Genetic predisposition to preeclampsia is conferred by fetal DNA variants near *FLT1*, a gene involved in the regulation of angiogenesis

Kathryn J. Gray, MD, PhD; Richa Saxena, PhD; S. Ananth Karumanchi, MD

Preeclampsia risk is influenced by both the mother's genetic background and the genetics of her fetus; however, the specific genes responsible for conferring preeclampsia risk have largely remained elusive. Evidence that preeclampsia has a genetic predisposition was first detailed in the early 1960s, and overall preeclampsia heritability is estimated at ~55%. Many traditional gene discovery approaches have been used to investigate the specific genes that contribute to preeclampsia risk, but these have largely not been successful or reproducible. Over the past decade, genome-wide association studies have allowed for significant advances in the understanding of the genetic basis of many common diseases. Genome-wide association studies are predicated on the idea that the genetic basis of many common diseases are complex and polygenic with many variants, each with modest effects that contribute to disease risk. Using this approach in preeclampsia, a large genome-wide association study recently identified and replicated the first robust fetal genomic region associated with excess risk. A screen of >7 million genetic variants in 2658 offspring from preeclamptic women and 308,292 population controls identified a single association signal close to the *Fms-like tyrosine kinase 1* gene, on chromosome 13. *Fms-like tyrosine kinase 1* encodes soluble Fms-like tyrosine kinase 1, a splice variant of the vascular endothelial growth factor receptor that exerts antiangiogenic activity by inhibiting signaling of proangiogenic factors. The *Fms-like tyrosine kinase 1* pathway is central in preeclampsia pathogenesis because excess circulating soluble Fms-like tyrosine kinase 1 in the maternal plasma leads to the hallmark clinical features of preeclampsia, including hypertension and proteinuria. The success of this landmark fetal preeclampsia genome-wide association study suggests that well-powered, larger maternal and fetal genome-wide association study will be fruitful in identifying additional common variants that implicate causal preeclampsia genes and pathways. Such efforts will rely on the continued development of large preeclampsia consortia focused on preeclampsia genetics to obtain adequate sample sizes, detailed clinical phenotyping, and matched maternal-fetal samples. In summary, the fetal preeclampsia genome-wide association study represents an exciting advance in preeclampsia biology, suggesting that dysregulation at the *Fms-like tyrosine kinase 1* locus in the fetal genome (likely in the placenta) is a fundamental molecular defect in preeclampsia.

Key words: angiogenic imbalance, fetal genetics, Fms-like tyrosine kinase 1, genome-wide association study, heritability, hypertension, preeclampsia, vascular endothelial growth factor

Preeclampsia is observed in approximately 3–5% of pregnancies worldwide. A mother's predisposition to preeclampsia is multifactorial and influenced by a host of factors including her age, race, parity, coexistent medical conditions, and pregnancy and partner characteristics, among others. Interestingly, both a mother's genetic background and the genetics of her fetus, impact risk.

To date, the specific genes responsible for conferring preeclampsia risk have largely remained elusive, until recently when the large genome-wide association study (GWAS) titled, Variants in the fetal genome near FLT1 locus are associated with risk of preeclampsia, by McGinnis et al[1] identified and replicated the first robust fetal genomic region associated with excess risk. Here we detail the history of genetic studies of preeclampsia, starting from the studies that initially suggested preeclampsia is heritable to the current findings, and describe the implications of this landmark fetal preeclampsia GWAS.

From the Brigham and Women's Hospital, Harvard Medical School (Dr Gray), the Broad Institute of Harvard and Massachusetts Institute of Technology and Massachusetts General Hospital (Dr Saxena), Harvard Medical School, and Beth Israel Deaconess Medical Center, Harvard Medical School (Dr Karumanchi), Boston, MA.

Received Oct. 4, 2017; revised Nov. 6, 2017; accepted Nov. 7, 2017.

Dr Karumanchi is a coinventor on patents related to preeclampsia biomarkers that are held at Beth Israel Deaconess Medical Center. He has financial interest in Aggamin LLC and reports having served as a consultant to Roche Diagnostics and Thermofisher Scientific. The other authors have no conflict of interest.

Corresponding author: S. Ananth Karumanchi, MD. sananth@bidmc.harvard.edu

0002-9378/$36.00

https://doi.org/10.1016/j.ajog.2017.11.562

Maternal and fetal genomes influence the risk of preeclampsia

Evidence that preeclampsia has a genetic predisposition was first detailed in the early 1960s by Chesley et al,[2] a group of researchers who were working at the Margaret Hague Maternity Hospital (Jersey City, NJ), who observed familial aggregation of cases of preeclampsia and eclampsia. By reviewing hospital records of women with eclampsia who had delivered at the hospital, they found an increased rate of preeclampsia and eclampsia in pregnancies of the sisters, daughters, and granddaughters of these women compared with the daughters-in-law of the same women.[3]

