, Cooley L, et al. Filamins as integrators of cell mechanics and signalling. *Nat Rev Mol Cell Biol.* 2001;2(2):138-145.
31. van der Flier A, Sonnenberg A. Structural and functional aspects of filamins. *Biochim Biophys Acta.* 2001;1538(2-3):99-117.
32. Takafuta T, Saeki M, Fujimoto TT, Fujimura K, Shapiro SS. A new member of the LIM protein family binds to filamin B and localizes at stress fibers. *J Biol Chem.* 2003;278(14):12175-12181.
33. Xiao G, Cheng H, Cao H, et al. Critical role of filamin-binding LIM protein 1 (FBLP-1)/migfilin in regulation of bone remodeling. *J Biol Chem.* 2012;287(25):21450-21460.
34. Garcia-Castellanos R, Nielsen NS, Runager K, et al. Structural and Functional Implications of Human Transforming Growth Factor beta-Induced Protein, TGFBIp, in Corneal Dystrophies. *Structure.* 2017;25(11):1740-1750 e1742.
35. Yamamoto S, Okada M, Tsujikawa M, et al. A kerato-epithelin (betaig-h3) mutation in lattice corneal dystrophy type IIIA. *Am J Hum Genet.* 1998;62(3):719-722.
36. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44(D1):D733-745.
37. Han KE, Choi SI, Kim TI, et al. Pathogenesis and treatments of TGFBI corneal dystrophies. *Prog Retin Eye Res.* 2016;50:67-88.

38. Lakshminarayanan R, Chaurasia SS, Murugan E, et al. Biochemical properties and aggregation propensity of transforming growth factor-induced protein (TGFBIp) and the amyloid forming mutants. *Ocul Surf.* 2015;13(1):9-25.
39. Escribano J, Hernando N, Ghosh S, Crabb J, Coca-Prados M. cDNA from human ocular ciliary epithelium homologous to beta ig-h3 is preferentially expressed as an extracellular protein in the corneal epithelium. *J Cell Physiol.* 1994;160(3):511-521.
40. Karring H, Runager K, Valnickova Z, et al. Differential expression and processing of transforming growth factor beta induced protein (TGFBIp) in the normal human cornea during postnatal development and aging. *Exp Eye Res.* 2010;90(1):57-62.
41. Kawamoto T, Noshiro M, Shen M, et al. Structural and phylogenetic analyses of RGD-CAP/beta ig-h3, a fasciclin-like adhesion protein expressed in chick chondrocytes. *Biochim Biophys Acta.* 1998;1395(3):288-292.
42. Skonier J, Neubauer M, Madisen L, Bennett K, Plowman GD, Purchio AF. cDNA cloning and sequence analysis of beta ig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-beta. *DNA Cell Biol.* 1992;11(7):511-522.
43. Kim JE, Kim SJ, Lee BH, Park RW, Kim KS, Kim IS. Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-beta-induced gene, betaig-h3. *J Biol Chem.* 2000;275(40):30907-30915.
44. Son HN, Nam JO, Kim S, Kim IS. Multiple FAS1 domains and the RGD motif of TGFBI act cooperatively to bind alphavbeta3 integrin, leading to anti-angiogenic and anti-tumor effects. *Biochim Biophys Acta.* 2013;1833(10):2378-2388.
45. Shibuya M. Structure and dual function of vascular endothelial growth factor receptor-1 (Flt-1). *Int J Biochem Cell Biol.* 2001;33(4):409-420.
46. Moens S, Goveia J, Stapor PC, Cantelmo AR, Carmeliet P. The multifaceted activity of VEGF in angiogenesis - Implications for therapy responses. *Cytokine Growth Factor Rev.* 2014;25(4):473-482.
47. Smith GA, Fearnley GW, Harrison MA, Tomlinson DC, Wheatcroft SB, Ponnambalam S. Vascular endothelial growth factors: multitasking functionality in metabolism, health and disease. *J Inherit Metab Dis.* 2015;38(4):753-763.
48. Shibuya M. VEGF-VEGFR Signals in Health and Disease. *Biomol Ther (Seoul).* 2014;22(1):1-9.
49. Karumanchi SA, Bdolah Y. Hypoxia and sFlt-1 in preeclampsia: the "chicken-and-egg" question. *Endocrinology.* 2004;145(11):4835-4837.
