

UCSF

UC San Francisco Previously Published Works

Title

A Genome-Wide Association Study (GWAS) for Bronchopulmonary Dysplasia

Permalink

<https://escholarship.org/uc/item/9kt5853z>

Journal

Pediatrics, 132(2)

ISSN

0031-4005

Authors

Wang, Hui
St. Julien, Krystal R
Stevenson, David K
et al.

Publication Date

2013-08-01

DOI

10.1542/peds.2013-0533

Peer reviewed

A Genome-Wide Association Study (GWAS) for Bronchopulmonary Dysplasia



WHAT'S KNOWN ON THIS SUBJECT: Twin studies suggest that bronchopulmonary dysplasia (BPD) is heritable; however, only a small number of genetic loci have been associated with BPD and these explain only a limited amount of this heritability.



WHAT THIS STUDY ADDS: A genome-wide association study of singleton infants (899 BPD cases and 827 controls) of 25 to 30 weeks' gestational age did not identify single-nucleotide polymorphisms associated with BPD at the genome-wide significance level but did identify polymorphisms warranting further study.

AUTHORS: Hui Wang, PhD,^a Krystal R. St. Julien, BS,^{b,c} David K. Stevenson, MD,^a Thomas J. Hoffmann, PhD,^d John S. Witte, PhD,^d Laura C. Lazzaroni, PhD,^e Mark A. Krasnow, MD, PhD,^{b,c} Cecele C. Quaintance, MSN,^a John W. Oehlert, MS,^a Laura L. Jelliffe-Pawlowski, PhD,^f Jeffrey B. Gould, MD,^{a,g} Gary M. Shaw, DrPH,^a and Hugh M. O'Brodovich, MD^a

Departments of ^aPediatrics, ^bBiochemistry, and ^cPsychiatry, and ^dHoward Hughes Medical Institute, Stanford University School of Medicine, Palo Alto, California; ^eDepartment of Epidemiology and Biostatistics, and Institute for Human Genetics, University of California San Francisco, San Francisco, California; ^fCalifornia Genetic Disease Screening Program of the California Department of Public Health, Richmond, California; and ^gCalifornia Perinatal Quality Care Collaborative, California

KEY WORDS

genome-wide association study (GWAS), chronic lung disease, genetic predisposition to disease, premature, very low birth weight infant

ABBREVIATIONS

BPD—bronchopulmonary dysplasia
BW—birth weight
CPQCC—California Perinatal Quality Care Collaborative
GA—gestational age
gDNA—genomic DNA
GWAS—genome-wide association study
IPPV—intermittent positive pressure ventilation
NBS—newborn screening bloodspots
OR—odds ratio
PC—principal component
PMA—postmenstrual age
PNA—postnatal age
SNP—single-nucleotide polymorphism
VLBW—very low birth weight

(Continued on last page)

abstract



OBJECTIVE: Twin studies suggest that heritability of moderate-severe bronchopulmonary dysplasia (BPD) is 53% to 79%, we conducted a genome-wide association study (GWAS) to identify genetic variants associated with the risk for BPD.

METHODS: The discovery GWAS was completed on 1726 very low birth weight infants (gestational age = 25⁰–29^{6/7} weeks) who had a minimum of 3 days of intermittent positive pressure ventilation and were in the hospital at 36 weeks' postmenstrual age. At 36 weeks' postmenstrual age, moderate-severe BPD cases ($n = 899$) were defined as requiring continuous supplemental oxygen, whereas controls ($n = 827$) inhaled room air. An additional 795 comparable infants (371 cases, 424 controls) were a replication population. Genomic DNA from case and control newborn screening bloodspots was used for the GWAS. The replication study interrogated single-nucleotide polymorphisms (SNPs) identified in the discovery GWAS and those within the HumanExome beadchip.

RESULTS: Genotyping using genomic DNA was successful. We did not identify SNPs associated with BPD at the genome-wide significance level (5×10^{-8}) and no SNP identified in previous studies reached statistical significance (Bonferroni-corrected P value threshold .0018). Pathway analyses were not informative.

