Published in final edited form as:
Obstet Gynecol. 2011 November ; 118(5): 1081–1089. doi:10.1097/AOG.0b013e31823389bb.

Role of African Ancestry and Gene-Environment Interactions in Predicting Preterm Birth

Hui-Ju Tsai[1,2,*], **Xiumei Hong**[1], **Jinbo Chen**[3], **Xin Liu**[1], **Colleen Pearson**[4], **Katherin Ortiz**[4], **Emmet Hirsch**[5,6], **Linda Heffner**[7], **Daniel E. Weeks**[8], **Barry Zuckerman**[4], and **Xiaobin Wang**[1]

[1]Mary Ann and J. Milburn Smith Child Health Research Program, Children's Memorial Hospital and Children's Memorial Research Center; Department of Pediatrics, Feinberg School of Medicine, Northwestern University; Chicago, IL, USA

[2]Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Taiwan

[3]Department of Biostatistics and Epidemiology, School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[4]Department of Pediatrics, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA

[5]Department of Obstetrics and Gynecology, North Shore University Health System, Evanston, IL, USA

[6]Department of Obstetrics and Gynecology, Pritzker School of Medicine, University of Chicago, Chicago, IL, USA

[7]Department of Obstetrics and Gynecology, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA

[8]Department of Human Genetics and Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

Objective—To estimate whether African ancestry, specific gene polymorphisms, and gene-environment interactions could account for some of the unexplained preterm birth variance within blacks.

Methods—We genotyped 1,509 African ancestry informative markers, cytochrome P-450 1A1 (*CYP1A1*) and glutathione S-transferases Theta 1 (*GSTT1*) variants in 1,030 self-reported black mothers. We estimated the African ancestral proportion using the ancestry informative markers for all 1,030 self-reported black mothers. We examined the effect of African ancestry and *CYP1A1* and *GSTT1* smoking interactions on preterm birth cases as a whole and within its subgroups: very preterm birth (gestational age less than 34 weeks); and late preterm birth (gestational age greater than 34 and less than 37 weeks). We applied logistic regression and receiver operating characteristic (ROC) curve analysis, separately, to evaluate if African ancestry and *CYP1A1*- and *GSTT1*-smoking interactions could make additional contributions to preterm birth beyond epidemiological factors.

[*]Correspondence to: Hui-Ju Tsai, MPH, PhD, Associate Investigator, Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, 35, Keyan Road, Zhunan Town, Miaoli County 350, Taiwan, Phone: +886-37-246-166 ext. 36150; Fax: +886-37-583-467, tsaihj@nhri.org.tw; htsai@childrensmemorial.org; h-tsai@northwestern.edu.

Financial Disclosure: The authors did not report any potential conflicts of interest. Additional tables and an additional figure are available online at http://links.lww.com/xxx.

Results—We found significant associations of African ancestry with preterm birth (22% vs. 31%, OR=1.11; 95%CI: 1.02–1.20) and very preterm birth (23% vs. 33%, OR=1.17; 95%CI: 1.03–1.33), but not with late preterm birth (22% vs. 29%, OR=1.06; 95%CI: 0.97–1.16). In addition, the ROC curve analysis suggested that African ancestry and *CYP1A1*- and *GSTT1*-smoking interactions made substantial contributions to very preterm birth beyond epidemiologic factors.

Conclusion—Our data underscore the importance of simultaneously considering epidemiological factors, African ancestry, specific gene polymorphisms and gene-environment interactions to better understand preterm birth racial disparity and to improve our ability to predict preterm birth, especially very preterm birth.

Introduction

Preterm birth (PTB) is commonly defined as delivery before 37 weeks of gestation. PTB is a serious public health problem[1], threatening all ethnic groups, but particularly Black populations.

The contemporary African American population is a genetic admixture between Africans and Europeans[5]. Genetic ancestry has been linked to a number of health outcomes with a known Black-White disparity, including asthma, cardiovascular diseases, prostate cancer, lung function and PTB[11–13]. However, it remains largely unknown whether genetic ancestry can further explain individual variation in self-reported Black women[14].

In an earlier report[14], using 57 ancestry informative markers (AIMs) in 812 women, we found that African ancestry was associated with PTB and its related traits. However, our previous ancestral estimates may have not been highly accurate due to the fact that we only used a set of 57 AIMs. Therefore, our current study is necessary to strengthen and expand on our previous findings, and is bound to provide more convincing evidence for the potential usefulness of genetic ancestry in the preterm birth research. In addition, our team previously identified interactions of two genes, cytochrome P-450 1A1 (*CYP1A1)* and glutathione S-transferases Theta 1 (*GSTT1),* and maternal smoking associated with PTB in the same population[18–20].

