Divergent Patterns of Mitochondrial and Nuclear Ancestry Are Associated with the Risk for Preterm Birth

Nicholas Crawford, PhD[1,2,*], D'Arcy Prendergast, BSc[3,*], John W. Oehlert, MSc[4], Gary M. Shaw, PhD[4], David K. Stevenson, MD[4], Nadav Rappaport, PhD[5], Marina Sirota, PhD[5], Sarah A. Tishkoff, PhD[1,2], and Neal Sondheimer, MD, PhD[1,6]

Objective To examine linkages between mitochondrial genetics and preterm birth by assessing the risk for preterm birth associated with the inheritance of nuclear haplotypes that are ancestrally distinct from mitochondrial haplogroup.
Study design Genome-wide genotyping studies of cohorts of preterm and term individuals were evaluated. We determined the mitochondrial haplogroup and nuclear ancestry for individuals and developed a scoring for the degree to which mitochondrial ancestry is divergent from nuclear ancestry.
Results Infants with higher degrees of divergent mitochondrial ancestry were at increased risk for preterm birth (0.124 for preterm vs 0.105 for term infants; $P < .05$). This finding was validated in 1 of 2 replication cohorts. We also observed that greater degrees of divergent ancestry correlated with earlier delivery within the primary study population, but this finding was not replicated in secondary cohorts born preterm.
Conclusions Individuals with divergent patterns of mitochondrial and nuclear ancestry are at increased risk for preterm birth. These findings may in part explain the higher rates of preterm birth in African Americans and in individuals with a matrilineal family history of preterm birth. *(J Pediatr 2017;■■:■■-■■).*

Preterm birth, delivery before 37 completed weeks of gestation, is a critical public health problem that has resisted sustained efforts to develop a deeper understanding of its causes.[1] Epidemiologic studies have strongly indicated that the risk of having a preterm infant is heritable. One distinctive feature of this heritability is that it appears to be maternally biased; pregnancy history studies of a woman's relatives through her mother, but not through her father, are correlated with greater risk.[2,3]

The risk of preterm birth is also unequally distributed between different ethnic groups. African American women are at increased risk for delivering preterm compared with the population as a whole.[4] Determining the etiologies responsible for this heightened risk is complex, because it can be difficult to distinguish environmental and genetic influences. Attempts to elucidate such influences have evaluated carefully controlled populations, such as members of the military, to conclude that African ancestry is an independent risk factor.[5,6] Specifying genetic risk factors in a related group is greatly complicated by false attribution, but 1 possible link between Americans of African descent is that African haplogroup mitochondrial DNA (mtDNA) is highly prevalent in this group.[7]

These 2 risk categories—matrilineal pedigree and mitochondrial ancestry—are consistent with a mitochondria-inherited risk factor. There is also evidence that mitochondrial function is important for the maintenance of pregnancy. Chemically induced damage to mitochondrial activity is sufficient to disrupt pregnancy in a mouse model[8] and mitochondrial parameters such as membrane potential are regulated in embryonic tissues.[9] Furthermore, women who carry the pathogenic mtDNA mutation mt.3243A>G have a higher risk for preterm birth, although they are clinically unaffected.[10] However, attempts to link preterm birth to specific polymorphisms in mtDNA in case-control studies failed to identify polymorphisms associated with preterm birth.[11] Thus, it remains unclear how a mtDNA-specific inheritance pattern could be associated with a common outcome like preterm birth.

In this study, we considered that risk might derive from the interplay of mitochondrial and nuclear inheritance. The mtDNA and nuclear genome must coordinate to create the multisubunit complexes required for the electron transport chain and the generation of cellular energy. Nuclear DNA is biparentally

BBC	Genome Wide Association Study of Preterm Birth/Boston Birth Cohort
BPD	Genome-Wide Association Study for Bronchopulmonary Dysplasia
DNBC	Danish National Birth Cohort Study
GENEVA-AA	Genome-Wide Association Studies of Prematurity and Its Complications
GPN	Genomic and Proteomic Network for Preterm Birth Research
HRS	Genetics Resource with the Health and Retirement Study
mtDNA	Mitochondrial DNA

From the [1]Department of Genetics; [2]Department of Biology, The University of Pennsylvania, Philadelphia, PA; [3]Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; [4]Department of Pediatrics, Stanford University, Palo Alto; [5]Department of Pediatrics, University of California San Francisco, San Francisco, CA; and [6]Department of Pediatrics, The University of Toronto, Toronto, Ontario, Canada

*These authors contributed equally.