CONCLUSIONS: We did not identify genomic loci or pathways that account for the previously described heritability for BPD. Potential explanations include causal mutations that are genetic variants and were not assayed or are mapped to many distributed loci, inadequate sample size, race ethnicity of our study population, or case-control differences investigated are not attributable to underlying common genetic variation. *Pediatrics* 2013;132:290–297

Bronchopulmonary dysplasia (BPD), a disorder characterized by impairment of alveolarization, remains a leading cause of morbidity and mortality in premature infants.¹ The risk of developing BPD rises with decreasing gestational age (GA) or birth weight (BW).^{2,3} Infants with BPD require prolonged stays in NICUs, and after discharge many infants require supplemental oxygen therapy and experience frequent hospitalizations.

Although mechanisms responsible for BPD have been investigated,² and there is evidence for mediators and pathways,³ there has been little progress in decreasing the incidence of BPD in very low birth weight (VLBW) infants (BW <1500 g). Because 2 separate twin studies^{4,5} indicated that genetic factors were a major risk for developing BPD, investigators have used different strategies to identify heritable factors, including identification of frequency differences in single-nucleotide polymorphisms (SNPs) in candidate genes. Although there have been reports of promising genes, such as SFTPB, these efforts have been largely unsuccessful (reviewed in refs 6 and 7). To investigate potential genetic etiologies of BPD in VLBW infants, we conducted a population-based genome-wide association study (GWAS) among cases and controls in California to identify disease-susceptible genes and generate biological hypotheses.

METHODS

Patient Population and Phenotype Definitions

Records of infants for this case-control study of singletons were identified from the California Perinatal Quality Care Collaborative (CPQCC; <http://www.cpqcc.org/>)⁸ database during calendar years 2005 to 2008. The CPQCC prospectively collects clinical data from 128 member hospitals representing more than 90% of all NICU admissions in California. CPQCC conducts yearly data abstractor training at California locations. Each

record has range and logic checks during data collection and before data closeout. Records with excessive missing data are audited.

Inclusion criteria included GA 25⁰ to 29^{6/7} weeks, BW <1500 g, and a minimum of 3 days intermittent positive pressure ventilation (IPPV) during the infant's hospitalization up to 36 weeks' postmenstrual age (PMA). The database did not differentiate between nasal ventilation and tracheal ventilation. The minimum of 3 days IPPV was included as 1 inclusion criterion so that this "environmental factor" would be consistent for both cases and controls. We used standard National Institute of Child Health and Human Development/National Institutes of Health criteria for diagnosis of mild, moderate, and severe BPD.^{1,9} Since Lavoie et al⁵ demonstrated that heritability was associated with moderate-severe, but not mild BPD, we defined cases as infants requiring supplemental oxygen at 36 weeks' PMA and controls as infants who were breathing room air at 36 weeks' PMA. The need for supplemental oxygen was determined by the practices of the individual NICU and physiologic assessments¹⁰ were not routinely carried out. Infants were excluded if they did not meet inclusion criteria, were 1 of a multiple birth, had major congenital abnormalities, if they died or left the hospital before 36 weeks' PMA, or if supplemental oxygen status at 36 weeks' PMA was not known. We also excluded infants with major surgery, as interventions, including assisted ventilation, may have introduced bias. Patent ductus arteriosus ligation was not an exclusion criterion. Control infants met all inclusion criteria and were in the hospital but did not require supplemental oxygen at 36 weeks' PMA (Supplemental Methods).

Research Ethic Boards

The Institutional Review Board of Stanford University and the Health and the Welfare Agency Committee for the

Protection of Human Subjects of the State of California approved this study. California newborn screening bloodspots (NBS) may be used for anonymous research studies unless parents refuse.

Genotyping

The California Department of Public Health linked study subjects' CPQCC information to their NBS. Genomic DNA (gDNA) was extracted from bloodspots by using the protocol described by St Julien et al¹¹ and genotyped (Illumina HumanOmni2.5 beadchip, San Diego, CA). Nonamplified gDNA was used and the genotype calls were made by using GenomeStudio software (Illumina, 2011) after quality control procedures (Supplemental Methods).