50. Red-Horse K, Zhou Y, Genbacev O, et al. Trophoblast differentiation during embryo implantation and formation of the maternal-fetal interface. *J Clin Invest.* 2004;114(6):744-754.
51. Banu N, Teichman J, Dunlap-Brown M, Villegas G, Tufro A. Semaphorin 3C regulates endothelial cell function by increasing integrin activity. *FASEB J.* 2006;20(12):2150-2152.
52. Ellis LM. The role of neuropilins in cancer. *Mol Cancer Ther.* 2006;5(5):1099-1107.
53. Antipenko A, Himanen JP, van Leyen K, et al. Structure of the semaphorin-3A receptor binding module. *Neuron.* 2003;39(4):589-598.
54. Lu S, Wang J, Chitsaz F, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 2020;48(D1):D265-D268.

55. Semenza GL. Regulation of mammalian O2 homeostasis by hypoxia-inducible factor 1. *Annu Rev Cell Dev Biol.* 1999;15:551-578.
56. Semenza GL. Hypoxia-inducible factors in physiology and medicine. *Cell.* 2012;148(3):399-408.
57. Wang GL, Jiang BH, Rue EA, Semenza GL. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proc Natl Acad Sci U S A.* 1995;92(12):5510-5514.
58. Jiang BH, Rue E, Wang GL, Roe R, Semenza GL. Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1. *J Biol Chem.* 1996;271(30):17771-17778.
59. Peng J, Zhang L, Drysdale L, Fong GH. The transcription factor EPAS-1/hypoxia-inducible factor 2alpha plays an important role in vascular remodeling. *Proc Natl Acad Sci U S A.* 2000;97(15):8386-8391.
60. Keith B, Johnson RS, Simon MC. HIF1alpha and HIF2alpha: sibling rivalry in hypoxic tumour growth and progression. *Nat Rev Cancer.* 2011;12(1):9-22.
61. Albers RE, Kaufman MR, Natale BV, et al. Trophoblast-Specific Expression of Hif-1alpha Results in Preeclampsia-Like Symptoms and Fetal Growth Restriction. *Sci Rep.* 2019;9(1):2742.
62. Wu D, Potluri N, Lu J, Kim Y, Rastinejad F. Structural integration in hypoxia-inducible factors. *Nature.* 2015;524(7565):303-308.
63. Gately MK, Renzetti LM, Magram J, et al. The interleukin-12/interleukin-12-receptor system: role in normal and pathologic immune responses. *Annu Rev Immunol.* 1998;16:495-521.
64. Oppmann B, Lesley R, Blom B, et al. Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. *Immunity.* 2000;13(5):715-725.
65. Sharma A, Satyam A, Sharma JB. Leptin, IL-10 and inflammatory markers (TNF-alpha, IL-6 and IL-8) in pre-eclamptic, normotensive pregnant and healthy non-pregnant women. *Am J Reprod Immunol.* 2007;58(1):21-30.
66. Zhang G, Mazurkie AS, Dunaway-Mariano D, Allen KN. Kinetic evidence for a substrate-induced fit in phosphonoacetaldehyde hydrolase catalysis. *Biochemistry.* 2002;41(45):13370-13377.
67. Ji T, Zhang C, Zheng L, Dunaway-Mariano D, Allen KN. Structural Basis of the Molecular Switch between Phosphatase and Mutase Functions of Human Phosphomannomutase 1 under Ischemic Conditions. *Biochemistry.* 2018;57(25):3480-3492.
68. Matthijs G, Schollen E, Pardon E, et al. Mutations in PMM2, a phosphomannomutase gene on chromosome 16p13, in carbohydrate-deficient glycoprotein type I syndrome (Jaeken syndrome). *Nat Genet.* 1997;16(1):88-92.
69. Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem.* 1985;54:631-664.
70. Dai J, Finci L, Zhang C, et al. Analysis of the structural determinants underlying discrimination between substrate and solvent in beta-phosphoglucomutase catalysis. *Biochemistry.* 2009;48(9):1984-1995.

71. Lu Z, Dunaway-Mariano D, Allen KN. HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. *Biochemistry.* 2005;44(24):8684-8696.
72. Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN. Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from Lactococcus lactis. *Biochemistry.* 2002;41(26):8351-8359.
