Seminars in Fetal & Neonatal Medicine 21 (2016) 89–93

Contents lists available at ScienceDirect

Seminars in Fetal & Neonatal Medicine

journal homepage: www.elsevier.com/locate/siny

Review

The genomics of prematurity in an era of more precise clinical phenotyping: A review

Tracy A. Manuck [a, b, *]

[a] Department of Obstetrics and Gynecology, Division of Maternal–Fetal Medicine, University of North Carolina – Chapel Hill, Chapel Hill, NC, USA
[b] Department of Obstetrics and Gynecology, Division of Maternal–Fetal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA

Keywords:
Infant
Premature
Premature birth
Phenotype
Genetic predisposition
Polymorphism

SUMMARY

Spontaneous preterm birth is a major public health problem, with a clear genetic component. Genetic association studies have evolved substantially in recent years, moving away from the traditional candidate gene analyses to newer approaches utilizing sophisticated analysis platforms to examine sequencing data, and shifting towards functional studies including methylation analysis. It is becoming increasingly evident that careful clinical phenotyping is crucial to high quality genetic association studies regardless of the assay or platform being used. Nonetheless, genetic studies of prematurity are hampered by numerous challenges including small sample sizes, incomplete phenotying, population stratification, and multiple comparisons. As the costs of sequencing and functional analyses continue to decrease, unbiased genome-wide assays will be more widely available. Researchers have met improved success recently when critically applying clinical phenotyping knowledge to group women prior to analyzing genotyping results. Eventually, as the analytic approaches evolve, it is likely that this methodology (combining precisely clinically phenotyped subjects with genome-wide data) will provide key information regarding the pathophysiology of prematurity, and provide potential new avenues for exploring innovative therapeutic strategies.

1. Introduction

Approximately 12% of infants born in the USA are preterm at <37 weeks of gestation. Preterm infants account for >70% of the neonatal morbidity and mortality in the USA, and are 40 times more likely to die in the neonatal period than their term counterparts [1]. Despite the magnitude of this clinical problem, preterm birth (PTB) remains a clinical and scientific enigma. Activation of gestational tissues and eventual preterm labor with delivery is likely a common final pathway triggered by multiple mechanisms, including hormonal mediation, inflammation and infection, and genetic factors [2]. There are many steps during the process at which an individual's genotype may affect phenotype in this complex and multistep cascade.

There is strong evidence of a genetic contribution to spontaneous PTB (sPTB). The strongest predictor of sPTB is a previous history of sPTB [3]. sPTB recurs in 35–50% of women, and tends to recur at similar gestational ages [4]. Likewise, the probability of sPTB increases with the number of prior sPTBs a woman has experienced, the most recent birth being the most predictive [5]. Women who themselves were born prematurely are at increased risk of sPTB [6,7]. The heritable nature of this complication is further supported by the findings of Winkvist and colleagues who determined that the risk of sPTB was elevated in women whose sisters had experienced sPTB (odds ratio: 1.94; 95% confidence interval: 1.26–2.99) [8]. Based on twin studies, the heritability of preterm birth is estimated to be 30% [9–11]. The racial disparity in the PTB rate cannot be explained by socio-economic factors alone, suggesting differences in risk-predisposing allele frequencies [12–16].

Numerous studies across a variety of platforms and analytic approaches have examined the relationship between genetics and prematurity. Unfortunately, results have often been inconsistent between populations and difficult to reproduce. Furthermore, though PTB is a multifactorial process leading to a final common pathway resulting in contractions, cervical dilation, and eventual delivery of the neonate, many studies of PTB have limited

* Address: UNC Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine, 3010 Old Clinic Building, CB#7516, Chapel Hill, NC 27599-7516, USA. Tel.: +1 919 966 1601; fax: +1 919 966 6377.
E-mail address: tmanuck@med.unc.edu.

http://dx.doi.org/10.1016/j.siny.2016.01.001
1744-165X/

additional phenotype information, and genetic analyses are based upon relatively heterogeneous groups of women. Recent investigations have shown that women may be grouped into distinct PTB phenotypes (including infection/inflammation, decidual hemorrhage, etc.) using clinical information. These clinical classifications can characterize groups of women (e.g. those delivering very preterm), and may provide further insight beyond the heterogenous PTB categorization [17]. Although significant overlap between phenotypes exists, distinct differences have been observed between phenotypes, as those delivering at the earliest gestational ages and Blacks have notably different phenotypic profiles compared to others [17,18]. It has been hypothesized that suboptimal clinical phenotyping is a major contributing factor to the challenges of reproducibility that have characterized genetic investigations of prematurity.

The purpose of this review is to examine current investigative techniques for genetic studies, and to summarize current knowledge regarding the contribution of genomic studies to PTB, focusing on and highlighting phenotype-specific genetic investigations.