Statistical Analyses

To identify potential factors for risk adjustment in analyses between SNPs and BPD, we first examined univariate relationships between variables reported to be associated with the risk of BPD.¹² Variables with $P < .20$ were entered into a logistic model and those with $P \leq .05$ were retained in models using backward selection. Based on variables and their results, potential associations between measured SNPs and BPD were evaluated with an additive genetic model using logistic regression, adjusted for genetic ancestry, gender, and BW. Bivariate analyses of a number of factors, as well as findings from the literature, indicated that both gender and BW were significantly associated with BPD. To estimate genetic ancestry, we calculated principal components (PCs) from a subset of 114 764 (see Supplemental Information) SNPs¹³ and adjusted for PCs that were significantly associated with BPD. We also performed self-reported race/ethnicity stratified analyses within African American, Hispanic, and Caucasian individuals. In addition to exploring genetic risks under an additive effect model, we explored risks under a dominant model. We also explored potential confounding on BPD risks associated with clinical practice variation

across NICUs by using a propensity score model and found no major confounding effect. To refine association signals, an additional 9 217 535 SNPs imputed from the 1000 Genomes Project were analyzed as additive dosages.¹³ In addition to single SNP analyses, gene sets and pathways (Kyoto Encyclopedia of Genes and Genomes, Reactome, Biocarta, and Gene Ontology, Supplemental Methods) were assessed for association with BPD.

Replication Study

After analyzing our discovery GWAS data, we selected 5673 SNPs for replication in an independent replication data set of 371 BPD cases and 424 controls based on their *P* values and odds ratios (ORs; Supplemental Methods). Infants were selected from CPQCC calendar years 2009 to 2010 by using the same criteria as in the discovery GWAS. All DNA replication samples were genotyped by using the Illumina HumanExome beadchip with custom SNP content added for the 5673 SNPs. Non-amplified gDNA was used and genotype calls were made by using GenomeStudio software (Illumina, 2011). Replication study data were analyzed both individually and in combination with the discovery set, the latter being termed joint analysis.

RESULTS

Patient Characteristics

From 2005 to 2008 there were more than 2 million births in California and 61 695 infants were admitted to the NICUs that report to the CPQCC. After applying inclusion and exclusion criteria, we identified 1063 controls and 1091 cases with moderate-severe BPD (total = 2154). NBS were matched to 851 controls and 922 infants with moderate-severe BPD (total = 1773). See Supplemental Fig 3 for number of infants at each step. The 2154 enrolled infants demonstrated the expected increased risk of moderate-severe BPD as GA and BW decreased (Table 1). Because we included only VLBW infants who had a minimum of 3 days

IPPV, the overall proportion of moderate-severe BPD is higher than what would be expected if all VLBW neonates had been included. From 2005 to 2010, the incidence of BPD in all premature infants who weighed 501 to 1500 g at birth and were entered in the CPQCC database was 32%. Self-reported race/ethnicity followed a frequency pattern similar to all of California's births (Table 1, Supplemental Fig 4). Factors known to increase or decrease BPD risk¹² were observed (Table 1, Supplemental Tables 3 and 4). Univariate and bivariate analyses of a number of factors, as well as findings from the literature, indicated that both gender and BW were significantly associated with BPD.

Discovery GWAS

Genomic DNA extracted from 1726 bloodspots (827 controls and 899 cases) was used in the GWAS discovery study and 1 795 103 SNPs with minor allele frequency ≥ 0.01 were successfully genotyped. The first 3 components of PC analysis identified 4 race/ethnic groups, representing African American, Hispanic, Caucasian, and Asian (Fig 1A). There were differences between genotyped race/ethnic background and self-identified race/ethnicity background (Fig 1B). The genomic inflation factor ($\lambda_{GC} = 1.0051$)¹⁴ and the quantile-quantile plot revealed no evidence of inflation of test statistics due to population stratification (Supplemental Fig 5).