73. Barnes MR, Gray IC. *Bioinformatics for geneticists.* Chichester, West Sussex, England ; Hoboken, N.J.: Wiley; 2003.
74. Kim S, Chen J, Cheng T, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 2019;47(D1):D1102-D1109.
75. Laskowski RA, Jablonska J, Pravda L, Varekova RS, Thornton JM. PDBsum: Structural summaries of PDB entries. *Protein Sci.* 2018;27(1):129-134.
76. Lee CCW, Munuganti RSN, Peacock JW, et al. Targeting Semaphorin 3C in Prostate Cancer With Small Molecules. *J Endocr Soc.* 2018;2(12):1381-1394.
77. Jiang Q, Arnold S, Heanue T, et al. Functional loss of semaphorin 3C and/or semaphorin 3D and their epistatic interaction with ret are critical to Hirschsprung disease liability. *Am J Hum Genet.* 2015;96(4):581-596.
78. Silvaggi NR, Zhang C, Lu Z, Dai J, Dunaway-Mariano D, Allen KN. The X-ray crystal structures of human alpha-phosphomannomutase 1 reveal the structural basis of congenital disorder of glycosylation type 1a. *J Biol Chem.* 2006;281(21):14918-14926.
79. Sitras V, Paulssen RH, Gronaas H, et al. Differential placental gene expression in severe preeclampsia. *Placenta.* 2009;30(5):424-433.
80. Del Valle-Perez B, Martinez VG, Lacasa-Salavert C, et al. Filamin B plays a key role in vascular endothelial growth factor-induced endothelial cell motility through its interaction with Rac-1 and Vav-2. *J Biol Chem.* 2010;285(14):10748-10760.
81. Maynard SE, Min JY, Merchan J, et al. Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. *J Clin Invest.* 2003;111(5):649-658.
82. Venkatesha S, Toporsian M, Lam C, et al. Soluble endoglin contributes to the pathogenesis of preeclampsia. *Nat Med.* 2006;12(6):642-649.
83. Nelissen EC, van Montfoort AP, Dumoulin JC, Evers JL. Epigenetics and the placenta. *Hum Reprod Update.* 2011;17(3):397-417.
84. Robinson WP, Price EM. The human placental methylome. *Cold Spring Harb Perspect Med.* 2015;5(5):a023044.
85. Bianco-Miotto T, Mayne BT, Buckberry S, Breen J, Rodriguez Lopez CM, Roberts CT. Recent progress towards understanding the role of DNA methylation in human placental development. *Reproduction.* 2016;152(1):R23-30.
86. Huppertz B. Trophoblast differentiation, fetal growth restriction and preeclampsia. *Pregnancy Hypertens.* 2011;1(1):79-86.
87. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev.* 2011;25(10):1010-1022.
88. Gamage T, Schierding W, Hurley D, et al. The role of DNA methylation in human trophoblast differentiation. *Epigenetics.* 2018;13(12):1154-1173.
89. Blair JD, Yuen RK, Lim BK, McFadden DE, von Dadelszen P, Robinson WP. Widespread DNA hypomethylation at gene enhancer regions in placentas associated with early-onset pre-eclampsia. *Mol Hum Reprod.* 2013;19(10):697-708.

90. Leavey K, Wilson SL, Bainbridge SA, Robinson WP, Cox BJ. Epigenetic regulation of placental gene expression in transcriptional subtypes of preeclampsia. *Clin Epigenetics.* 2018;10:28.
91. Wang Y, Zhao S, Loyd S, Groome LJ. Increased urinary excretion of nephrin, podocalyxin, and betaig-h3 in women with preeclampsia. *Am J Physiol Renal Physiol.* 2012;302(9):F1084-1089.
92. SA K, S R, RN T. Angiogenesis and preeclampsia. In: *Chesley's Hypertensive Disorders in Pregnancy.* 4 ed. Oxford, UK: Academic Press; 2014:113-132.
93. Yang WJ, Hu J, Uemura A, Tetzlaff F, Augustin HG, Fischer A. Semaphorin-3C signals through Neuropilin-1 and PlexinD1 receptors to inhibit pathological angiogenesis. *EMBO Mol Med.* 2015;7(10):1267-1284.
94. Fuh G, Garcia KC, de Vos AM. The interaction of neuropilin-1 with vascular endothelial growth factor and its receptor flt-1. *J Biol Chem.* 2000;275(35):26690-26695.