In this study, we first examined whether the previously observed association was retained between African ancestral proportion and PTB and its subphenotypes (including very PTB and late PTB) by genotyping several more AIMs in a larger set of subjects to estimate African ancestral proportion. Second, we assessed the association between African ancestry and PTB, very PTB and late PTB in self-reported Black women after controlling for pertinent risk factors. Finally, we examined whether including African ancestral proportion, gene polymorphisms and GxE interactions could make additional and independent contributions to predicting PTB, very PTB, and late PTB beyond known epidemiological risk factors.

Materials and Methods

Study Subjects and Data Collection

This study included a subset of 1,030 Black mothers (518 PTB cases and 512 controls) enrolled in an ongoing case-control study of preterm birth at the Boston Medical Center (BMC)[21]. The current study overlaps with our earlier study[14], in that 441 of the 812 subjects in our first study also participated in this study. The enrollment period was from 1998 to 2008. Case mothers were those who delivered singleton, live births occurring at less than 37 weeks of gestation, and controls were defined as mothers delivering at greater than or equal

to 37 weeks of gestation with birth weight appropriate for gestational age as defined by the National Center for Health Statistics/CDC guidelines (birth weight between 2,500 and 4,000 grams)[22]. Pregnancies resulting in multiple births and newborns with major birth defects were excluded. A detailed description of the study population was previously published[21]. The Institutional Review Boards of the Boston University Medical Center, the Massachusetts Department of Public Health, and Children's Memorial Hospital in Chicago approved the study protocol. All participants gave written informed consent.

Definition of Phenotype

Preterm birth (PTB) was evaluated as both a binary (<37 weeks of gestation vs. ≥37 weeks of gestation) and a continuous (gestational age) variable. Gestational age was assessed using an algorithm based on last menstrual period and the result of early ultrasound (<20 weeks of gestation). The last menstrual period estimate was used only if confirmed by ultrasound within 7 days or if no ultrasound estimate was obtained; otherwise, the ultrasound estimate was used. This approach has been used in previous studies[21,23].

In this study, we used a cut-off point of <34 weeks to define very preterm birth (very PTB), and 34.0–36.86 weeks to define late preterm birth (late PTB), which has been used by other groups[24,25]. In addition, we categorized PTB cases as spontaneous PTB if they occurred secondarily to documented active preterm labor (uterine contractions with cervical effacement and dilation at less than 37 weeks), or preterm premature rupture of membranes (PPROM) (< 37 weeks without uterine contractions) or both uterine contractions and PPROM occurring simultaneously; or as indicated, including PTB that was defined as delivery, which was not preceded by the presence of uterine contractions and/or rupture of membranes. A detailed description was previously published[14].

Genotyping Data

We genotyped a total of 1,509 AIMs previously identified as highly informative between African and European ancestry[12,26] for 1,030 Black mothers (518 PTB cases and 512 matched controls). Specifically, we applied the Illumina African American Panel as our genotyping platform (http://www.illumina.com/products/african_american_admixture_panel.ilmn). For quality control, four duplicate DNA samples were placed on each 96-well plate. The concordance rate of these duplicate samples was >99.5%.

In addition, we genotyped genetic markers within the two genes, *CYP1A1* and *GSTT1,* separately. Detailed information regarding DNA extraction, PCR condition and quality check of *CYP1A1* and *GSTT1* has been described elsewhere[19,21]

Statistical Analysis

Several analytical approaches were used in this study. First, we obtained ancestral estimates for each subject using the Structure program[27,28]. Second, we compared the equality of ancestral distributions between PTB and term controls, very PTB and term controls, and very PTB and late PTB, respectively, using the Kolmogorov-Smirnov statistic. Third, using stepwise model selection in a logistic regression framework, we identified a set of significant risk factors and further tested the association of significant risk factors, African ancestry, *CYP1A1, GSTT1,* and their interactions with smoking with PTB, very PTB and late PTB, individually. Last, the receiver operating characteristic (ROC) curve analysis and the corresponding *c* statistic were obtained, and a non-parametric statistic was applied to assess the discrimination ability of different predictive models. Below, we provide detailed information about each analytical approach applied in this study.

First, we applied an admixture model implemented in the program Structure to estimate individual admixture proportions using our panel of 1,509 AIMs[27,28]. Specifically, the admixture model assumes that each individual inherits some proportion of their ancestry from each ancestral population. To compute African ancestral estimates, we input genotyping data from both ancestral populations (Africans and Europeans) (of note, genotyping data for both ancestral populations were from the International HapMap project (http://hapmap.ncbi.nlm.nih.gov/)), specified as *known populations*, and from admixed subjects, specified as an *unknown population*. We then assumed an admixture model and used default values for other parameters provided by Structure with 5,000 burn-in and 50,000 further iterations through the Markov chain Monte Carlo (MCMC) algorithm. In addition, we generated plots to compare the distribution of African ancestral proportion between PTB and term controls, very PTB and term controls, and very PTB and late PTB, respectively. We also carried out the non-parametric test using the Kolmogorov-Smirnov statistic for the equality of distributions.