Assuming a genome-wide significance threshold¹⁵ of 5×10^{-8} , no SNPs were significantly associated with moderate-severe BPD (Fig 2). The SNP with the smallest *P* value was rs8089528 ($P = 8.64 \times 10^{-7}$), an intergenic SNP on chromosome 18. Additional SNPs with small *P* values included rs118078182 ($P = 1.30 \times 10^{-6}$), an intronic SNP in collagen, type XXIII, alpha1 (*COL23A1*), and rs12571250 ($P = 1.24 \times 10^{-6}$), an intron in the bicaudal C homolog 1 (*BICC1*) gene. We further evaluated our GWAS discovery cohort and identified 259 "super-control"

infants who were breathing room air at both 28 days' postnatal age (PNA) and 36 weeks' PMA and 568 infants with mild BPD^{1,9} who were receiving supplemental oxygen at 28 days' PNA but breathing room air at 36 weeks' PMA. Assuming a genome-wide significance threshold¹⁵ of 5×10^{-8} , we did not find any SNPs that were significantly associated with moderate-severe BPD relative to either mild BPD or the super-control group.

Replication Study

The replication population-based study cohort of 371 moderate-severe BPD cases and 424 controls shared similar demographic and clinical characteristics as the discovery GWAS cohort. No SNP from the discovery stage replicated, and joint analysis of discovery and replication cohorts did not yield any genome-wide significant SNPs (Supplemental Fig 6). Nevertheless, our joint analysis identified 6 SNPs with $P \leq 10^{-5}$, a criterion used by the National Human Genome Research Institute (<http://www.genome.gov/26525384>). The strongest signals from the joint analysis were rs556493 ($P = 1.42 \times 10^{-6}$, OR = 0.74), an intronic SNP in syntaxin binding protein 5 (STXBP5), and rs12356475 ($P = 1.54 \times 10^{-6}$, OR = 0.70), an intronic SNP in catenin (cadherin-associated protein), $\alpha 3$ (CTNNA3) (Table 2). Nine SNPs with a $P \leq 10^{-5}$ that were detected in the discovery stage did not replicate (Table 2). SNPs with $P \leq 10^{-5}$ and in linkage disequilibrium with the 3 index SNPs in Table 2 are tabulated in Supplemental Table 5.

SNPs Identified in Previous Studies

We examined 27 SNPs previously associated with BPD (Supplemental Table 6). None reached replication-wide significance ($P < .05/27 = 0.0018$). Supplemental Table 7 shows SNPs with *P* values < 0.1 in our discovery GWAS analysis or race/ethnicity subgroup analysis (adjusted for PCs); these include (*SPOCK 2*) (rs1245560), vascular endothelial growth factor A (*VEGFA*) (rs699947 and rs833061), superoxide

