

# Candidate gene linkage approach to identify DNA variants that predispose to preterm birth

Elise N.A. Bream<sup>1</sup>, Cara R. Leppellere<sup>1</sup>, Margaret E. Cooper<sup>2</sup>, John M. Dagle<sup>1</sup>, David C. Merrill<sup>3</sup>, Kaare Christensen<sup>4</sup>, Hyagriv N. Simhan<sup>5</sup>, Chin-To Fong<sup>6</sup>, Mikko Hallman<sup>7</sup>, Louis J. Muglia<sup>8</sup>, Mary L. Marazita<sup>2,9</sup> and Jeffrey C. Murray<sup>1</sup>

**BACKGROUND:** The aim of this study was to identify genetic variants contributing to preterm birth (PTB) using a linkage candidate gene approach.

**METHODS:** We studied 99 single-nucleotide polymorphisms (SNPs) for 33 genes in 257 families with PTBs segregating. Nonparametric and parametric analyses were used. Premature infants and mothers of premature infants were defined as affected cases in independent analyses.

**RESULTS:** Analyses with the infant as the case identified two genes with evidence of linkage: *CRHR1* ( $P = 0.0012$ ) and *CYP2E1* ( $P = 0.0011$ ). Analyses with the mother as the case identified four genes with evidence of linkage: *ENPP1* ( $P = 0.003$ ), *IGFBP3* ( $P = 0.006$ ), *DHCR7* ( $P = 0.009$ ), and *TRAF2* ( $P = 0.01$ ). DNA sequence analysis of the coding exons and splice sites for *CRHR1* and *TRAF2* identified no new likely etiologic variants.

**CONCLUSION:** These findings suggest the involvement of six genes acting through the infant and/or the mother in the etiology of PTB.

Preterm birth (PTB) is a major public health issue accounting for 3 million deaths worldwide each year. Despite a slight decrease in the incidence recently, PTB has increased from about 10% to more than 12.5% of births over the past two decades in the United States (1). Improvements in neonatal care have contributed to an increase in survival rates of preterm infants (2) in countries with optimal infant health-care delivery. However, despite these advances, PTB is still associated with substantial rates of morbidities, including chronic lung disease, patent ductus arteriosus, retinopathy of prematurity, intracranial hemorrhage, and cerebral palsy, especially in extremely preterm infants (3). These complications, in addition to PTB itself, are the largest risk factors for infant mortality in the United States (4).

The majority (72%) of PTBs are spontaneous (5) with unknown etiology (6). One substantial risk factor for PTB is genetic predisposition (7,8). An infant has an increased risk of

being premature if the mother was born prematurely (9); if a maternal aunt had a premature infant (10); and especially, if the mother had a prior PTB (11). Twin studies suggest that the heritability of PTB ranges from 15 to 40% (6). A few other maternal risk factors implicated in spontaneous PTB include low socioeconomic status, black race, younger age, intrauterine infection, inflammation (6), low prepregnancy weight (12), low cholesterol levels (13,14), and substance abuse (15,16). Conditions such as preeclampsia or fetal distress may lead to induction of labor or cesarean delivery before 37 wk gestation, resulting in an indicated PTB. Studies have shown that the mother transmits much of the genetic risk for spontaneous PTB, with smaller contributions from the father and fetus (17–19).

There are a variety of approaches to identifying genes associated with a complex trait. A candidate gene approach takes advantage of the known biology associated with labor and delivery whereas a genome-wide approach can implicate new physiologic pathways. In addition to known biology, conservation of evolutionary mechanisms can also be applied to human parturition timing to suggest additional genes (20). There are strong arguments for candidate gene studies to continue being used in the study of complex disease (21). Linkage studies have the ability to detect rare, higher-risk variants and can identify causal genes when allelic heterogeneity prevents genome-wide association from succeeding (22). We hypothesized that identification of new genes containing variants contributing to familial cases of PTB would be possible using a candidate gene linkage approach.

