Insights into genetic susceptibility in the etiology of spontaneous preterm birth

Sasha E Parets¹
Anna K Knight²
Alicia K Smith¹,²

¹Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA;
²Genetics and Molecular Biology Program, Emory University, Atlanta, GA, USA

Abstract: Preterm birth (PTB; <37 weeks of gestation) is a complex disorder, whose etiology is influenced by a variety of factors. A greater understanding of the biological mechanisms that contribute to PTB will facilitate identification of those at increased risk and may inform new treatments. To accomplish this, it is vital to elucidate the heritability patterns of this condition as well as the environment and lifestyle factors that increase risk for PTB. Identifying individual genes that contribute to the etiology of PTB presents particular challenges, and there has been little agreement among candidate gene and genome-wide studies performed to date. In this review we will evaluate recent genetic studies of spontaneous PTB, discuss common themes among their findings, and suggest approaches for future studies of PTB.

Keywords: PTB, GWAS, linkage, candidate gene, African–American, epigenetic

Introduction

Despite ongoing research and the advancement of health-care systems in the US, the rate of preterm birth (PTB) remains high. Prevention strategies have not been successful at reducing the PTB rate, and interventions for PTB after initiation of labor are only intermittently successful. Risk factors that increase the likelihood of a woman delivering preterm include her age, race, and current smoking status. Additionally, women who have had a previous preterm delivery are at higher risk of subsequent preterm deliveries.¹–³

Over the past several years, clinical advancements have primarily occurred in managing the acute clinical needs of neonates who are born preterm.⁴ Despite this progress, PTB is the primary cause of mortality in the 1st year,⁵ and the morbidities of being born preterm can last throughout life as children born preterm are more likely to develop chronic diseases.²,⁶ The dire consequences of this condition cannot be fully addressed until the mechanisms that contribute to PTB are identified. This review will summarize the recent progress of genetic studies of PTB and discuss the insight they provide into its etiology.

PTB heritability

There are three primary categories of PTB. Approximately 30%–35% of all PTB are considered medically indicated and result from specific condition such as preeclampsia or fetal growth restriction affecting either the mother or fetus.² Preterm premature rupture of the membranes (pPROMs) occurs in 25%–30% of PTB cases and is likely to occur in cases of infection, placental abruption, or if there are any anatomical abnormalities in the mother.² Spontaneous PTB accounts for the remaining 35%–45%...
of PTB cases and has no known etiology.\textsuperscript{7} Genetic studies may be particularly useful for identifying biological pathways that contribute to spontaneous PTB.

Risk for PTB is not evenly distributed in the population. African–Americans are 1.5 times more likely to deliver preterm and more than twice as likely to deliver before 32 weeks gestation compared to Caucasians.\textsuperscript{2,8} A review by Anum et al\textsuperscript{10} highlighted other factors that could contribute to the increased risk of PTB among African–Americans including anemia during pregnancy, low serum folate levels, vitamin D deficiency, poor weight gain during pregnancy, or high prepregnancy body mass index (BMI). However, it is important to note that this increased risk of PTB among African–Americans appears to be independent of socioeconomic status or other social factors,\textsuperscript{11,12} as evidenced by the disparity remaining in military settings that have the same access to healthcare.\textsuperscript{10}

One of the greatest risk factors for PTB is personal or familial history;\textsuperscript{2} and heritability studies have attempted to quantify how much of the variation in spontaneous PTB can be attributed to genetic differences. The estimates of PTB heritability in different populations vary substantially (17%–36%) depending on the types of PTB included and the nature of the cohort studied.\textsuperscript{13–16} For example, a study in a cohort of Australian twins estimated that the genetic contribution to PTB is 27% in any pregnancy and 17% for the first pregnancy.\textsuperscript{15} Another study specifically estimated the heritability of spontaneous PTB as high as 34%.\textsuperscript{17}

A recent heritability study for PTB evaluated 1,784,172 noniatrogenic births from the Utah Population Database, of which 7.6% were spontaneous PTB.\textsuperscript{18} Overall, this comprehensive analysis conducted by Wu et al\textsuperscript{18} suggested that heritability of spontaneous PTB was less than 25%, which is relatively low when compared to other complex disorders.\textsuperscript{18} This study also emphasizes the profound role the environment plays in mediating the risk for PTB, as 75% of the variance could be attributed to the environment, overlapping and independent contributions to PTB.\textsuperscript{18}