Since this study was based on a "loosely age-matched" design, we further took into account this "matching-adjusted age effect" to ensure that we did not misinterpret the results. We adjusted for this matching effect as follows. We first obtained the prevalence of PTB for each matching stratum. In the formula we let n_{case} be the number of PTB cases in the stratum a, $n_{control}$ be the number of controls in the same stratum and p_a be the prevalence of PTB in this stratum. We used this formula: $\log(n_{case}/n_{control}) - \log(p_a/(1-p_a))$ to compute an age-adjusted variable for each subject within each corresponding stratum[29]. Then we applied subsequent logistic regression while including this age-adjusted variable using an "offset" option implemented in STATA.

Next, we identified a set of risk factors using stepwise model selection (cutoff p value < 0.15). Specifically, the examined known risk factors of PTB including: maternal age (<20, 20–24, 25–29, 30–34 and ≥35 years), education (≤ middle school, high school, and > high school), parity (0, 1, and ≥2), marital status (married, other), maternal pre-pregnant body mass index (BMI) (<20, 20–24, 25–29, and ≥30 Kg/m^2), maternal smoking during pregnancy (current smoking, quitters and none smoking), illicit drug use (yes/no), stress (not stressed, average stress and very stressed) and number of years in the U.S. (born in the U.S., <5, and >5 years). In addition, we used logistic regression to examine the association of African ancestry proportion, *CYP1A1*, *GSTT1* and their interactions with maternal smoking, with PTB, very PTB and late PTB, respectively. The odds ratio (OR) is expressed for each 10% increment of African ancestry proportion.

We further assessed the model performance for the added value of African ancestry, *CYP1A1*, *GSTT1* and their interactions with maternal smoking for the prediction of PTB, very PTB and late PTB, individually, using the ROC curve analysis[30]. Specifically, we evaluated the predictability across different models by calculating the concordance (c) statistic, the most commonly used quantity for indicating the discrimination ability of different models. In addition, we applied a non-parametric statistic implemented in STATA to test the equality of the area under the curve across four sequential predictive models[31]. The set of risk factors assessed in the predictive models was as follows: the known epidemiological variables described above, African ancestral proportion, two previously identified genetic polymorphisms (*CYP1A1* and *GSTT1*) and their interactions with maternal smoking. Finally, to evaluate if African ancestry was co-linear with known PTB epidemiological risk factors, we computed the Pearson correlation coefficient between African ancestry and maternal age, BMI and years in the US, and the point-biserial correlation coefficient between African ancestry and education, marital status and parity. Data analyses were performed using statistical packages R 2.10.0 (http://www.r-project.org) and Intercooled STATA 11.0 (College Station, TX).

Results

A total of 1,030 Black mothers (518 PTB cases and 512 term controls) were included in the study. The 518 PTB cases were further divided into 211 very PTB and 307 late PTB. Table 1 shows the demographic, clinical and genetic characteristics of the study subjects, stratified by PTB, very PTB, late PTB and term controls. The average and corresponding standard deviation (SD) of African proportions was 0.88 (SD = 0.15) in PTB cases, 0.88 (SD = 0.12) in very PTB, 0.87 (SD = 0.16) in late PTB and 0.85 (SD = 0.16) in controls. The distribution of African ancestry between PTB cases and controls is provided in Figure 1 in the Appendix, available online at http://links.lww.com/xxx. In addition, the Kolmogorov-Smirnov statistic indicated that the distributions of African ancestral proportion did not differ significantly in PTB vs. term controls ($P = 0.78$), very PTB vs. term controls ($P = 0.20$), and very PTB vs. late PTB ($P = 0.18$).

For the known PTB epidemiological risk factors, we applied stepwise model selection to identify a subset of important risk factors, which were included in the subsequent predictive models (Table 2). Consistent with previous findings, maternal smoking during pregnancy and overall stress were significantly associated with PTB, whereas illicit drug use was borderline significant. In addition, we examined the association of African ancestry and *CYP1A1*, *GSTT1*, and their interactions with maternal smoking with PTB, very PTB and late PTB, separately. . Notably, African ancestry was significantly associated with PTB (22% vs. 31%, OR=1.11 for every 10% increment; 95% CI: 1.02–1.20) and very PTB (23% vs. 33%, OR=1.17; 95% CI: 1.03–1.33), but not with late preterm birth (22% vs. 29%, OR=1.06; 95%CI: 0.97–1.16), whereas *CYP1A1*-maternal smoking was significantly associated with PTB (OR=1.83; 95% CI: 1.20–2.81) and late PTB (OR=2.03; 95% CI: 1.30–3.19), but not *GSTT1*-maternal smoking (Table 2).

