Integrated Genomic Analyses in Bronchopulmonary Dysplasia

Namasivayam Ambalavanan, MD[1], C. Michael Cotten, MD[2], Grier P. Page, PhD[3], Waldemar A. Carlo, MD[1], Jeffrey C. Murray, MD[4], Soumyaroop Bhattacharya, MS, MEd[5], Thomas J. Mariani, PhD[5], Alain C. Cuna, MD[6], Ona M. Faye-Petersen, MD[7], David Kelly, MD[7], and Rosemary D. Higgins, MD[8], on behalf of the Genomics and Cytokine Subcommittees of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Neonatal Research Network*

Objective To identify single-nucleotide polymorphisms (SNPs) and pathways associated with bronchopulmonary dysplasia (BPD) because O_2 requirement at 36 weeks' postmenstrual age risk is strongly influenced by heritable factors.

Study design A genome-wide scan was conducted on 1.2 million genotyped SNPs, and an additional 7 million imputed SNPs, using a DNA repository of extremely low birth weight infants. Genome-wide association and gene set analysis was performed for BPD or death, severe BPD or death, and severe BPD in survivors. Specific targets were validated via the use of gene expression in BPD lung tissue and in mouse models.

Results Of 751 infants analyzed, 428 developed BPD or died. No SNPs achieved genome-wide significance ($P < 10^{-8}$), although multiple SNPs in adenosine deaminase, CD44, and other genes were just below $P < 10^{-6}$. Of approximately 8000 pathways, 75 were significant at false discovery rate (FDR) <0.1 and $P < .001$ for BPD/death, 95 for severe BPD/death, and 90 for severe BPD in survivors. The pathway with lowest FDR was miR-219 targets ($P = 1.41E-08$, FDR 9.5E-05) for BPD/death and phosphorous oxygen lyase activity (includes adenylate and guanylate cyclases) for both severe BPD/death ($P = 5.68E-08$, FDR 0.00019) and severe BPD in survivors ($P = 3.91E-08$, FDR 0.00013). Gene expression analysis confirmed significantly increased miR-219 and CD44 in BPD.

Conclusions Pathway analyses confirmed involvement of known pathways of lung development and repair (CD44, phosphorus oxygen lyase activity) and indicated novel molecules and pathways (adenosine deaminase, targets of miR-219) involved in genetic predisposition to BPD. *(J Pediatr 2015;166:531-37)*.

See editorial, p 521

Bronchopulmonary dysplasia (BPD) is common in extremely preterm infants, and genetic factors may account for much of the variance in risk for BPD.[1] Targeted candidate gene analyses suggest single-nucleotide polymorphisms (SNPs) in certain cytokines, surfactant proteins, and related molecules[2,3] but not others[4] are associated with BPD. Hadchouel et al[5] identified the *SPOCK2* gene as associated with BPD in a genome-wide association study (GWAS) that evaluated the entire genome in an unbiased manner. However, Wang et al[6] did not find SNPs associated with BPD in a GWAS.

Most complex diseases (such as BPD) involve gene-environment interactions and interactions among different loci. However, conventional single marker analysis does not explicitly look for interactions among different genes in the same biological pathway that have a multiplicative or a threshold effect.[7] Most GWAS that focus on analysis of single markers lack the power to identify the small contribution of most genetic variants.[8] Pathway-based approaches, which consider multiple contributing factors in the same biological pathway, complement the single-marker approach and provide understanding of GWAS data in many diseases.[9]

ADARB2	Adenosine deaminase
BPD	Bronchopulmonary dysplasia
FDR	False discovery rate
GWAS	Genome-wide association study
SNP	Single-nucleotide polymorphism

From the [1]Department of Pediatrics, University of Alabama at Birmingham, Birmingham, AL; [2]Department of Pediatrics, Duke University, Durham, NC; [3]RTI International, Atlanta, GA; [4]Department of Pediatrics, University of Iowa, Iowa City, IA; [5]Division of Neonatology and Program in Pediatric Molecular and Personalized Medicine, University of Rochester, Rochester, NY; [6]Department of Pediatrics, University of Missouri-Kansas City School of Medicine, Kansas City, MO; [7]Department of Pathology, University of Alabama at Birmingham, Birmingham, AL; and [8]*Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD

*List of members of the Genomics and Cytokine Subcommittees of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Neonatal Research Network is available at www.jpeds.com (**Appendix**).

Supported by the National Institutes of Health (General Clinical Research Center M01 RR30, M01 RR32, M01 RR39, M01 RR70, M01 RR80, M01 RR633, M01 RR750, M01 RR997, M01 RR6022, M01 RR7122, M01 RR8084, M01 RR16587, UL1 RR24979) and the *Eunice Kennedy Shriver* NICHD (U01 HD36790, U10 HD21364, U10 HD21373, U10 HD21385, U10 HD21397, U10 HD21415, U10 HD27851, U10 HD27853, U10 HD27856, U10 HD27871, U10 HD27880, U10 HD27881, U10 HD27904, U10 HD34216, U10 HD40461, U10 HD40492, U10 HD40498, U10 HD40689, U10 HD53109). J.M. received assistance for the GENEVA study from the National Human Genome Research Institute (U01 HG4423). Data collected at participating sites of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Neonatal Research Network were transmitted to RTI International, the data coordinating center for the network, which stored, managed, and analyzed the data for this study. The authors declare no conflicts of interest.

http://dx.doi.org/10.1016/j.jpeds.2014.09.052

In this study, we used a GWAS combined with pathway-based approaches to increase our understanding of the role of genetics in BPD susceptibility and integrated these results with gene expression comparing BPD with control subjects and a newborn mouse model of hyperoxia exposure simulating BPD. We hypothesized that SNPs in biological pathways involved in lung development and injury will be enriched in infants who develop BPD or die. The combined outcome of BPD or death was used because death is a competing outcome for BPD, ie, infants who die early cannot develop BPD, even though they may be at the greatest risk of BPD.