York et al\textsuperscript{19} questioned the conclusion that extended pedigrees would be informative for mapping genes related to PTB. In their extensive commentary, they highlight their work and the work of others using twin and family data to discuss methodological factors that should be considered by heritability studies of gestational age (GA) and PTB.\textsuperscript{19,20} Interestingly, these studies, which evaluated the heritability of GA in population-based records of primarily term deliveries, also support the conclusion that the environment plays the primary role in PTB risk while the genetic components contribute only modestly.\textsuperscript{21,22}

**Contributions of maternal and fetal genomes**

Pregnancy involves a semi-allograftic relationship between a mother and her fetus.\textsuperscript{21} The fetus is genetically distinct and recognized as such by the maternal immune system. Maternal physiology is adapted to tolerate and nurture the development of the fetus.\textsuperscript{24,25} Both the mother and fetus take an active role in the parturition process.\textsuperscript{8} However, the contribution of the mother and fetus toward the initiation of labor, or more generally toward PTB, remains an active debate.

Although there are undoubtedly multiple mechanisms that may contribute to the etiology of PTB, some argue that genetic studies of PTB should consider both the maternal and fetal genomes, recognizing that they may have both overlapping and independent contributions to PTB. Others argue that primary genetic contributions to PTB are inherent to the maternal genome.\textsuperscript{26,27} One such study arguing for the maternal genome’s contribution to PTB demonstrated that there was a recurrence risk of 1.85 for offspring of mothers who were born early preterm to be born preterm themselves, while there was a recurrence risk of 1.06 for offspring of fathers who were born early preterm to be born preterm themselves. Thus, they argue that risk is likely to be transmitted primarily through the maternal genome, with a minimal contribution from the fetal genome.\textsuperscript{26} Additionally, Plunkett et al\textsuperscript{27} found that a parent-of-origin model was the most likely mode of inheritance of PTB and that the maternal contribution to PTB is greater than that of the paternal contribution. However, they note that a maternal-only inheritance model did not outperform the model accounting for the contributions of both parents.\textsuperscript{27} This is consistent with the observations of Wu et al,\textsuperscript{18} who found that the heritability of GA estimated by comparing GA of fathers and their offspring was less than half the estimate made by comparing GA of mothers and their offspring.\textsuperscript{18} Furthermore, when examining the heritability of GA, population-based studies
also report that the maternal contribution is greater than the fetal contribution.21,22

The maternal and fetal genomes share approximately 50% sequence similarity, and it can be difficult to distinguish between the effects of maternal genes and the effects of maternally-inherited genes. This is further complicated by the potential contributions of the maternal genome in regulating the intrauterine environment in which the fetus develops and with which the fetal genome interacts. The biases inherent of different heritability study designs are summarized in other papers.19,20 Overall, it is likely that individual heritability estimates vary widely in different contexts and among different cohorts, suggesting that identification of genes that increase risk for PTB may be more productive from a public health standpoint than further refinement of heritability estimates in different cohorts. However, it remains important to consider which genome is being evaluated when trying to draw conclusions across published genetic studies of PTB.

Candidate gene studies
Candidate gene studies, which compare the frequency of a predetermined genetic variant in unrelated spontaneous PTB cases and term birth controls, were among the earliest genetic studies of PTB. Candidate genes are selected to test a specific hypothesis related to the biological pathways believed to contribute to a phenotype of interest, such as PTB, but are limited to the current understanding of the biological processes underlying that phenotype.28,29 Candidate gene studies for PTB have primarily focused on genes associated with inflammation or the transition of the uterus into a contractile state.30–39