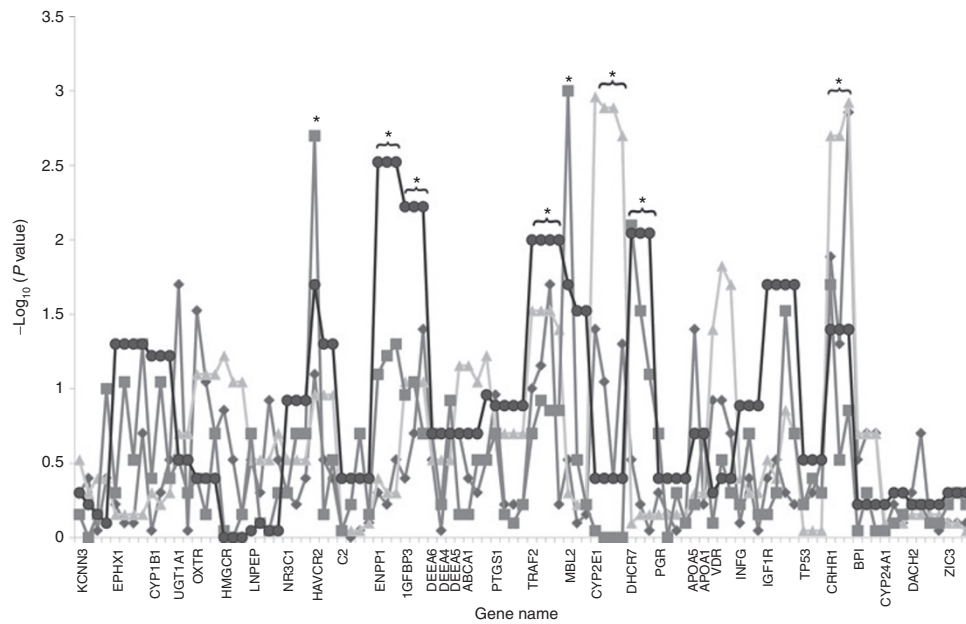
## RESULTS

A total of 257 extended families were chosen including 492 premature infants forming 297 affected-relative pairs (260 infant affected pairs and 37 mother affected pairs), see **Table 1** for a summary by study site. The mean family size for typed members was  $10.9 \pm 3.4$  (median = 10; range = 6–27), with a mean of  $2.0 \pm 0.9$  for typed premature infants (median = 2; range = 1–5) and a mean of  $1.2 \pm 0.4$  for typed mothers of premature infants

<sup>1</sup>Department of Pediatrics, University of Iowa, Iowa City, Iowa; <sup>2</sup>Department of Oral Biology and Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania; <sup>3</sup>Department of Obstetrics and Gynecology, Wake Forest University School of Medicine, Winston-Salem, North Carolina; <sup>4</sup>Department of Epidemiology, University of Southern Denmark, Odense, Denmark; <sup>5</sup>Department of Obstetrics and Gynecology, University of Pittsburgh, Pittsburgh, Pennsylvania; <sup>6</sup>Department of Pediatrics, University of Rochester School of Medicine and Dentistry, Rochester, New York; <sup>7</sup>Department of Pediatrics, University of Oulu, Oulu, Finland; <sup>8</sup>Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee; <sup>9</sup>Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania. Correspondence: Jeffrey C. Murray (jeff-murray@uiowa.edu)

**Table 1.** Number and type of participants by study site

	Iowa City, IA (n)	Pittsburgh, PA (n)	Rochester, NY (n)	Wake Forest, NC (n)	Denmark (n)	Total (n)
Samples	459	137	64	549	238	1,447
Extended families	91	28	13	94	31	257
Affected individuals	173	58	30	168	63	492
Infant affected-relative pairs	79	40	13	89	39	260
With sibling	61	34	9	72	7	183
With other relative	18	6	4	17	32	77
Mother affected-relative pairs	4	0	2	8	23	37



**Figure 1.** Nonparametric linkage (NPL) single-point (SP) and multipoint (MP) analysis results using either the infant or mother as the affected case are graphed by SNP centimorgan position starting with chromosome 1. Diamonds, infant NPL SP results; triangles, infant NPL MP results; squares, mother NPL SP results; circles, mother NPL MP results. \* $P < 0.01$ . SNP, single-nucleotide polymorphism.

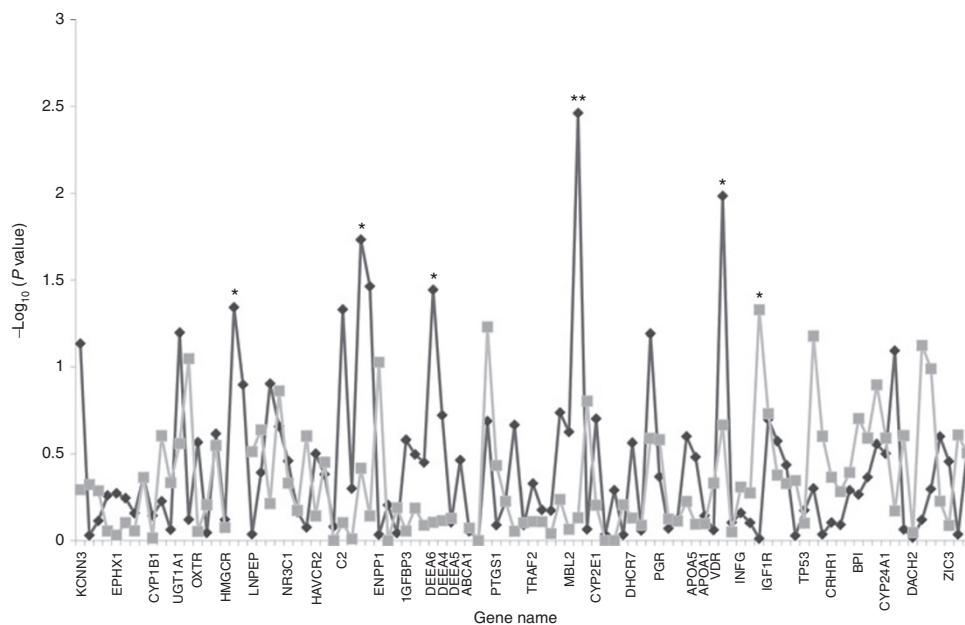
(median = 1; range = 1–3) per pedigree. An initial power analysis indicated that the sample size for this study was adequate to detect evidence of linkage with modest locus heterogeneity and Mendelian models. The data for all genes examined for either a fetal or maternal effect using nonparametric linkage analysis and a transmission disequilibrium test (TDT) are summarized in **Figures 1** and **2**, respectively. The entire data set along with parametric linkage results, which did not reveal any significant findings, is available in **Supplementary Tables S1** and **S2** online. Two single-nucleotide polymorphisms (SNPs)—rs1876831 (*CRHR1*,  $P = 5.0 \times 10^{-11}$ ) and rs573549 (*APOA1*,  $P = 0.0004$ )—violated Hardy-Weinberg equilibrium and were not included in the analysis.