TABLE 1 Controls and Cases of Moderate-Severe BPD in Discovery GWAS

Characteristics	Controls (<i>n</i> = 827) <i>n</i> (%)	Cases (<i>n</i> = 899) <i>n</i> (%)	Odds Ratio (<i>n</i> = 1726) OR (95% CI)
Maternal age, y			
<18	38 (4.6)	41 (4.6)	1.1 (0.7–1.8)
18–20	91 (11.0)	115 (12.8)	1.3 (0.9–1.8)
21–25	165 (20.0)	208 (23.1)	1.3 (1.0–1.7)
26–30	202 (24.4)	199 (22.1)	ref
31–35	191 (23.1)	196 (21.8)	1.0 (0.8–1.4)
36–40	117 (14.1)	114 (12.7)	1.0 (0.7–1.4)
>40	23 (2.8)	26 (2.9)	1.1 (0.6–2.1)
Maternal race/ethnicity (self-reported)			
African American	117 (14.1)	108 (12.0)	0.7 (0.5–1.0)
Hispanic	448 (54.2)	460 (51.2)	0.8 (0.7–1.0)
Asian/Pacific Islander	74 (8.9)	93 (10.3)	1.0 (0.7–1.5)
Other	14 (1.7)	22 (2.4)	1.3 (0.6–2.5)
Caucasian	174 (21.0)	216 (24.0)	ref
Mother's years of education			
<12	241 (30.7)	291 (34.3)	1.1 (0.8–1.4)
High school graduate or GED	215 (27.4)	239 (28.2)	ref
>12	330 (42.0)	318 (37.5)	0.9 (0.7–1.1)
Gender			
Male	405 (49.0)	495 (55.1)	1.3 (1.1–1.5)
Female	422 (51.0)	404 (44.9)	ref
Gestational age, wk			
25(0/7)–25(6/7)	136 (16.4)	255 (28.4)	3.3 (2.3–4.7)
26(0/7)–26(6/7)	187 (22.6)	234 (26.0)	2.2 (1.6–3.1)
27(0/7)–27(6/7)	203 (24.5)	188 (20.9)	1.6 (1.1–2.3)
28(0/7)–28(6/7)	168 (20.3)	146 (16.2)	1.5 (1.1–2.2)
29(0/7)–29(6/7)	133 (16.1)	76 (8.5)	ref
Birth weight, g			
150–499	5 (0.6)	15 (1.7)	5.4 (1.9–15.0)
500–749	99 (12.0)	224 (24.9)	4.1 (3.0–5.4)
750–1000	345 (41.7)	449 (49.9)	2.3 (1.9–2.9)
1001–1499	378 (45.7)	211 (23.5)	ref
IUGR			
Yes	50 (6.3)	72 (8.4)	1.4 (0.9–2.0)
No	744 (93.7)	783 (91.6)	ref
SGA			
SGA	69 (8.3)	142 (15.8)	2.1 (1.5–2.8)
NGA	758 (91.7)	757 (84.2)	ref
Surfactant given in the delivery room			
Yes	281 (34.1)	353 (39.4)	1.2 (1.0–1.5)
No	544 (65.9)	544 (60.7)	ref
Surfactant given at any time			
Yes	746 (90.2)	846 (94.2)	1.8 (1.2–2.5)
No	81 (9.8)	52 (5.8)	ref
Early bacterial sepsis and/or meningitis (\leq day 3)			
Yes	27 (3.3)	29 (3.2)	1.0 (0.6–1.7)
No	799 (96.7)	869 (96.8)	ref
Late sepsis and/or meningitis ($>$ day 3)			
Yes	94 (11.4)	192 (21.4)	2.1 (1.6–2.8)
No	732 (88.6)	706 (78.6)	ref
Grade IVH			
0	550 (67.3)	543 (61.0)	ref
1	121 (14.8)	160 (18.0)	1.3 (1.0–1.7)
2	72 (8.8)	84 (9.4)	1.2 (0.8–1.7)
3	42 (5.1)	55 (6.2)	1.3 (0.9–2.0)
4	32 (3.9)	48 (5.4)	1.5 (0.96–2.4)
Grade ROP			
0	412 (53.0)	296 (33.7)	ref
1	155 (19.9)	173 (19.7)	1.6 (1.2–2.0)

dismutase 2, mitochondrial (*SOD2*) (rs5746136), toll-interleukin 1 receptor (TIR) domain containing adaptor protein (*TIRAP*) (rs8177374), mannose-binding lectin (protein C) 2, soluble (*MBL2*) (rs5030737). In our set-based analysis, surfactant protein D (*SFTPD*), interleukin 18 (interferon- γ -inducing factor) (*IL18*), superoxide dismutase 3, extracellular (*SOD3*), matrix metalloproteinase 16 (membrane-inserted) (*MMP16*), and selectin L (*SELL*) had $P < .1$ (Supplemental Table 6).

Genetic Models and Covariables

GWAS analyses included the examination of differing genetic models and covariables as possible influences on risk of moderate-severe BPD. Results associated with a dominant genetic model largely agreed with those from the additive model, as expected (the additive model is still highly powered for a dominant genetic model); P values associated with 12 SNPs that were larger than 10^{-5} when analyzed under an additive model became less than 10^{-5} when analyzed under a dominant model in joint analyses (Supplemental Table 8). We also tested models without adjustments for gender and BW. Although there was no dramatic change in the result, P values of 5 SNPs were substantially smaller (up to 1.4 orders of magnitude), suggesting confounding of these SNPs with gender and BW (Supplemental Table 9). Further examination showed that confounding was associated with BW. The higher-order term of BW showed no association with moderate-severe BPD in the discovery data set and significant association in the replication data set. However, adding the higher-order term of BW had a negligible effect on discovery results. Analyses adjusting for NICUs using a propensity score did not reveal substantively different results despite prevalence differences of moderate-severe BPD across NICUs. Year of birth did not affect findings.