Moreover, we performed stepwise model selection by stratifying PTB into the subtypes of spontaneous PTB and indicated PTB to identify a set of important PTB risk factors. Interestingly, the same set of important risk factors was identified among PTB, very PTB, late PTB and spontaneous PTB, but not indicated PTB (Tables 1 and 2 in the Appendix, available online at http://links.lww.com/xxx).

Furthermore, we performed ROC curve analysis. We evaluated four different models for predicting PTB, very PTB and late PTB, separately, as follows: 1) Model 1: a set of known PTB epidemiological risk factors, including: smoking, illicit drug use and overall stress, which were significant factors identified from the above association tests; 2) Model 2: Model 1 plus African ancestry; 3) Model 3: Model 1 plus *GSTT1* and interaction of *GSTT1* and maternal smoking, *CYP1A1* and interaction of *CYP1A1* and maternal smoking; 4) Model 4: Model 3 plus African ancestry. As shown in Table 3, Model 4 (containing African ancestry, genetic findings and GxE interactions) showed the highest area under curve at 0.66 (95% CI: 0.61–0.70) among the predictive models for very PTB. Likewise, Model 4 (containing African ancestry, genetic findings and GxE interactions) for PTB and late PTB, individually, also showed the best discrimination ability among the four presented models. Similarly, we performed ROC curve analysis and evaluated the four predictive models described above for spontaneous PTB and indicated PTB, respectively. We also used another set of important risk factors identified for indicated PTB and carried out the corresponding ROC curve analysis (Tables 3–5 in the Appendix, http://links.lww.com/xxx). The results in Tables 5–6 in the Appendix (http://links.lww.com/xxx), showed that the best discrimination ability was observed in Model 4 (containing African ancestry, genetic findings and GxE interactions) for spontaneous PTB.

We further applied a non-parametric statistic to test the equality of the area under curve. There was significant improvement when adding African ancestry, genetic findings and GxE interactions in Model 4 (Table 4). The results in Table 4 indicated that Model 4 was significantly better than all the other models (Model 1 vs. Model 4, $p = 0.002$ for PTB; $p = 0.0007$ for very PTB; $p = 0.004$ for late PTB).

We further examined whether African ancestral proportion was associated with the two PTB susceptible genetic polymorphisms (*CYP1A1* and *GSTT1*). We found that neither of these genetic polymorphisms was significantly associated with African ancestral proportion (Table 1 in the Appendix, http://links.lww.com/xxx). In addition, we tested whether African ancestry was co-linear with known PTB epidemiological risk factors such as age, BMI, education, marital status, parity and years in the US. Interestingly, we did not find any substantial degree of correlation between African ancestry and the above known PTB epidemiological risk factors (data not shown).

Discussion

While the impact of genetic ancestry and its potential confounding effect have been examined in several common complex diseases such as asthma, breast cancer, prostate cancer and cardiovascular disease, few studies have assessed the impact of genetic ancestry in PTB and very PTB. This study has evaluated the distributions of genetic ancestry among PTB, very PTB, and late PTB compared to term controls in a sample of 1,030 U.S. Black women. We further examined the association of genetic ancestry with PTB, very PTB, and late PTB in self-reported Black subjects. Moreover, we evaluated if African ancestry, two PTB susceptibility genes and their interactions with maternal smoking could make additional and independent contributions to predicting PTB, very PTB and late PTB beyond known epidemiological risk factors of PTB.

This study has evidenced several important findings. First, we demonstrated that African ancestry was significantly associated with PTB (OR=1.11; 95% CI: 1.02–1.20) and very PTB (OR=1.17; 95% CI: 1.03–1.33), but not with late PTB (OR=1.06; 95%CI: 0.97–1.16) in the logistic regression models. However, no difference was observed in known epidemiological factors between the very preterm birth group and the late preterm birth group in this study. It indicated that those known epidemiological factors may have similar impact in both preterm subgroups. Interestingly, compared to PTB, previous studies also reported a greater Black-White disparity for very PTB[2]. This finding indicates that the influence of genetic ancestral background may have a greater impact on very PTB than on late PTB. Additionally, it is likely that the effect of gene-environment interaction may be also stronger in very PTB. However, at present, the underlying mechanism of genetic ancestral influence in very PTB remains largely unexplored. It will be of importance to confirm this finding in another study and further conduct functional investigation. Second, we showed that the model including African ancestry, specific gene polymorphisms and their interactions with maternal smoking provided the best discrimination ability with an AUC of 0.66 (95% CI: 0.61–0.70) for very PTB (Table 3).