With the preterm infant as the affected case, a nonparametric linkage analysis of all candidate genes revealed two linkage peaks. These multipoint linkage peaks were on chromosome 10 (*CYP2E1*,  $P = 0.0011$ – $0.002$ ) and chromosome 17 (*CRHR1*,  $P = 0.0012$ – $0.002$ ). *CRHR1* also had significant single-point linkage peaks ( $P = 0.05$ – $0.0014$ ). With the mother of a premature infant as the affected case, seven linkage peaks were identified. Four were multipoint linkage peaks on chromosome 6

(*ENPPI*,  $P = 0.003$ ), chromosome 7 (*IGFBP3*,  $P = 0.006$ ), chromosome 9 (*TRAF2*,  $P = 0.01$ ), and chromosome 11 (*DHCR7*,  $P = 0.009$ ). Three were single-point linkage peaks on chromosome 5 (*HAVCR2*, rs12654265,  $P = 0.002$ ), chromosome 10 (*MBL2*, rs2136892,  $P = 0.001$ ), and chromosome 11 (*DHCR7*, rs1790318,  $P = 0.008$ ).

Using TDT, we identified eight associated SNPs that are nominally significant ( $P < 0.05$ ). However, none fell within the most significant linkage peaks based on affection status, and none were significant when accounting for multiple comparisons using a Bonferroni correction. With the mother as case, there was one suggested association with rs10878774 in *INFG* ( $P = 0.047$ ). With the infant as case, a suggested association was seen for rs2303152 in *HMGCR* ( $P = 0.046$ ); rs605203 ( $P = 0.047$ ), rs7746553 ( $P = 0.019$ ), and rs592229 ( $P = 0.034$ ) in *C2*; rs44589901 in *DEFA6* ( $P = 0.036$ ); rs11003136 in *MBL2* ( $P = 0.003$ ); and rs4760648 in *VDR* ( $P = 0.010$ ).

An analysis was performed that stratified premature individuals based on type of labor. There were 251 individuals with spontaneous, 40 with induced, 36 with no labor (cesarean section), and all others with unknown type of labor. The reason for induction



**Figure 2.** Transmission disequilibrium test (TDT) analysis results using either the infant or the mother as the affected case are graphed by SNP centimorgan position starting with chromosome 1. Diamonds, infant results; squares, mother results. \* $P < 0.05$ ; \*\* $P < 0.01$ . SNP, single-nucleotide polymorphism.

was not known for all individuals in the data set. There was no significant difference in mean ( $P = 0.07$ ) or median ( $P = 0.19$ ) gestational age between those with spontaneous, induced, and no labor. When the unknown group was added, a statistically significant difference was seen as compared with the no labor group (ANOVA mean  $P = 0.02$ , median  $P = 0.01$ ), but no difference was seen in comparison with the spontaneous and induced groups. Using the preterm infant with spontaneous labor as the affected case, a nonparametric linkage analysis revealed two multipoint peaks and six single-point peaks. The multipoint peaks were on chromosome 9 (*TRAF2*,  $P = 0.03$ ) and chromosome 20 (*BPI*,  $P = 0.03$ ). The single-point peaks were in *HMGCR* (rs3931914,  $P = 0.04$ ), *PTGS1* (rs10513401,  $P = 0.03$ ), *TRAF2* (rs10781522,  $P = 0.04$ ), *CRHR1* (rs7225082,  $P = 0.014$ ), and *BPI* (rs5743507,  $P = 0.04$ , and rs4358188,  $P = 0.04$ ). TDT identified eight suggested associations: rs7746553 in *C2* ( $P = 0.02$ ), rs4458901 in *DEFA6* ( $P = 0.003$ ), rs2515617 in *ABCA1* ( $P = 0.04$ ), rs10781522 ( $P = 0.04$ ) and rs4880166 ( $P = 0.04$ ) in *TRAF2*, rs11003136 in *MBL2* ( $P = 0.02$ ), rs1630498 in *DHCR7* ( $P = 0.01$ ), and rs1893505 in *PGR* ( $P = 0.02$ ).