2. Overview of genetic approaches

The study of genomic DNA allows investigators to correlate genotype with phenotype, and has evolved substantially over the past several decades. Traditionally, studies have used candidate gene approaches to evaluate whether certain alleles confer an increased or decreased risk of certain diseases. Candidate gene studies represent the largest body of work across the medical literature with regard to genetic association studies, and involve an a-priori hypothesis regarding biologic function and genetic association. This type of study may be 'hit or miss' depending on whether the 'right' gene was selected by the researchers.

In contrast, examination of the entire exome or entire genome offers a more comprehensive and unbiased approach. High-throughput genotyping, including most genome-wide association study (GWAS) platforms, are now widely accessible to investigators. Although GWAS "scans the genome," this technique examines only a subset of single nucleotide polymorphisms (SNPs) in the genome that are in linkage disequilibrium with other, non-genotyped SNPs. Unfortunately, GWAS has little power to detect rare, potentially causal, genetic variants [19]. The combination of unsuccessful research attempts to find a single causal gene and the multifactorial nature of spontaneous PTB makes it likely that multiple genetic variants and gene–environment interactions that contribute to this disorder have yet to be elucidated. Recently, advanced technology has driven down the costs of sequencing, and it is now feasible for investigators to study every base pair in the human genome. The possibilities with regard to sequencing analyses are endless; the challenges lie in the analysis and interpretation of these large volumes of data [20].

Epigenetic modifications provide another mechanism by which genes may interact with the environment; they affect gene expression by inducing structural changes in DNA that are maintained through cell division, respond to changes in the environment, yet are potentially reversible and can be targets for disease therapy [21]. DNA methylation at cytosine–guanine dinucleotides (CpG sites) – the most commonly studied epigenetic modification in humans – guides temporal and tissue-specific gene expression during fetal development and tissue differentiation. Even subtle environmental changes may induce epigenetic changes and have effects on phenotype. Methylation is tissue specific; results from blood, placental or cervical tissue, for example, typically cannot be directly compared. However, methylation studies in blood may be a reliable correlate of physiologic processes in other tissues [22]. Methylation analyses can be candidate gene and very focused, or can be genome wide. Limited studies of epigenetics in obstetrics have demonstrated identifiable differences between women delivering preterm and those with term deliveries only, across a spectrum of tissue types (placenta, maternal blood, cord blood) [23–28]. Many PTB risk factors result in DNA methylation differences – for example, methylation patterns are associated with stress, diet, smoking, inflammatory cytokine levels, and medication exposure [29–32].

Sophisticated analysis software has become increasingly widespread for researchers to analyze genetic data in the framework of physiologic pathways through systems biology. Such programs include Ingenuity Pathway Analysis (IPA; Qiagen, Valencia, CA, USA), DAVID (National Institute of Allergy and Infectious Diseases, NIH), Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway, and others. The majority of these programs function by using complex algorithms that analyze the role of a list of genes provided by the user, to search for connected functions and pathways, or overrepresentation of a particular type of gene function (e.g. genes involved in cell signaling) or ontologies.

Additional techniques not directly related to DNA are beyond the scope of this review.

A summary of example gene pathways and representative genes frequently associated with each main preterm birth phenotype, including preterm labor with intact membranes, is shown in Table 1. Each phenotype, with related genetic associations, is discussed in additional detail below.

3. Genetics of preterm labor with intact membranes

Although the largest number of studies has focused on "idiopathic PTB," this phenotype is the most heterogeneous and should be considered a diagnosis of exclusion. In the most general sense, studies of spontaneous PTB often encompass those with PTB due to cervical insufficiency, preterm premature rupture of membranes (PPROM), placental abruption, uterine overdistension, or a combination of these. Women without any PTB risk factors frequently will have an incomplete phenotypic evaluation during pregnancy. For example, routine endovaginal cervical length screening is not always performed on low risk women. Therefore, it is possible that some women thought to have "idiopathic" PTB may indeed have cervical insufficiency. This may occur if a woman has a clinically unrecognized short cervix, which then progresses to preterm cervical dilation and eventually preterm labor. When the woman presents with advanced dilation, it is then impossible to discern the inciting event leading to the advanced preterm labor. Thus, there is considerable overlap between definitions of sPTB with intact membranes and PPROM, as many studies use different thresholds to define PPROM (minimum interval between rupture and delivery, or minimum interval between rupture and development of frequent uterine contractions). The distinction between sPTB with intact membranes and abruption is also challenging, given that decidual hemorrhage is a recognized risk factor for sPTB, and bloody show may be difficult to distinguish from an early clinical abruption. In all situations, incomplete clinical data collection and clinical phenotyping, particularly in the case of retrospective analyses, may introduce heterogeneity into the analysis and limit statistical power to detect genetic differences.

Nonetheless, despite these limitations, several genetic studies have been conducted using samples from women with clearly phenotyped PTB. These studies have most consistently found