Over time, candidate gene selection has become more comprehensive and has begun incorporating different hypotheses of PTB etiology. Romero et al40 examined 775 single nucleotide polymorphisms (SNPs) in 190 candidate genes to identify variants in maternal and fetal samples that associate with preterm delivery and PTB risk. The genes and variants were selected from a literature review of genes previously evaluated for association with spontaneous PTB, preterm labor, pPROM, small for GA, or preeclampsia. In maternal samples, they identified an association between preterm delivery and a SNP in interleukin-6 receptor (IL6R; rs8192282), which binds IL6 and is associated with immune signaling.44 Genetic variation in IL6R has been associated with preterm labor.40 A recent study found that rs8192282 is part of a haplotype block that may influence soluble IL6R levels to modulate inflammatory disease risk.45 Finally, Romero et al46 identified haplotypes in four genes (IL6R; collagen, type IV, alpha 3 – COL4A3; lactotransferrin – LTF; and fibroblast growth factor 1 – FGFI) associated with delivering preterm in maternal samples and haplotypes in three genes (insulin-like growth factor 2 – IGF2, interleukin-2 – IL2, collagen, type IV, alpha 1 – COL4A1) associated with PTB in fetal samples. This study concludes that both the fetal and maternal genomes may contribute to PTB risk.40

A systemic review of the literature identified 189 SNPs in 84 genes that had been associated with PTB.46 A meta-analysis of these studies supported significant, albeit modest, associations between PTB and maternal SNPs in interleukin-1 receptor antagonist (IL1RN), beta-2 adrenergic receptor (ADRB2), interferon gamma (IFNG), and factor 2 (F2). Of these genes, IL1RN and IFNG are involved in immune response, ADRB2 is involved in G protein-coupled receptor, and F2 is involved in coagulation. In the fetal genome, the meta-analysis suggested a single SNP (rs1799963) in F2 also associated with PTB.46

Overall, candidate gene association studies in unrelated cases and controls have neither resulted in overwhelming reproducibility in maternal or fetal samples, nor have they provided substantial insight into the etiology of spontaneous PTB. Thus, investigators have begun focusing on genetic approaches that do not require an a priori understanding of the biological mechanisms underlying spontaneous PTB.

Genome-wide studies
Genome-wide association studies (GWASs) are valuable for examining the associations between genetic variants and diseases with complex etiologies, such as PTB.47 GWASs are unbiased and allow for generation of new hypotheses. However, they require large sample sizes to garner the statistical power necessary to account for the number of independent tests performed.

One such GWAS examined paired maternal–infant samples from early PTB in 959 mothers and 979 infants compared to term births in 960 mothers and 985 infants.48 No SNP in the analysis of the maternal genome remained associated with PTB after correction for multiple testing. However, SNPs from the fetal genome, located on chromosome 6
in ribonuclease T2 (RNASET2) and in the extended major histocompatibility complex, were associated with PTB at experiment-wide significance levels. Both genes are involved in regulation of the immune response, although their connection with PTB remains unclear. Despite this initial finding, the investigators were unable to replicate the association in an independent fetal cohort of 243 cases and 149 controls.

Uzun et al nested a candidate gene study inside a GWAS. Using maternal GWAS data from the GENEVA (Gene Environment Association Studies Initiative) cohort, they first examined candidate genes that were previously associated with PTB. They found no association between PTB and genes from the literature after correction for multiple tests. They next evaluated all of the SNPs from the GWAS and identified one that remained associated after multiple test correction. However, that SNP was not located near any gene and has no known regulatory function. Using a curated SNP and gene list (N=515 genes), they then preformed gene set enrichment analysis for biological process, molecular function, and cellular components and reported gene enrichment for 30 pathways, compared to 39 pathways identified using the GWAS data. Many of these pathways were associated with inflammation and metabolic disorders, and the authors suggest that these pathways may provide clues as to the gene–gene interactions that may contribute to PTB.

Another GWAS evaluated 1,000 PTBs and 1,000 term births from the Danish National Birth Cohort and found no association with any autosomal SNP in the fetal genome. The authors suggest that family-based linkage studies may be more informative of genetic variants associated with PTB risk. Overall, GWAS has had limited success in identifying potentially causal genetic variants associated with PTB, potentially due to study limitations including inadequate power, variations in sample collection and analyses methods, and the heterogeneous mechanism and definitions of PTB. It is not clear how best to improve the future likelihood of success in identifying PTB-associated variants. Some suggest that larger studies with careful phenotypic characterization will be required to identify genes of subtle effect sizes. However, it is also possible that studies of unrelated cases and controls may not be the ideal design for gene identification and that family-based approaches would be more fruitful. Indeed, the inconsistencies among findings from case–control studies performed to date suggest that other methods may prove more fruitful in identifying the genetic contribution to PTB.