Linkage haplotype analysis was performed for *CRHR1* because it had the strongest linkage signal. The three SNPs in *CRHR1* generated seven of a possible eight haplotypes. These genotypes were treated as “super alleles” numbered 1 through 8. Because haplotypes 6 and 7 had low frequencies, they were pooled and redefined. Both parametric and nonparametric analyses were performed (nonparametric  $P = 0.0024$ ). The dominant model had a log of odds score of 0.82 and heterogeneity log of odds score of 1.32 ( $\alpha = 0.630$ ). The recessive model was not significant (log of odds =  $-18.1$ , heterogeneity log of odds score = 0.329 with  $\alpha = 0.121$ ). With TDT haplotype association analysis, no individual haplotype was significant ( $P > 0.28$ ), with the global  $P = 0.83$ .

An initial analysis of 33 SNPs in *OXTR*, *PGR*, *VDR*, *CRHR1*, *PTGS1*, *KCNN3*, *TRAF2*, *IGF1R*, and *NR3C1* was performed on a subset of the population (412 premature infants forming 230 affected-relative pairs) in which *CRHR1* and *TRAF2* showed evidence of linkage ( $P = < 0.01$ ). Because of strong evidence in the literature to support a causal genetic variant within these genes, we sequenced their coding regions rather than the 10–20 centimorgan chromosomal region surrounding the identified linkage peaks. Although previously reported SNPs were identified in the coding regions, no novel missense, frame-shift, or nonsense mutations were detected.

## DISCUSSION

Identification of a genetic contribution to PTB would allow detection of at-risk pregnancies and might also suggest environmental contributors to PTB. This could provide tools to prolong gestation by tailoring obstetrical management to individual genetic susceptibilities. In the past, interventions for preventing PTB have proven largely unsuccessful (23), but by identifying specific individual pathophysiologic mechanisms of PTB, new strategies can be developed. In this study, we used a linkage candidate gene and sequencing approach in an attempt to identify chromosomal regions that may contain genes involved in the etiology of PTB. Recently, genome-wide association studies have had enormous success in identifying genes associated with complex traits, but they have not been reported for PTB. In addition, association will not identify those genes for which allelic heterogeneity is responsible for the heritability even though those alleles might have greater impact on the phenotype in a given family than common variants typically have. Linkage is the best approach to detecting this class of variant. Although previous studies have supported a stronger maternal contribution

to PTB, it is also thought that PTB may be due to the role of genes present in the mother/uterus, baby/placenta, or a combination of both (8,24). This is the first large linkage study focusing on both the mother and the fetus as potential risk cases, so the potential linkages identified will be signals to be examined in larger studies using more markers and a greater number of families than were available in this study. Our findings suggest the involvement of *CRHR1* or *CYP2E1* mediated by the infant and/or *ENPP1*, *IGFBP3*, *DHCR7*, or *TRAF2* mediated by the mother in the etiology of PTB. *CRHR1*, *DHCR7*, and *TRAF2*, in particular, are members of pathways identified in previous studies and have biologic plausibility for playing a role in PTB.

*CRHR1* encodes one of the two receptors found in humans to which corticotropin-releasing hormone (CRH) binds (25). *CRHR1* is expressed in the pituitary (26), endometrium (27), myometrium, and placenta (25), among other locations. The coding sequence of *CRHR1* is highly conserved, with only six missense variants reported in the National Heart, Lung, and Blood Institute/Grand Opportunity Exome Sequencing Project database (<http://evs.gs.washington.edu/EVS/>) of over 1,000 sequenced Europeans. Placental CRH is part of a feed-forward loop in both mother and fetus. CRH stimulates the release of adrenocorticotrophic hormone from the pituitary, which leads to the release of glucocorticoids from the adrenal glands, promoting production of more CRH (26,27). Plasma CRH levels undergo an exponential increase during pregnancy, peaking at the time of delivery (26) due to increased placental production and decreased CRH binding protein concentrations (27). Women having a PTB undergo a more rapid increase (26), establishing different patterns of CRH levels as early as the end of the first trimester, suggesting that the length of gestation is predetermined and the onset of parturition is triggered when CRH levels peak (26). Genetic variants in *CRHR1* have also been shown to have an association with susceptibility to bacterial vaginosis, which is a risk factor for PTB (28).

7-Dehydrocholesterol reductase catalyzes the final step in the synthesis of cholesterol. Cholesterol is an important substrate in the synthesis of many hormones, including placental progesterone (13), which is critical for successful reproduction. A physiologic hypercholesterolemia has been demonstrated to occur later in pregnancy and is thought to be a mechanism for pregnancy maintenance (29). A study by Steffen *et al.* showed fetal polymorphisms in *DHCR7*, as well as other genes involved in cholesterol metabolism, to be associated with birth weight and PTB. In addition, a strong association was seen between low total cholesterol levels during pregnancy in Caucasian women and PTB (13).