Previous studies have identified a certain number of PTB-related genetic variants in the hope that these discoveries will provide deeper insights into the dissection of the etiology of PTB, and ultimately lead to the development of new therapies and/or preventive strategies[18,19,32]. However, for many complex diseases including PTB, limited studies have applied these genetic discoveries to risk assessment in clinical practice. In this study, we included African ancestry along with genetic findings, gene-smoking interactions on PTB and the known epidemiological factors in predictive models for PTB, very PTB and late PTB, respectively. Notwithstanding the results suggesting that incorporating African ancestry only or genetic

findings and their interaction with smoking only offer merely limited improvements in our ability to predict PTB, very PTB and late PTB beyond known PTB epidemiological risk factors, the results also showed that the model including African ancestry, PTB susceptible genes and gene-smoking interactions provided the highest AUC of 0.66 (95% CI: 0.61–0.70) for very PTB among the sequential predictive models (Table 3).

This study has several strengths. First, the accuracy of estimating ancestral proportion is strongly affected by the number of AIMs. In this study, we genotyped a panel of 1,509 AIMs, which provided a robust estimation of African ancestry. Second, this is a large sample of inner city Black women assembled to study the influence of African ancestry on PTB and its subgroups: very PTB and late PTB. Third, we applied ROC/AUC methods to evaluate sequential models of African ancestry, genetic contributions and GxE interactions beyond known epidemiological risk factors. Our results suggested that genetic ancestry, gene polymorphisms and their interactions with maternal smoking together can improve predictability for very PTB, although our predictive model is far from perfect and there is tremendous work that remains. Finally, our predictive models, which integrated epidemiological factors, genetic discoveries and GxE interactions, illuminate a future direction towards a better and more accurate predictive model of PTB in clinical practice.

On the other hand, this study also has some limitations. Although we have included major known risk factors of PTB in this population, as reflected in relatively modest discrimination ability, it is likely that additional risk factors are necessary to take into account in the predictive models. In addition, we only included two genes and their interactions in the predictive models. Recent studies have reported a potential interaction of cytokine polymorphisms and bacterial vaginosis in PTB development as well as the interaction of inflammatory-response regulatory polymorphisms and bacterial vaginosis[35,36]. Therefore, we anticipate that there are other important genetic factors and GxE interactions yet to be considered or discovered. The ultimate goal of our work is to develop a highly accurate and predictive model for PTB that can be used in clinical and public health settings. However, while our predictive models presented in this study were far from perfect given their modest discrimination ability, this was only the first step towards our goal. Looking forward, the performance of our predictive models needs to be validated and evaluated in independent samples.

In summary, consistent with our previous report, African ancestry was significantly associated with an increased risk of PTB and very PTB in this sample of inner city Black mothers. Our data underscore the need to simultaneously consider African ancestry, important gene polymorphisms and GxE interactions in order to better understand preterm racial disparity and to improve our ability to predict, treat and prevent PTB, especially very PTB.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported in part by grants from the National Institute of Child Health and Human Development (R01 HD41702 and R21HD066471), National Institute of Environmental Health Sciences (R01ES11682, R21ES11666), and March of Dimes Birth Defects Foundation (20-FY98-0701, 20-FY02-56 and #21-FY07-605).

The authors thank Dr. Chin-Fu Hsiao and Tsung-Kai Chang for statistical consultation, Lingling Fu and Camper Liu for data management, Tami R. Bartell for English editing, the nursing staff of Labor and Delivery at Boston Medical Center for their continuous support and assistance to the study, Ann Ramsay for administrative support, and all of the participating mothers and their families.