95. Tal R. The role of hypoxia and hypoxia-inducible factor-1alpha in preeclampsia pathogenesis. *Biol Reprod.* 2012;87(6):134.
96. Cowden Dahl KD, Fryer BH, Mack FA, et al. Hypoxia-inducible factors 1alpha and 2alpha regulate trophoblast differentiation. *Mol Cell Biol.* 2005;25(23):10479-10491.
97. Daniel Y, Kupferminc MJ, Baram A, et al. Plasma interleukin-12 is elevated in patients with preeclampsia. *Am J Reprod Immunol.* 1998;39(6):376-380.
98. Ching T, Ha J, Song MA, et al. Genome-scale hypomethylation in the cord blood DNAs associated with early onset preeclampsia. *Clin Epigenetics.* 2015;7:21.
99. Cemgil Arikan D, Aral M, Coskun A, Ozer A. Plasma IL-4, IL-8, IL-12, interferon-gamma and CRP levels in pregnant women with preeclampsia, and their relation with severity of disease and fetal birth weight. *J Matern Fetal Neonatal Med.* 2012;25(9):1569-1573.
100. Bachmayer N, Rafik Hamad R, Liszka L, Bremme K, Sverremark-Ekstrom E. Aberrant uterine natural killer (NK)-cell expression and altered placental and serum levels of the NK-cell promoting cytokine interleukin-12 in pre-eclampsia. *Am J Reprod Immunol.* 2006;56(5-6):292-301.
101. Lv H, Tong J, Yang J, et al. Dysregulated Pseudogene HK2P1 May Contribute to Preeclampsia as a Competing Endogenous RNA for Hexokinase 2 by Impairing Decidualization. *Hypertension.* 2018;71(4):648-658.
102. Hannon T, Innes BA, Lash GE, Bulmer JN, Robson SC. Effects of local decidua on trophoblast invasion and spiral artery remodeling in focal placenta creta - an immunohistochemical study. *Placenta.* 2012;33(12):998-1004.
103. Nancy P, Tagliani E, Tay CS, Asp P, Levy DE, Erlebacher A. Chemokine gene silencing in decidual stromal cells limits T cell access to the maternal-fetal interface. *Science.* 2012;336(6086):1317-1321.
104. Zhang S, Lin H, Kong S, et al. Physiological and molecular determinants of embryo implantation. *Mol Aspects Med.* 2013;34(5):939-980.
105. Cohen M, Wuillemin C, Irion O, Bischof P. Role of decidua in trophoblastic invasion. *Neuro Endocrinol Lett.* 2010;31(2):193-197.
106. Mousa AA, Archer KJ, Cappello R, et al. DNA methylation is altered in maternal blood vessels of women with preeclampsia. *Reprod Sci.* 2012;19(12):1332-1342.
107. Marquardt T, Denecke J. Congenital disorders of glycosylation: review of their molecular bases, clinical presentations and specific therapies. *Eur J Pediatr.* 2003;162(6):359-379.

108. Wurm D, Loffler G, Lindinger A, Gortner L. Congenital disorders of glycosylation type Ia as a cause of mirror syndrome. *J Perinatol.* 2007;27(12):802-804.
109. Vidaeff AC, Pschirrer ER, Mastrobattista JM, Gilstrap LC, 3rd, Ramin SM. Mirror syndrome. A case report. *J Reprod Med.* 2002;47(9):770-774.
110. Flood-Nichols SK, Kazanjian AA, Tinnemore D, et al. Aberrant glycosylation of plasma proteins in severe preeclampsia promotes monocyte adhesion. *Reprod Sci.* 2014;21(2):204-214.
111. Ji L, Brkic J, Liu M, Fu G, Peng C, Wang YL. Placental trophoblast cell differentiation: physiological regulation and pathological relevance to preeclampsia. *Mol Aspects Med.* 2013;34(5):981-1023.
112. Hansen AT, Bernth Jensen JM, Hvas AM, Christiansen M. The genetic component of preeclampsia: A whole-exome sequencing study. *PLoS One.* 2018;13(5):e0197217.
113. Bartsch E, Medcalf KE, Park AL, Ray JG, High Risk of Pre-eclampsia Identification G. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. *BMJ.* 2016;353:i1753.