*TRAF2* plays a role in the tumor necrosis factor (TNF) signal transduction pathway. In this pathway, TNF binds its receptor to recruit caspase 8, initiating apoptosis (30). Activation of this pathway via binding to TNF receptor 1 on fetal membranes has been implicated in the etiology of premature rupture of membranes (31). TNF also activates the transcription factor nuclear factor  $\kappa$ B, which interacts with inhibitor-of-apoptosis proteins to block caspase 8 (32). TNF binding to TNF receptor 2 on

fetal membranes activates nuclear factor  $\kappa$ B via TNF receptor-associated factor 2 (TRAF2), leading to increased production of inflammatory cytokines. This subsequently increases production of prostaglandins, which can initiate preterm labor (31).

There is less evidence to support *CYP2E1*, *ENPP1*, and *IGFBP3* in the etiology of PTB, but they may play a role based on known maternal risk factors. Members of the cytochrome P450 family are involved in the detoxification and metabolism of a variety of substrates as well as synthesis of cholesterol, steroids, and other lipids. *CYP2E1*, specifically, encodes a protein that is induced by ethanol and pathologic states such as fasting, diabetes, obesity, and high-fat diet (33). It also metabolizes specific substrates including ethanol and nitrosamines (34), pre-mutagens found in cigarette smoke. *ENPP1* encodes a protein responsible for cleaving pyrophosphate and phosphodiester bonds of nucleotides and nucleotide sugars, which are a source of chemical energy and play an important role in metabolism. Phosphate removal can interrupt the activity of nucleotides, resulting in deranged metabolism. *IGFBP3* encodes a protein that binds the majority of circulating insulin-like growth factors, which are thought to play a role in fetal and postnatal growth (35). Increases in maternal serum levels have been associated with increasing gestational age (36), and decreased levels have been shown to be present in deliveries before 32 wk gestation (35). An increase in inflammatory cytokines has been shown to decrease levels of insulin-like growth factor-binding protein 3 (IGFBP3) (37), and inflammation is a well-known pathway implicated in PTB (6).

For this study, the number of affected-relative pairs was limiting, particularly for mothers of premature infants. Additional families could provide power to make a genome-wide linkage analysis practical. A second limitation was that roughly one-third of our cohort had an unknown type of labor. Therefore, even though a separate analysis was performed looking at spontaneous PTB only, the results should be interpreted keeping in mind that the sample size as well as the power of the study were both significantly reduced. In addition, there were only 40 infants with known induced PTB, and the reason for augmentation was not known for all these infants, thereby preventing us from performing a separate analysis of these individuals. For future studies, it would be important to recruit families for which the type of labor and reasons of augmentation, if applicable, are known in order to have a more informative cohort to use for stratified analyses.

Additional sequencing could better characterize the significant linkage peaks identified in this study. It would be important to look at regulatory elements for *CRHR1* and *TRAF2* because only coding regions were examined. In addition, the coding regions and regulatory elements for the other genes with significant linkage peaks (*CYP2E1*, *ENPP1*, *IGFBP3*, and *DHCR7*) should be sequenced. An alternative approach would be to saturate the chromosomal regions surrounding these genes with additional markers and include additional samples for a fine-mapping genetic association study with increased power. Whole-exome sequencing using familial cases may also provide valuable insight. Once we have further defined genetic variants and likely

environmental contributors, analyses can be performed to look at the interactions and to adjust for confounding variables.

In summary, we have identified several candidate genes/regions that may harbor rare variants contributing to PTB, with one, *CRHR1*, having the strongest data and in which the effect is modulated via the fetus.

## METHODS

### DNA Sample Collection

Cases were defined as singleton preterm infants (delivery at <37 completed weeks of gestation) admitted to one of our centers in Iowa City, IA; Pittsburgh, PA; Rochester, NY; Wake Forest, NC; or the island of Funen in Denmark. We included both indicated and spontaneous deliveries. Gestational age was estimated by the first day of the last menstrual period and confirmed by

obstetrical examination, including ultrasound when indicated. Signed informed consent, approved by the institutional review board (no. 199911068) at the University of Iowa, was obtained from all families. DNA was extracted from cord blood or buccal swabs collected for the infants and venous blood, saliva samples, or buccal swabs collected for relatives. Demographic information and additional phenotype data were collected through an interview with the mother and medical chart review.

### Family Selection

Families were included if samples were available for a minimum of two premature individuals or mothers of premature infants, excluding multiples, or one premature infant with at least one full-term sibling or cousin. An infant affected-relative pair was defined as any pair of premature individuals, excluding infant/parent pairs, within a family. A mother affected-relative pair was defined as a pair of sisters both having premature infants.