References

1. Mattison DR, Damus K, Fiore E, Petrini J, Alter C. Preterm delivery: a public health perspective. Paediatr Perinat Epidemiol. Jul; 2001 15(Suppl 2):7–16. [PubMed: 11520396]
2. IOM report. Preterm birth: cause, consequences, and prevention. Washington, D.C: The national academies press; 2006. Institute of medicine of the national academies (Committee on understanding premature birth and assuring healthy outcomes board on health sciences policy).
3. Macones GA, Parry S, Elkousy M, Clothier B, Ural SH, Strauss JF 3rd. A polymorphism in the promoter region of TNF and bacterial vaginosis: preliminary evidence of gene-environment interaction in the etiology of spontaneous preterm birth. Am J Obstet Gynecol. Jun; 2004 190(6): 1504–1508. discussion 1503A. [PubMed: 15284722]
4. Crider KS, Whitehead N, Buus RM. Genetic variation associated with preterm birth: a HuGE review. Genet Med. Nov–Dec; 2005 7(9):593–604. [PubMed: 16301860]
5. Parra EJ, Marcini A, Akey J, et al. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet. Dec; 1998 63(6):1839–1851. [PubMed: 9837836]
6. Pfaff CL, Parra EJ, Bonilla C, et al. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am J Hum Genet. Jan; 2001 68(1): 198–207. [PubMed: 11112661]
7. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. Genome Res. Dec; 2002 12(12):1805–1814. [PubMed: 12466284]
8. Shriver MD, Parra EJ, Dios S, et al. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet. Apr; 2003 112(4):387–399. [PubMed: 12579416]
9. Shriver MD, Mei R, Parra EJ, et al. Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. Hum Genomics. 2005; 2(2):81–89. [PubMed: 16004724]
10. Smith MW, O'Brien SJ. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. Nat Rev Genet. Aug; 2005 6(8):623–632. [PubMed: 16012528]
11. Kumar R, Seibold MA, Aldrich MC, et al. Genetic ancestry in lung-function predictions. N Engl J Med. Jul 22; 2010 363(4):321–330. [PubMed: 20647190]
12. Freedman ML, Haiman CA, Patterson N, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A. Sep 19; 2006 103(38): 14068–14073. [PubMed: 16945910]
13. Reiner AP, Carlson CS, Ziv E, Iribarren C, Jaquish CE, Nickerson DA. Genetic ancestry, population sub-structure, and cardiovascular disease-related traits among African-American participants in the CARDIA Study. Hum Genet. Jun; 2007 121(5):565–575. [PubMed: 17356887]
14. Tsai HJ, Yu Y, Zhang S, et al. Association of genetic ancestry with preterm delivery and related traits among African American mothers. Am J Obstet Gynecol. Jul; 2009 201(1):94 e91–10. [PubMed: 19446788]
15. Hoggart CJ, Parra EJ, Shriver MD, et al. Control of confounding of genetic associations in stratified populations. Am J Hum Genet. Jun; 2003 72(6):1492–1504. [PubMed: 12817591]
16. Enoch MA, Shen PH, Xu K, Hodgkinson C, Goldman D. Using ancestry-informative markers to define populations and detect population stratification. J Psychopharmacol. Jul; 2006 20(4 Suppl): 19–26. [PubMed: 16785266]
17. Patterson N, Hattangadi N, Lane B, et al. Methods for high-density admixture mapping of disease genes. Am J Hum Genet. May; 2004 74(5):979–1000. [PubMed: 15088269]
18. Hao K, Wang X, Niu T, et al. A candidate gene association study on preterm delivery: application of high-throughput genotyping technology and advanced statistical methods. Hum Mol Genet. Apr 1; 2004 13(7):683–691. [PubMed: 14976157]
19. Tsai HJ, Liu X, Mestan K, et al. Maternal cigarette smoking, metabolic gene polymorphisms, and preterm delivery: new insights on GxE interactions and pathogenic pathways. Hum Genet. May; 2008 123(4):359–369. [PubMed: 18320229]
20. Yu Y, Tsai HJ, Liu X, et al. The joint association between F5 gene polymorphisms and maternal smoking during pregnancy on preterm delivery. Hum Genet. Jan; 2009 124(6):659–668. [PubMed: 19020903]

21. Wang X, Zuckerman B, Pearson C, et al. Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight. Jama. Jan 9; 2002 287(2):195–202. [PubMed: 11779261]
22. Hamill PV, Drizd TA, Johnson CL, Reed RB, Roche AF, Moore WM. Physical growth: National Center for Health Statistics percentiles. Am J Clin Nutr. Mar; 1979 32(3):607–629. [PubMed: 420153]
23. Kramer MS, Platt R, Yang H, et al. Secular trends in preterm birth: a hospital-based cohort study. Jama. Dec 2; 1998 280(21):1849–1854. [PubMed: 9846780]
24. Engle WA, Tomashek KM, Wallman C. "Late-preterm" infants: a population at risk. Pediatrics. Dec; 2007 120(6):1390–1401. [PubMed: 18055691]
25. Mathews TJ, Menacker F, MacDorman MF. Infant mortality statistics from the 2002 period: linked birth/infant death data set. Natl Vital Stat Rep. Nov 24; 2004 53(10):1–29. [PubMed: 15622996]
26. Reich D, Patterson N, Ramesh V, et al. Admixture mapping of an allele affecting interleukin 6 soluble receptor and interleukin 6 levels. Am J Hum Genet. Apr; 2007 80(4):716–726. [PubMed: 17357077]
27. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. Aug; 2003 164(4):1567–1587. [PubMed: 12930761]
28. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. Jun; 2000 155(2):945–959. [PubMed: 10835412]
29. Breslow NE, Zhao LP. Logistic regression for stratified case-control studies. Biometrics. Sep; 1988 44(3):891–899. [PubMed: 3203135]
30. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. Apr; 1982 143(1):29–36. [PubMed: 7063747]
31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. Sep; 1988 44(3): 837–845. [PubMed: 3203132]
32. Jakobsdottir J, Gorin MB, Conley YP, Ferrell RE, Weeks DE. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. PLoS Genet. Feb.2009 5(2):e1000337. [PubMed: 19197355]
33. Ober C, Vercelli D. Gene-environment interactions in human disease: nuisance or opportunity? Trends Genet. Jan 7.2011
34. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. May; 2008 9(5):356–369. [PubMed: 18398418]
35. Jones NM, Holzman C, Friderici KH, et al. Interplay of cytokine polymorphisms and bacterial vaginosis in the etiology of preterm delivery. J Reprod Immunol. Dec; 2010 87(1–2):82–89. [PubMed: 20965572]
36. Gomez LM, Sammel MD, Appleby DH, et al. Evidence of a gene-environment interaction that predisposes to spontaneous preterm birth: a role for asymptomatic bacterial vaginosis and DNA variants in genes that control the inflammatory response. Am J Obstet Gynecol. Apr; 2010 202(4): 386, e381–386. [PubMed: 20350647]