Stratified analyses by self-reported race/ethnicity (and adjusted for PCs) identified a total of 21 SNPs with $P \leq 10^{-5}$ in both

TABLE 1 Continued

Characteristics	Controls (<i>n</i> = 827) <i>n</i> (%)	Cases (<i>n</i> = 899) <i>n</i> (%)	Odds Ratio (<i>n</i> = 1726) OR (95% CI)
2	135 (17.4)	234 (26.7)	2.4 (1.9–3.1)
3	75 (9.6)	170 (19.4)	3.2 (2.4–4.4)
4	1 (0.1)	5 (0.6)	6.9 (0.8–59.9)
Pneumothorax			
Yes	26 (3.1)	45 (5.0)	1.6 (0.99–2.7)
No	801 (96.9)	854 (95.0)	ref
PDA severity			
No PDA	282 (40.3)	231 (28.4)	ref
Treated only with indomethacin	331 (47.3)	344 (42.3)	1.3 (1.0–1.6)
Treated with indomethacin, then surgery	53 (7.6)	149 (18.3)	3.4 (2.4–4.9)
Direct to surgery	34 (4.9)	89 (10.9)	3.2 (2.1–4.9)
Postnatal steroids given for CLD			
Yes	44 (5.4)	196 (22.0)	5.0 (3.5–7.0)
No	777 (94.6)	769 (78.0)	ref

CI, confidence interval; CLD, chronic lung disease; GED, General Education Development; IUGR, intrauterine growth restriction; IVH, intraventricular hemorrhage; NGA, normal for GA; PDA, patent ductus arteriosus; ref, reference group for ORs; ROP, retinopathy of prematurity; SGA, small for GA.

GWAS discovery samples and combined samples (Supplemental Table 10). The strongest statistical signal was identified in Caucasians in discovery samples: rs6988306 (OR = 0.23, $P = 2.81 \times 10^{-7}$), an intron SNP in GRHL2. This SNP did not show statistical evidence of replication (replication OR = 0.81, $P = .52$; joint analysis OR = 0.36, $P = 2.8 \times 10^{-6}$). The SNPs identified in each race/ethnic subpopulation did not overlap.

Imputed SNPs and Pathway Analyses

Results from imputed SNPs in the discovery phase supported associations of genotyped SNPs and 5 imputed SNPs showed a stronger signal than their linked genotyped SNPs in joint analysis (Table 2). Our set-based analyses investigated how groups of SNPs affected risk of moderate-severe BPD. No gene or pathway was statistically associated with moderate-severe BPD (Bonferroni correction for P values: 5×10^{-6} for gene set analysis and 2×10^{-5} for pathway analysis).

Exome SNP Analyses

Associations with ~41 000 SNPs with minor allele frequency >0.01 present on the exome array were evaluated in replication samples. One SNP, a missense mutation in muscular Lamin A/C–interacting

protein, had $P < 10^{-5}$ (Supplemental Fig 7, Supplemental Table 11). No exome array SNPs reached statistical significance using set-based (gene) analyses.

DISCUSSION

To systematically investigate genetic influences on VLBW infants developing moderate-severe BPD, we used a large population-based approach to evaluate millions of SNPs by using both a GWAS and an exome-based genotyping platform. We neither identified SNPs associated with moderate-severe BPD at the genome-wide significance level (5×10^{-8}) nor replicated previous findings. This indicates that heritability of BPD is not attributable to 1, or only a few, of the large number of ancestrally conserved genetic variants assayed herein. We did identify 15 SNPs with $P \leq 10^{-5}$ in either our discovery stage or joint analyses, which will help guide future studies trying to characterize potential genetic risks for moderate-severe BPD.^{4,5} Our novel demonstration that nonamplified gDNA from NBS can be used for a GWAS has significant implications for the millions of stored NBS.