Preeclampsia heritability from mothers to daughters was then replicated in a much larger group of women in the Swedish Birth Registry[4,5] and extended to delineate the importance of fetal genetics, as well as maternal genetics, in disease heritability.[6-8] Based on this work, preeclampsia heritability is estimated at ~55%, with both maternal and fetal contributions to risk (estimated at 35% and 20%, respectively).[4-8]

Traditional gene discovery approaches to identify preeclampsia genes

To determine the specific genetic markers (ie, DNA variations) that confer preeclampsia risk, many traditional gene discovery approaches have been used including linkage studies and candidate gene association studies. In linkage studies, genetic markers of disease (ie, specific chromosomal regions) are studied in related individuals to determine which chromosomal regions segregate with disease, and positional cloning is then used to identify the specific gene or genetic variant of interest.[9]

In candidate gene association studies, the DNA sequence of biologically plausible genes is determined, and allele frequencies (ie, the specific base at a given position within the gene) are compared between cases and controls.[10] For preeclampsia, linkage approaches have implicated at least 8 different chromosomal regions,[11-14] and further study of these regions have identified variants within the genes, *ACVR2A*[15] and *ERAP2*,[16] that may predispose to preeclampsia.

Preeclampsia candidate gene approaches have largely focused on variants within the following: (1) the renin-angiotensin system; (2) coagulation factors; (3) oxidative stress pathways; (4) dyslipidemia; and (5) immunoregulatory components, in particular within the HLA region. Despite extensive work, these approaches have largely not been successful or reproducible.[17] As for many complex diseases, traditional gene discovery in preeclampsia has suffered from errors of inadequate sample size, inaccurate clinical phenotyping, lack of correction for multiple testing, poorly matched control groups, hidden ethnic bias, failure to replicate in an independent population, positive publication bias, and random error.[18]

Genome-wide association studies for unbiased gene discovery

Over the past decade, GWASs have allowed for dramatic advances in the understanding of the genetic basis of common disease, including type 2 diabetes, obesity, hypertension, autoimmune disease, and psychiatric traits.[19] GWAS is predicated on the idea that the genetic basis of many common diseases are complex and polygenic with many variants, each with modest effects, that contribute to disease risk.

The unit of genetic variation assessed is the single-nucleotide polymorphism (SNP), which is a single-base pair change in the DNA sequence. SNPs are common in the human genome and are used in GWASs as markers of genomic regions. Most SNPs have minimal biological effects themselves. In a GWAS, genome-wide SNP data are used to look for SNPs associated with a given disease or trait.

The ability of any given GWAS to succeed depends on the sample size, the panel of genome-wide variants genotyped, the genetic architecture (ie, the effect size and allele frequency of risk variants), how genetic variants contributing to risk segregate in the population, and the heterogeneity of the disease.[19] For example, a heterogeneous disease with many genetic risk loci, each with modest effects on disease, will require a GWAS with a much larger sample size to discover variants with genome-wide significant effects ($P < 5 \times 10^{-8}$, accounting for multiple testing) than a more homogeneous disease with a few genetic risk loci of large effect.

Following the initial results of a GWAS for any given trait, several additional steps are necessary. Given the possibility of false discovery when surveying the whole genome, genome-wide significant SNPs must be replicated in an independent population. Because SNPs are only markers of a given genomic region, functional follow-up of replicated SNPs is required to determine the biological relevance of each genetic locus to disease. Because GWASs test genetic variation across the genome, they are, by nature, unbiased, with the potential to discover novel underlying disease biology. In many cases, GWASs have uncovered genes distinct from those previously chosen in hypothesis-driven candidate gene studies, highlighting the importance of the unbiased approach.

Despite the success of a GWAS for understanding many complex diseases, GWASs of obstetric traits have lagged behind. Several challenges in pregnancy-related genetic research (including heterogeneous phenotypes, lack of large cohorts with detailed pregnancy information, exclusion of pregnant women from research studies, and the involvement of at least 2 genetically distinct individuals (mother and fetus[es]) in each obstetric outcome have made it difficult to obtain the sample sizes needed for well-powered GWASs of obstetric traits. In preeclampsia, 3 recent small-scale maternal preeclampsia GWASs have been reported,[20-22] but the DNA variants identified have not been replicated.

Discovery of the first robust preeclampsia genetic association: Fms-like tyrosine kinase 1 (FLT1) fetal DNA variants

The recently published fetal GWAS, Variants in the fetal genome near *FLT1* are associated with risk of preeclampsia, by McGinnis et al[1] is the first preeclampsia GWAS to identify a genetic risk variant with genome-wide significance with convincing replication in an independent cohort. This GWAS finding provides compelling evidence that alterations near the *FLT1* locus in the human fetal genome are causal in the development of preeclampsia. It is striking that this first well-powered unbiased fetal GWAS homes in on the *FLT1* genomic region, given the body of literature devoted to the role of the FLT1 pathway in preeclampsia pathogenesis (see following text).

In the discovery phase of this fetal preeclampsia GWAS, >7 million genetic variants were assessed in 2658 offspring from preeclamptic women and 308,292 population controls of European descent (from Iceland and the United Kingdom).