Methods

Patients included were a subset of infants enrolled in the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Neonatal Research Network's Cytokines study that enrolled infants 401-1000 g at birth, <72 hours' old, and free of major congenital anomalies.[10] The study was approved by institutional review boards at participating centers, and written informed consent was obtained from parent(s). Additional review by the institutional review board allowed the GWAS genotyping results with a limited phenotype data to be included in the National Human Genome Research Institute Database of Genotypes and Phenotypes.

DNA was extracted from the earliest age blood spot collected on filter paper. Whole-genome amplification was used for samples that did not provide adequate genomic DNA. Genotyping was performed on the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, California). BPD was defined by supplemental O_2 at 36 weeks' postmenstrual age. Severe BPD was defined as therapy with O_2 >21% for at least 28 days plus the use of ≥30% O_2 and/or positive pressure (ventilation or nasal continuous positive airway pressure) at 36 weeks' postmenstrual age.[11] Death was defined as in-hospital death before discharge.

Ancestry was classified as black (African-American), white (non-Hispanic Caucasians), Hispanic (Hispanic Caucasian), and others, including Asian and multiracial, using GWAS-Tools[12] to generate eigenvalues for the entire dataset. Imputation was run using beagle 3.3.1. A total of 769 757 SNPs were used for imputation with 7 500 443 SNPs being imputed.[13]

Analysis of SNPs was performed via 2 complementary methods: a standard GWAS analysis followed by a pathway analysis. SNPs were analyzed using PLINK[14] using logistic regression under an additive model. Three models were run: BPD or death vs survival without BPD, severe BPD or death vs survival without severe BPD, and severe BPD in survivors vs survivors without severe BPD. The regression model included covariates for gestational age, small for gestational age, sex, Apgar at 5 minutes <5, antenatal steroids, and the race ethnicity eigenvalues 1-4. The top 10 SNPs (by lowest *P*-value) for each of the 3 models were mapped to genes.

We assigned genes to pathways (gene sets) using the Molecular Signatures Database (http://www.broadinstitute.org/gsea/msigdb/collections.jsp). SNPs were assigned to gene(s) based on being exonic, intronic, untranslated region, or within 20 kb of the ends of the gene model. Pathways were analyzed using gene set enrichment analysis.[15]

Gene expression values for individual members of pathways considered most important were extracted from an existing dataset describing genome-wide expression in lung tissue obtained from BPD cases or controls and assessed for differential expression.[16] Two selected molecules (miR-219 and CD44) were further evaluated by TaqMan Gene Expression assays (Life Technologies, Grand Island, New York) from RNA isolated via the QIAGEN RNeasy FFPE kit (QIAGEN, Valencia, California) from paraffin-embedded, formalin-fixed samples of lungs collected at autopsy from extremely preterm infants (24-28 weeks' gestation) who died soon after birth, term stillborn infants, and preterm infants who died from BPD at term corrected age (36-44 weeks' postmenstrual age; n = 4/group).

Three molecules (miR-219, adenosine deaminase [ADARB2], and CD44) were selected for further evaluation in a mouse model. Gene expression was evaluated at different points during alveolar septation and hyperoxia exposure, using samples from studies approved by the UAB Institutional Animal Care and Use Committee.[17,18] RNA was isolated from lung homogenates for real-time reverse-transcription polymerase chain reaction using specific primers.[19]

Results

The GWAS cohort included 834 infants whose DNA samples were successfully genotyped. A total of 172 (20%) samples required whole-genome amplification; 751 infants met inclusion criteria with adequate information on BPD phenotype and genotyping (>97% call rate). Characteristics of the study cohort are listed in **Table I** (available at www.jpeds.com). As expected, infants who developed outcomes of interest (BPD/death; severe BPD/death; severe BPD in survivors) were more immature, of lower birth weight, more likely to be male, mechanically ventilated, and ventilated for a longer duration compared with those who did not develop these outcomes.

GWAS Analysis

None of the SNPs was significant at the genome-wide significance level ($P < 10^{-8}$). The analysis for top 10 SNPs for BPD/death (**Table II**) identified 4 SNPs in ADARB2, 2 SNPs in CD44, 1 in NSMC4A, 1 in WDR45L, and 2 associated with no known gene. Similarly, the top 10 SNPs for severe BPD/death were 4 SNPs in ADARB2, 1 in CD44, 1 in NSMC4A, 1 in NUAK1, 1 in KCNH7, and 2 associated with no known gene (**Table II**). The analysis for severe BPD in survivors also found ADARB2, CD44, NUAK1, KCNH7, and WDR45B, in addition to GRIP1 and GALNTL6 (**Table II**). Most of these SNPs had *P* values of 10^{-6} to 10^{-7}.