Admixture mapping uses the higher prevalence of PTB among those of African descent to identify regions of potential disease association that have higher frequencies of African alleles compared to European alleles. On average, individuals of African descent have less linkage disequilibrium and shorter haplotype blocks compared to individuals of European descent. A recent study used admixture to map a region on chromosome 7 (7q21–7q22) that may be associated with PTB in African–Americans. This region contains genes involved in metabolism, inflammation, calcium regulation, and collagen, all of which may be relevant to PTB. Admixture mapping in families may provide even more power to detect susceptibility loci.

**Family-based studies**

In contrast to GWASs, linkage studies use families to attempt to identify genomic regions that segregate with a trait or disorder of interests. Linkage refers to the tendency of two genomic loci to be inherited together, at a greater a frequency than random chance. A recent linkage study by Haataja et al evaluated 120 women from 20 large families with at least two recurrent spontaneous preterm deliveries. Seven large families were selected for linkage analysis from northern Finland, a region characterized by relative genetic homogeneity that facilitates linkage mapping. This analysis suggested that a region of 15q26.3 was linked to PTB in the infants of these families. No region was independently linked to PTB in the maternal samples, but the authors noted some evidence of linkage to 15q26.3, potentially supporting the linkage signal in the infants. This region includes the gene that encodes the insulin-like growth factor receptor 1, which the authors hypothesize may be involved in the signaling pathways related to parturition.

Another linkage study performed by the same group examined the X chromosome in relation to spontaneous PTB in the same Finnish cohort. In fetal samples, Karjalainen et al reported linkage between PTB and Xq13.1, a region containing genes that encode the androgen receptor (AR) and interleukin-2 receptor gamma subunit (IL2RG). In a case–control analysis in an independent population they examined the association between spontaneous PTB and SNPs in AR and IL2RG; however, no SNP or haplotype associated with PTB status. However, upon examination of the CAG repeat of AR in fetal samples, they discovered that longer CAG repeats (≥26) associate with a greater risk of spontaneous PTB when compared to those with the shortest repeats (≤19). The authors suggest that repeat length variation may associate with length of gestation and pregnancy complications.
by altering the regulation of AR, which plays vital roles in pregnancy and development. However, an association between AR repeat length and PTB was not observed in maternal samples.56

**Epigenetics**

Since genetic studies have not yet identified genetic variants that increase risk of PTB or uncovered genes that inform its etiology, epigenetics has been proposed as a possible mechanism that incorporates environmental risk as well as the intergenerational risk.57 Epigenetics refers to the regulation of gene expression without changes in the underlying sequence of DNA. One such regulatory mark is DNA methylation at the 5′ position of cytosine. Recent studies have demonstrated that methylation may depend, in part, on the underlying genotype as methylation patterns from different tissue types are correlated within the same individual, and monozygotic twins show a higher rate of correlation than dizygotic twins.58,59 These areas of genetic–epigenetic correlation may underlie gene–environment interactions and have been associated with numerous complex traits and disorders.60–63 Thus, epigenetic regulation cannot be fully separated from the genomic sequence, and epigenetic differences have been associated with early life events, such as PTB and maternal stress, thus strengthening the hypothesis that changes in the epigenome may associate with, or even play a causal role in PTB.64–67

Epigenetic studies in murine models provide key insight into the intergenerational transmission of PTB. A recent study has examined the effects of maternal and grandmaternal stress on GA by exposing pregnant mice to a stressor for up to three generations. This study concluded that stress in previous generations may associate with shorter gestation, potentially through altered miRNA expression.68 This finding, if translated to human populations, would have enormous implications on how individual behaviors may contribute to the risk of PTB for subsequent generations.