Table 1

Demographic, clinical and genetic characteristics of Black mothers* at the Boston Medical Center, grouped by degree of preterm birth (PTB)**.

Characteristics	PTB (N=518)	Very PTB (N=211)	Late PTB (N=307)	Control (N=512)
Demographic characteristics				
	Mean ± SD			
Age (yr)	28.67 ± 6.92	28.68 ± 7.07	28.67 ± 6.81	28.68 ± 6.92
Yr in the US (yr) [a]	7.16 ± 7.08 (N=194)	8.69 ± 8.03 (N=72)	6.21 ± 6.32 (N=123)	6.51 ± 6.21 (N=263)
Prepregnancy BMI (Kg/m^2)	26.54 ± 6.17	26.76 ± 6.04	26.37 ± 6.25	26.66 ± 5.80
Parity (%)				
0	43.05	42.65	43.59	38.28
1	25.68	25.59	25.21	31.05
>= 2	31.27	31.75	31.20	30.66
Education (%)				
< High School	25.78	25.73	21.46	30.63
High School	41.99	43.69	39.91	31.03
Some college and above	32.23	30.58	38.63	38.34
Place of birth (%)				
US	53.20	55.77	51.28	37.72
Other	46.80	44.23	48.72	62.28
Smoking (%)				
Never	75.53	76.08	75.11	85.35
Quitter	6.21	6.70	6.01	6.64
Current	18.25	17.22	18.88	8.01
Illicit drug use (%)				
No	81.94	81.34	83.26	90.57
Yes	18.06	18.66	16.74	9.43
Lifetime stress (%)				
Not stressed	29.63	29.61	30.04	36.83
Average	53.22	50.49	54.51	52.87
Very stressed	17.15	19.90	15.45	10.30
Stress during pregnancy (%)				
Not stressed	29.04	28.64	29.18	39.80
Average	45.42	47.57	44.21	42.57
Very stressed	25.54	23.79	26.61	17.62
Clinical characteristics				
Gestational age (wk)	33.29 ± 3.58	29.76 ± 3.09	35.71 ± 0.81	39.48 ± 1.30
Birthweight (g)	2042.73 ± 768.00	1367.25 ± 552.33	2508.67 ± 506.32	3323.38 ± 497.99
Hypertensive disorders (%)				

Characteristics	PTB (N=518)	Very PTB (N=211)	Late PTB (N=307)	Control (N=512)
No	69.05	60.95	76.92	92.55
Yes	30.95	39.05	23.08	7.45
Histologic chorioamnionitis (%)				
No	73.59	58.29	84.85	85.74
Yes	26.41	41.71	15.15	14.26
Diabetes Mellitus (DM) (%)				
No	89.75	89.10	91.42	95.11
Yes	4.06	3.79	3.00	0.78
Gestational DM	6.19	7.11	5.58	4.11
Placenta abruption (%)				
No	91.89	85.78	97.01	99.21
Yes	8.11	14.22	2.99	0.79
Placenta Previa (%)				
No	95.75	94.31	97.44	99.21
Yes	4.25	5.69	2.56	0.79
Genetic characteristics				
African ancestry 0.88 (0.15); proportion (mean (sd); median)	0.88 (0.12); 0.92	0.87 (0.16); 0.92	0.85 (0.16); 0.92	0.89
Genotypes of *GSTT1* (%)				
Present	75.26	76.03	74.71	70.23
Absent	24.74	23.97	25.29	29.77
Genotypes of *CYP1A1* (%)				
AA	59.03	61.16	57.49	59.22
Aa/aa	40.97	38.84	42.51	40.78

Note:

*Based on self-reported Black race.

**PTB: preterm birth with gestational age <37 weeks; very PTB: preterm birth with gestational age <34 weeks; late PTB: preterm birth with gestational age 34.0–36.86 weeks.