**Table 2.** List of genes and single-nucleotide polymorphisms studied

Gene	Chromosome	Reference sequence numbers			
<i>KCNN3</i>	1	rs10128027	rs1051614	rs11584635	rs12058931
<i>EPHX1</i>	1	rs4559463	rs1051740	rs2740170	rs1009668
<i>CYP1B1</i>	2	rs163076	rs1056836	rs4670813	
<i>UGT1A1</i>	2	rs4663333	rs10203852		
<i>OXTR</i>	3	rs237887	rs4686301	rs237897	
<i>HMGCR</i>	5	rs3931914	rs2303152	rs5908	
<i>LNPEP</i>	5	rs4869315	rs3849749	rs4869317	rs13175726
<i>NR3C1</i>	5	rs258747	rs4128753	rs11167814	
<i>HAVCR2</i>	5	rs12654265	rs13170556	rs1036199	
<i>C2</i>	6	rs605203	rs2734335	rs7746553	rs592229
<i>ENPP1</i>	6	rs943003	rs1409181	rs7754561	
<i>IGFBP3</i>	7	rs2453839	rs3110697	rs2132571	
<i>DEFA6</i>	8	rs4458901			
<i>DEFA4</i>	8	rs2239668			
<i>DEFA5</i>	8	rs10095331			
<i>ABCA1</i>	9	rs2066716	rs4149313	rs2000069	rs2515617
<i>PTGS1</i>	9	rs10513401	rs5788	rs4240474	rs2778624
<i>TRAF2</i>	9	rs2811761	rs10781522	rs3750512	rs4880166
<i>MBL2</i>	10	rs2136892	rs11003136	rs11003231	
<i>CYP2E1</i>	10	rs9418990	rs2070673	rs2249695	rs1536826
<i>DHCR7</i>	11	rs1790318	rs1792325	rs1630498	
<i>PGR</i>	11	rs1042839	rs503362	rs1942836	rs1893505
<i>APOA5</i>	11	rs2071522			
<i>APOA1</i>	11	rs1729410			
<i>VDR</i>	12	rs2239182	rs4760648	rs4516035	
<i>INFG</i>	12	rs2098395	rs2069718	rs10878774	
<i>IGF1R</i>	15	rs1521480	rs4966036	rs4966038	rs2684811
<i>TP53</i>	17	rs8073498	rs1042522	rs3744262	
<i>CRHR1</i>	17	rs7225082	rs4458044	rs173365	
<i>BPI</i>	20	rs1341023	rs5743507	rs4358188	
<i>CYP24A1</i>	20	rs927650	rs2762941	rs2296241	
<i>DACH2</i>	X	rs6617219	rs5968924	rs12164321	rs4828153
<i>ZIC3</i>	X	rs2859253	rs2746112	rs1085654	

## Genotyping

Candidate genes were selected based on biologic plausibility, a review of current literature (14,20,26,31,34–36,38–40), and previous association study findings from our lab. The group included 8 genes (*BPI*, *C2*, *DEFA6*, *DEFA4*, *DEFA5*, *INFG*, *MBL2*, and *TRAF2*) in inflammatory pathways, 12 genes involved in hormonal regulation (*ABCA1*, *APOA1*, *APOA5*, *CRHR1*, *CYP1B1*, *DHCR7*, *HMGCR*, *LNPEP*, *NR3C1*, *OXTR*, *PGR*, and *PTGSI*), and 13 other genes (*CYP24A1*, *CYP2E1*, *DACH2*, *ENPP1*, *EPHX1*, *HAVCR2*, *IGF1R*, *IGFBP3*, *KCNN3*, *TP53*, *UGT1A1*, *VDR*, and *ZIC3*). A list of all 33 genes and their respective SNPs—99 of 101 are reported after excluding 2 not in Hardy–Weinberg equilibrium—are shown in [Table 2](#).

Genotyping of SNP markers was performed using Applied Biosystems (Foster City, CA) TaqMan chemistry. Within each gene, two to four on-demand SNP genotyping assays were chosen based on linkage disequilibrium data available from the International Hapmap project (<http://www.hapmap.org>) as well as for their haplotype characteristics, such as high heterozygosity and low correlation coefficient, to maximize heterozygosity. The average heterozygosity per locus was 0.85. Applied Biosystems provided standard conditions under which the reactions were run. Thermocycling was performed with conditions of 95°C for 10 min followed by 50 cycles alternating between 92°C for 15 s and 60°C for 1 min. Allele determination was done in the end point analysis mode on an Applied Biosystems 7900 HT Sequence Detection System machine with SDS 2.3 software (Applied Biosystems). Mendelian errors were checked, and individuals with >10% error rates were excluded from analyses. Genotypes were entered into Progeny (South Bend, IN), a laboratory database, and files were generated in linkage format for analysis.

## Data Analysis

Two different phenotypic outcomes, premature individuals and mothers of premature individuals, were defined in independent analyses. Single-point and multipoint nonparametric and parametric linkage analyses, as well as linkage haplotype analysis, were performed using the Merlin 1.1.2 software package (<http://www.sph.umich.edu/csg/abecasis/Merlin/index.html>). Two parametric linkage analysis models, autosomal recessive and autosomal dominant, were used assuming a disease allele frequency of 0.11 for both. We used penetrances of 0.80, 0.02, and 0.02 for the recessive model and 0.20, 0.20, and 0.02 for the dominant model for the wild-type homozygotes, heterozygotes, and homozygotes, respectively. The values selected were based on the rates of PTB in US Caucasian populations during the time frame of this study and the penetrances as arbitrary, but midrange, choices based on other complex trait models. Changes in penetrances did not greatly affect results.