A study by Paret et al69 evaluated the correlation of maternal and fetal DNA methylation patterns in African–Americans from a spontaneous PTB cohort. In this study, 5,171 CpG sites correlated between maternal and fetal samples, although most of the correlated CpG sites could be attributed to genetic variation. Furthermore, methylation of CpG sites in 57 genes was unique to PTB pairs.69 CpG sites that correlate between maternal and fetal pairs are enriched in pathways associated with metabolic, cardiovascular, and immune function.69 This study suggests that both genetic and epigenetic factors contribute to PTB, an observation that should be considered in the design of future studies.

**Conclusion**

Although the precise patterns of PTB heritability have yet to be determined, studies overwhelmingly suggest that genetic factors contribute to PTB risk. Candidate gene studies have provided an initial platform for exploring genetic variation associated with PTB, but have yielded few convincing contributions to explain PTB etiology. As summarized in Table 1, GWASs have generated some promising new leads. Studies that incorporate family-based designs versus unrelated spontaneous cases and controls are particularly appealing. Linkage studies in families have produced promising initial results, and further studies may help to identify genes involved in the etiology of spontaneous PTB. Similarly, epigenetic studies of families may be useful for incorporating the influence of a shared environment into ongoing studies of PTB. In time, such studies may facilitate biomarker development. Clearly, implementation of early preventative measures for women

**Table 1 Summary of genome-wide studies of PTB**

<table>
<thead>
<tr>
<th>Design</th>
<th>N</th>
<th>Major findings</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case–control GWAS</td>
<td>2,000 mothers</td>
<td>rs12682166 associated with PTB</td>
<td>51</td>
</tr>
<tr>
<td>Case–control GWAS</td>
<td>2,000 neonates</td>
<td>No association after multiple test correction</td>
<td>52</td>
</tr>
<tr>
<td>Case–control GWAS</td>
<td>1,919 mothers; 1,964 neonates</td>
<td>rs17527054 (RNASET2) and rs3777722 (MHC) associated with PTB in neonates; no associations in mothers</td>
<td>48</td>
</tr>
<tr>
<td>Admixture study in cases</td>
<td>177 mothers</td>
<td>7q21–22 identified as associated with PTB</td>
<td>54</td>
</tr>
<tr>
<td>Family-based linkage and case–control association</td>
<td>120 mothers; 348 mothers; 334 neonates</td>
<td>15q26.3 linked to PTB; haplotype analysis indicated IGF1R; IGF1R haplotype associated in neonates but not mothers</td>
<td>14</td>
</tr>
<tr>
<td>Family-based linkage and case–control association</td>
<td>89 family members; 473 mothers; 468 neonates</td>
<td>Xq13.1 (rs6525299) linked to PTB: androgen receptor repeat associated with PTB in neonates</td>
<td>56</td>
</tr>
<tr>
<td>EWAS</td>
<td>39 mothers; 39 neonates</td>
<td>Identified CpG sites that correlate between mother and neonate, some of which are specific to PTB pairs</td>
<td>69</td>
</tr>
</tbody>
</table>

**Abbreviations:** GWAS, genome-wide association study; EWAS, epigenome-wide association study; MHC, major histocompatibility complex; IGF1R, insulin-like growth factor receptor 1; PTB, preterm birth; RNASET2, ribonuclease T2.
most at risk would have enormous positive implications for both acute and long-term maternal and fetal health.

The genetic studies presented in this review each have a different set of limitations, and further research in this area is clearly required. It is, however, tempting to consider the similarities across the results of the studies published to date. Overall, studies using linkage,\textsuperscript{14} admixture,\textsuperscript{54} and epigenome-wide\textsuperscript{680} approaches independently identify genes involved in metabolism and immune function. Indeed, women who deliver preterm are at increased risk to develop chronic disorders as they age,\textsuperscript{70–78} suggesting that spontaneous PTB may result from subsyndromic metabolic or immune dysregulation or may be an early indicator of future risk for mothers and their fetuses.

**Acknowledgments**

This work was supported by grants from the National Institutes of Minority and Health Disparities (MD009064 to AKS).

**Disclosure**

Salary support for SEP and AKK was provided, in part, by National Institute of Environmental Health Sciences and National Institute of General Medical Sciences (T32ES012870) and National Institute of General Medical Sciences (T32GM008490), respectively. The authors report no other conflicts of interest in this work.

**References**