Funded by The March of Dimes Prematurity Research Center at the University of Pennsylvania (22-FY17-890). The authors declare no conflicts of interest.

https://doi.org10.1016/j.jpeds.2017.10.052

inherited and modern individuals may have highly admixed nuclear ancestry as mobility has increased. By contrast, mtDNA is inherited as a unified haplogroup in a matrilineal fashion, allowing modern individuals to trace their ancestry back to a single geographically isolated region. Differences in ancestral variants in mtDNA affect mitochondrial function.[12] Furthermore, studies in which the mtDNA from 1 mouse strain was transferred into a distinct nuclear background have identified an array of behavioral and metabolic changes that result from such a misalignment of mitochondrial haplogroup and nuclear haplotype.[13-15] In this study, we have developed tools to determine the alignment between nuclear and mitochondrial inheritance, and have used this factor to evaluate previously collected data on genetic risk factors for preterm birth, hypothesizing that differences in the ancestral inheritance of an infant's mitochondrial and nuclear genome would be associated with an increased risk of preterm birth.

Methods

We identified and analyzed all datasets within the database of Genotypes and Phenotypes (dbGAP) that were case-control studies of preterm birth, containing sufficient markers for nuclear and mitochondrial genotyping that were focused on mothers or infants with diverse ancestries. The primary analysis was performed in the *Eunice Kennedy Shriver* Institute of Child Health and Human Development Genomic and Proteomic Network for Preterm Birth Research (GPN) cohort (phs000714.v1.p1).[16] Validation of the effects was attempted in the Boston Birth Cohort (BBC; phs000332.v3.p2),[17] the Danish National Birth Cohort Study (DNBC; phs000103.v3.p1)[18] and the Genome-Wide Association Studies of Prematurity and Its Complications (GENEVA-AA; phs000353.v1.p1).[19] We also performed validation using the Genome-Wide Association Study for Bronchopulmonary Dysplasia (BPD).[20] Additional controls as needed were obtained from the Genetics Resource with the Health and Retirement Study (HRS; phs000428.v1.p1). All subjects with evaluable genotypes and phenotypes in each dataset were analyzed. The underlying characteristics of each dataset are provided in **Table I** (available at www.jpeds.com).

The GPN, BBC, DNBC, and HRS datasets were downloaded from dbGAP after approval and Research Ethics Board or Institutional Review Board review. The BPD Study data were evaluated under the approvals of the Institutional Review Board of Stanford University and the Health and Welfare Agency Committee for the Protection of Human Subjects of the State of California.

To determine haplogroup, mitochondrial polymorphisms were extracted from the datasets using the open-source whole genome association analysis toolkit, PLINK.[21] Haplogroups were calculated using Haplogrep.[22] L-branch haplogroups, U6 and U5b1, were scored (H) as 1 and non-L haplogroups were scored as 0. Subjects with haplogrouping confidence scores of ≤0.5 (low confidence) were excluded from further analysis.

Nuclear ancestry was determined using ADMIXTURE.[23] Briefly, autosomal markers were integrated with data from the 1000Genomes project from individuals of African, European, South Asian, East Asian, and Native American descent.[24] Minor allele frequencies (>0.05) were pruned for linkage disequilibrium using PLINK to obtain approximately 25 000 informative markers. ADMIXTURE was run using unsupervised analyses and the k-factor providing best resolution of African ancestry individuals was used for analysis.

Divergent ancestry was calculated with a focus on African ancestry. The degree of African nuclear ancestry (A_A) was defined as the ADMIXTURE-derived fraction of ancestry that aligned to the reference African individuals in the 1000Genomes project. Divergent ancestry was calculated as:

$$\text{Divergent ancestry} = |H - A_A|$$

Divergent ancestry ranged from 0 to 1, with matched ancestries yielding a score of 0 and completely unmatched ancestries yielding a score of 1.

Statistical Analyses

For datasets where preterm birth was studied as a quantitative trait using weeks of estimated gestational age (GPN, BBC) the results were evaluated by Pearson correlation. For datasets where preterm birth was studied as a qualitative trait (BPD, HRS), comparisons were made by 2-tailed Student t test. Confounders were evaluated using phenotypes reported in the parent datasets. Associations were tested using 1-way ANOVA or the Student t test, depending on the number of classes tested. Statistical testing was performed using Prism 6 (GraphPad, La Jolla, CA), except for logistic regression analyses, which were performed in R (R Core Team, Vienna, Austria). Bonferroni correction was applied to the statistical analysis of the BPD/HRS data to account for multiple testing.

Results

The *Eunice Kennedy Shriver* Institute of Child Health and Human Development GPN study was a multisite, multiethnic study of preterm birth risk of American women experiencing preterm birth at <34 weeks of gestation with comparison to term controls at >39 weeks of gestation.[16] No pregnancies with births between 34 and 39 weeks of gestation were present in the dataset. DNA samples from women and their infants were obtained. Gestational ages determined by a pediatrician were recorded for most infants. Cases and controls were spontaneously in labor, and women were excluded for polyhydramnios, uterine anomalies, cerclage, fetal aneuploidy, or lethal fetal anomaly. Controls were excluded if they had a prior history of preterm birth. Preterm birth cases and controls were balanced for maternal age, self-reported race, and history of prior pregnancy.

Unsupervised ADMIXTURE analysis of the mothers and infants was performed together with the 1000Genomes project individuals of African, European, South Asian, East Asian, and

FLA 5.5.0 DTD ■ YMPD9560_proof ■ December 13, 2017