[a]Only used non-US born subjects to calculate year in the US.

Table 2

Major risk factors of preterm birth (PTB), very PTB and late PTB* in Black mothers** at the Boston Medical Center.

	PTB			Very PTB			Late PTB		
	OR	**95% CI**	***p***	**OR**	**95% CI**	***p***	**OR**	**95% CI**	***p***
Epidemiologic Factors									
Smoking	1.3 3	1.08–1.64	0.007	1.2 8	0.98–1.66	0.07	1.3 8	1.09–1.75	0.007
Illicit drug use	1.4 8	0.97–2.25	0.07	1.6 0	0.95–2.68	0.08	1.4 1	0.88–2.27	0.15
Overall Stress	1.3 0	1.06–1.60	0.01	1.1 3	1.00–1.28	0.05	1.2 4	0.98–1.58	0.07
Genetic Ancestry									
Ancestry [a]	1.1 1	1.02–1.20	0.01	1.1 7	1.03–1.33	0.01	1.0 6	0.97–1.16	0.22
Gene-Environment Interactions									
CYP1A1 [b]	1.0 1	0.70–1.45	0.96	0.9 2	0.58–1.46	0.74	1.0 7	0.71–1.61	0.76
CYP1A1 x smoking	1.8 3	1.20–2.81	0.005	1.5 2	0.91–2.55	0.11	2.0 3	1.30–3.19	0.002
GSTT1 [c]	0.7 8	0.53–1.16	0.23	0.7 5	0.45–1.25	0.28	0.8 1	0.51–1.27	0.35
GSTT1 x smoking	1.2 8	0.81–2.03	0.28	1.5 6	0.93–2.61	0.09	1.0 6	0.61–1.86	0.83

Note:

* PTB: preterm birth with gestational age <37 weeks; very PTB: preterm birth with gestational age <34 weeks; late PTB: preterm birth with gestational age 34.0–36.86 weeks.

** Based on self-reported Black race.

[a] Ancestry is defined as African ancestry proportion and OR is expressed for each 10% increment of African ancestry.

[b] *CYP1A1* (AA) as the reference group.

[c] *GSTT1* (present) as the reference group.

Table 3

Evaluation of sequential predictive models of preterm birth (PTB)*, very PTB*, and late PTB* in Black mothers**, using receiver operating characteristic curve analysis/area under curve analysis.

Model	Covariates in the model	*c* statistic (95% CI)					
		PTB		Very PTB		Late PTB	
Model 1	Epidemiological variables only [a]	0.58	0.52–0.63	0.57	0.52–0.62	0.58	0.53–0.63
Model 2	Model 1 + ancestry	0.62	0.57–0.67	0.63	0.57–0.67	0.61	0.56–0.66
Model 3	Model 1+ *CYP1A1* + *CYP1A1** smoking + *GSTT1* + *GSTT1** smoking	0.61	0.56–0.66	0.61	0.56–0.66	0.61	0.56–0.66
Model 4	Model 1+ *CYP1A1* + *CYP1A1** smoking + *GSTT1* + *GSTT1** smoking only + ancestry	0.65	0.60–0.70	0.66	0.61–0.70	0.66	0.61–0.71

Note:

*PTB: preterm birth with gestational age <37 weeks; very PTB: preterm birth with gestational age <34 weeks; late PTB: preterm birth with gestational age 34.0–36.86 weeks.

**Based on self-reported Black race.

[a]Covariates included in model 1 are smoking, illicit drug use and overall stress.

Table 4

Area under curve for each predictive model shown in Table 3 and pairwise p-values for model comparisons of preterm birth (PTB)*, very PTB*, and late PTB* in Black mothers**

PTB				
	Model 1	**Model 2**	**Model 3**	**Model 4**
Model 1	0.58			
Model 2	0.09	0.62		
Model 3	0.09	0.78	0.61	
Model 4	0.002	0.05	0.02	0.65
Very PTB				
	Model 1	Model 2	Model 3	Model 4
Model 1	0.57			
Model 2	0.14	0.63		
Model 3	0.06	0.62	0.61	
Model 4	0.007	0.05	0.008	0.66
Late PTB				
	Model 1	Model 2	Model 3	Model 4
Model 1	0.58			
Model 2	0.08	0.61		
Model 3	0.21	0.78	0.61	
Model 4	0.004	0.05	0.02	0.66

Note:

*PTB: preterm birth with gestational age <37 weeks; very PTB: preterm birth with gestational age <34 weeks; late PTB: preterm birth with gestational age 34.0–36.86 weeks.

**Based on self-reported Black race.

[a]The values in the diagonal are the area under curve for each predictive model.

[b]The values in the off-diagonal are the *p*-values for testing the equality of the area under curve