Association testing was performed on the families with DNA samples available to form case–parent triads using the preterm infant or the mother of a preterm infant as case. We used the family-based association test (<http://www.biostat.harvard.edu/~fbat/fbat.htm>), a family-based TDT, to look for nonrandom allele transmission from parents to offspring.

## Sequencing

Primers were designed from public sequence to amplify the coding regions of *CRHR1* and *TRAF2* and are available on request. We sequenced 190 preterm infants from the linkage cohort and 105 mothers of preterm infants from a population-matched cohort in Helsinki, Finland. In addition, 162 parents of term infants from Iowa City, 29 Centre d'Étude du Polymorphisme Humain parents, and 85 mothers of term infants from Helsinki, Finland, were used as controls. PCR products were sent to Functional Biosciences (Madison, WI) for sequencing. Chromatograms were transferred to a UNIX workstation, base called with PHRED (v.0.961028), assembled with PHRAP (v. 0.960731), scanned by POLYPHRED (v. 0.970312), and viewed with CONSED (v. 4.0) (University of Washington, Seattle, WA).

## STATEMENT OF FINANCIAL SUPPORT

This work was supported by a grant from the Doris Duke Charitable Foundation to the University of Iowa to fund clinical research fellow E.N.A.B., as well as funding from the National Institutes of Health (grants RO1 HD-052953 and RO1 HD-057192-01A2) and March of Dimes (grants MOD 21-FY10-180 and MOD 6-FY11-261).

## ACKNOWLEDGMENTS

The authors thank all the families participating in the study as well as the staff responsible for obtaining consent and samples from participants, in particular, Laura Knosp, Sarah Wente, Gretchen Cress, Ruthann Schrock, Sara Scott, and Nancy Krutzfeld at the University of Iowa; Karen Debes at the Magee Womens Hospital; Lauren Smith and Eileen Blakely at the University of Rochester; Kristi Lanier at the Wake Forest University; and Dorte Grosen and Bibi Anshoj at the University of Southern Denmark. Susan Berends helped to identify extended families and Dina Ahram provided student support in the lab.

## SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/pr>

## REFERENCES

- Martin JA, Osterman MJ, Sutton PD. Are preterm births on the decline in the United States? Recent data from the National Vital Statistics System. NCHS Data Brief 2010;1–8.
- Fellman V, Hellstrom-Westas L, Norman M, et al. One-year survival of extremely preterm infants after active perinatal care in Sweden. JAMA 2009;301:2225–33.
- Austeng D, Blennow M, Ewald U, et al. Incidence of and risk factors for neonatal morbidity after active perinatal care: extremely preterm infants from Sweden (EXPRESS). Acta Paediatr 2010;99:978–92.
- Callaghan WM, MacDorman MF, Rasmussen SA, Qin C, Lackritz EM. The contribution of preterm birth to infant mortality rates in the United States. Pediatrics 2006;118:1566–73.
- Mercer BM, Goldenberg RL, Das A, et al. The preterm prediction study: a clinical risk assessment system. Am J Obstet Gynecol 1996;174:1885–93; discussion 1893–5.
- Muglia LJ, Katz M. The enigma of spontaneous preterm birth. N Engl J Med 2010;362:529–35.
- Varner MW, Esplin MS. Current understanding of genetic factors in preterm birth. BJOG 2005;112:Suppl 1:28–31.
- Menon R, Fortunato SJ, Thorsen P, Williams S. Genetic associations in preterm birth: a primer of marker selection, study design, and data analysis. J Soc Gynecol Invest 2006;13:531–41.
- Porter TF, Fraser AM, Hunter CY, Ward RH, Varner MW. The risk of preterm birth across generations. Obstet Gynecol 1997;90:63–7.
- Winkvist A, Mogren I, Högberg U. Familial patterns in birth characteristics: impact on individual and population risks. Int J Epidemiol 1998;27:248–54.
- Mercer BM, Goldenberg RL, Moawad AH, et al. The preterm prediction study: effect of gestational age and cause of preterm birth on subsequent obstetric outcome. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. Am J Obstet Gynecol 1999;181(5 Pt 1):1216–21.
- Luke B. The evidence linking maternal nutrition and prematurity. J Perinat Med 2005;33:500–5.
- Edison RJ, Berg K, Remaley A, et al. Adverse birth outcome among mothers with low serum cholesterol. Pediatrics 2007;120:723–33.
- Steffen KM, Cooper ME, Shi M, et al. Maternal and fetal variation in genes of cholesterol metabolism is associated with preterm delivery. J Perinatol 2007;27:672–80.
- Shah NR, Bracken MB. A systematic review and meta-analysis of prospective studies on the association between maternal cigarette smoking and preterm delivery. Am J Obstet Gynecol 2000;182:465–72.
- O'Leary CM, Nassar N, Kurinczuk JJ, Bower C. The effect of maternal alcohol consumption on fetal growth and preterm birth. BJOG 2009;116:390–400.
- Wilcox AJ, Skjaerven R, Lie RT. Familial patterns of preterm delivery: maternal and fetal contributions. Am J Epidemiol 2008;167:474–9.
- Boyd HA, Poulsen G, Wohlfahrt J, Murray JC, Feenstra B, Melbye M. Maternal contributions to preterm delivery. Am J Epidemiol 2009;170:1358–64.