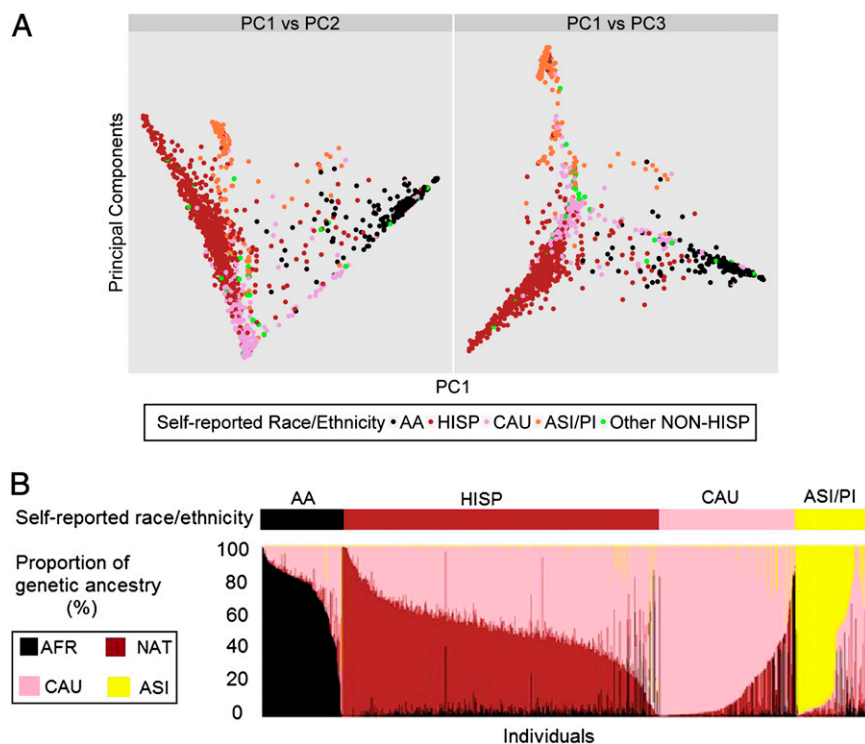
Our most promising SNP was rs118078182 in *COL23A1*, a transmembrane collagen expressed in the developing lung's mesenchyme.¹⁶ *COL23A1* has a cleavable canonical collagen domain localized to the

outside of the cell. *COL23A1* may anchor mesenchyme cells to the basement membrane through the collagen domain.

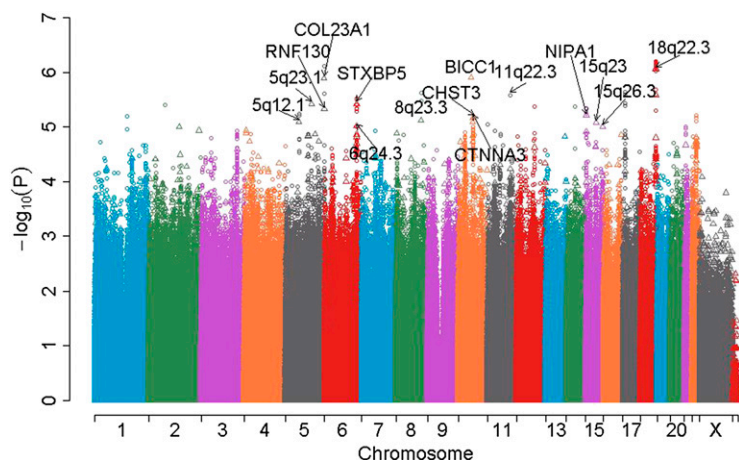
Although *COL23A* and the other SNPs did not replicate in our replication study, our replication sample size was limited and real signals may have gone undetected. Others suggest that a P value criterion of 5×10^{-8} for genome-wide significance is too stringent.¹⁵ Therefore, testing the types of SNPs described here for functional relevance has merit.

A GWAS on 43 BPD French patients using pooled DNA from cases and controls identified *sparc/osteonectin, cwcv, and kazal-like domains proteoglycan (testican) 2 (SPOCK 2)* as an at-risk gene.¹⁷ Hadchouel et al replicated their findings in Finnish infants.¹⁷ The association for *SPOCK 2* SNP rs1245560 was seen for both Caucasians ($P = .02$, OR = 1.85) and Africans ($P = .007$, OR = 2.43), whereas the association of rs1049269 was observed only in Caucasians ($P = .025$, OR = 1.79).¹⁷ The association was observed with moderate-severe, but not mild, BPD, which is consistent with previous observations.⁵ Our GWAS did not replicate Hadchouel et al's findings in the overall case-control analyses (OR = 1.0, $P = .97$ for rs1245560; and OR = 1.0, $P = .95$ for rs1049269). Analyses stratified by self-reported race/ethnicity (adjusted for PCs), showed a weak association for rs1245560 ($P = .08$, OR = 1.32) in only Caucasians, thus lending some support for *SPOCK2* association with BPD. The P values of these 2 SNPs in replication samples and the overall P value of the *SPOCK 2* gene showed no association with BPD. Because the ancestral proportions for racially admixed populations have been shown to influence lung function in adults¹⁸ and children,¹⁹ it is possible we did not observe an association with *SPOCK 2* because of the relatively small number of Caucasian and African American infants.