19. Plunkett J, Feitosa MF, Trusgnich M, et al. Mother's genome or maternally-inherited genes acting in the fetus influence gestational age in familial preterm birth. *Hum Hered* 2009;68:209–19.
20. Plunkett J, Doniger S, Orabona G, et al. An evolutionary genomic approach to identify genes involved in human birth timing. *PLoS Genet* 2011;7:e1001365.
21. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008;322:881–8.
22. Savitz DA. Invited commentary: disaggregating preterm birth to determine etiology. *Am J Epidemiol* 2008;168:990–2; discussion 993–4.
23. Slattery MM, Morrison JJ. Preterm delivery. *Lancet* 2002;360:1489–97.
24. Cotten CM, Ginsburg GS, Goldberg RN, Speer MC. Genomic analyses: a neonatology perspective. *J Pediatr* 2006;148:720–6.
25. Grammatopoulos DK. Placental corticotrophin-releasing hormone and its receptors in human pregnancy and labour: still a scientific enigma. *J Neuroendocrinol* 2008;20:432–8.
26. Smith R. Parturition. *N Engl J Med* 2007;356:271–83.
27. Kalantaridou SN, Zoumakis E, Makrigiannakis A, Lavasidis LG, Vrekoussis T, Chrousos GP. Corticotropin-releasing hormone, stress and human reproduction: an update. *J Reprod Immunol* 2010;85:33–9.
28. Ryckman KK, Simhan HN, Krohn MA, Williams SM. Predicting risk of bacterial vaginosis: the role of race, smoking and corticotropin-releasing hormone-related genes. *Mol Hum Reprod* 2009;15:131–7.
29. Tranquilli AL, Cester N, Giannubilo SR, Corradetti A, Nanetti L, Mazzanti L. Plasma lipids and physicochemical properties of the erythrocyte plasma membrane throughout pregnancy. *Acta Obstet Gynecol Scand* 2004;83:443–8.
30. Chen G, Goeddel DV. TNF-R1 signaling: a beautiful pathway. *Science* 2002;296:1634–5.
31. Menon R, Fortunato SJ. Fetal membrane inflammatory cytokines: a switching mechanism between the preterm premature rupture of the membranes and preterm labor pathways. *J Perinat Med* 2004;32:391–9.
32. Wang CY, Mayo MW, Korneluk RG, Goeddel DV, Baldwin AS Jr. NF-kappaB antiapoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation. *Science* 1998;281:1680–3.
33. Kessova I, Cederbaum AI. CYP2E1: biochemistry, toxicology, regulation and function in ethanol-induced liver injury. *Curr Mol Med* 2003;3:509–18.
34. Neafsey P, Ginsberg G, Hattis D, Johns DO, Guyton KZ, Sonawane B. Genetic polymorphism in CYP2E1: population distribution of CYP2E1 activity. *J Toxicol Environ Health B Crit Rev* 2009;12:362–88.
35. Cooley SM, Donnelly JC, Collins C, Geary MP, Rodeck CH, Hindmarsh PC. The relationship between maternal insulin-like growth factors 1 and 2 (IGF-1, IGF-2) and IGFBP-3 to gestational age and preterm delivery. *J Perinat Med* 2010;38:255–9.
36. Langford KS, Nicolaidis KH, Jones J, Abbas A, McGregor AM, Miell JP. Serum insulin-like growth factor-binding protein-3 (IGFBP-3) levels and IGFBP-3 protease activity in normal, abnormal, and multiple human pregnancy. *J Clin Endocrinol Metab* 1995;80:21–7.
37. De Benedetti F, Meazza C, Oliveri M, et al. Effect of IL-6 on IGF binding protein-3: a study in IL-6 transgenic mice and in patients with systemic juvenile idiopathic arthritis. *Endocrinology* 2001;142:4818–26.
38. Ehn NL, Cooper ME, Orr K, et al. Evaluation of fetal and maternal genetic variation in the progesterone receptor gene for contributions to preterm birth. *Pediatr Res* 2007;62:630–5.
39. Hirota Y, Daikoku T, Tranguch S, Xie H, Bradshaw HB, Dey SK. Uterine-specific p53 deficiency confers premature uterine senescence and promotes preterm birth in mice. *J Clin Invest* 2010;120:803–15.
40. Haataja R, Karjalainen MK, Luukkonen A, et al. Mapping a new spontaneous preterm birth susceptibility gene, IGF1R, using linkage, haplotype sharing, and association analysis. *PLoS Genet* 2011;7:e1001293.