Two SNPs (rs3771159 and rs3771171) in interleukin-18 have been associated with BPD.²⁰ We did not replicate these

**FIGURE 1**

Race/ethnicity of study population. A, PC analysis for population stratification. In the left panel, the first PC (PC1) is plotted against the second; in the right panel, the first PC is plotted against the third PC (PC2 and PC3). Each point represents an infant sample, and clusters of PCs represent population substructures. Colors represent self-reported race/ethnicities. Texts in plots indicate directions of genetic ancestries. AA, African American; ASI, Asian; CAU, Caucasian; HISP, Hispanic; NAT, North American Native; NON-HISP, Non-Hispanic; PI, Pacific Islander. B, Estimated proportions of genetic ancestries defined by SNP-typing for our study population. Colors represent ancestries. Each vertical line represents an individual infant, and the vertical spread of a specific color indicates the percentage of a specific ancestry in that individual. The self-reported race/ethnicity is aligned on top of the estimated genetic ancestries. The estimated genetic ancestry has discrepancies with self-reported race/ethnicity and displays a high level of population mixture. The small numbers of "Other" infants (see Table 1) are not illustrated. AA, African American; AFR, African; ASI, Asian; CAU, Caucasian; HISP, Hispanic; NAT, North American Native; PI, Pacific Islander.

**FIGURE 2**

Manhattan plot of GWAS discovery results. Measured SNPs are represented with triangles, and imputed SNPs are represented with circles. Results are plotted on negative log₁₀ scale. SNPs in Table 2 are labeled with their gene symbols or chromosomal regions.

associations. However, SNP rs61731845, a missense mutation in paralemmin 3 (*PALM3*), which binds to immunoglobulin interleukin-1 receptor-related molecule (*SIGIRR*),²¹ had an OR of 6.97 ($P = 8.5 \times 10^{-5}$ risk allele = G) in our discovery GWAS. It, however, did not replicate (OR = 1.09, $P = .88$). There were significant differences in patients of the previous study²⁰ and our study. The previous study evaluated 1091 Caucasian and African American infants who were born between 1989 and 2008, had a GA <35 weeks, and the BPD criterion used supplemental oxygen at 28 days' PNA. In contrast, our population was predominantly Hispanic, was VLBW, had a GA <30 weeks, and if infants were on supplemental oxygen at 28 days' PNA, yet on room air at 36 weeks' PMA, they were included in our control group. They were included in our control group because previous work demonstrated heritability of BPD was not observed in infants with mild BPD.⁵

Most genetic studies for BPD have been small, have included a limited set of genes, and primarily targeted Caucasians (reviewed in ref 6). Our GWAS included a highly diverse population (Fig 1) and has by far the largest sample size and number of SNPs interrogated to date for BPD. Confounding from genetic ancestries was accounted for in logistic regression by using PCs, a method that addresses population stratification in heterogeneous populations.²² However, relative to GWASs focusing on adult diseases, which have involved tens of thousands of patients, our sample size is relatively small. This limited power to detect small-moderate associations.

Our inability to detect loci at genome-wide significant signals indicates that genetic risk for moderate-severe BPD is not likely to be solely influenced by 1, or a small number, of major ancestrally conserved genetic variants. Our approach may have missed some other heritable genetic paradigm, such as rare mutations, epigenetic effects, joint effects of multiple SNPs, copy number variations, or

TABLE 2 SNPs With $P < 10^{-5}$ in Discovery GWAS and Joint Analyses of the GWAS and Replication

Chr.	SNP	Closest Genes	Gene Location	Allele ^a	MAF ^b (Cases/Ctrls)	GWAS		Replication		Joint Analysis	
						OR ^c	P	OR	P	OR (95% CI)	P
I	5q12.1	<i>ISCA1L, HTR1A</i>	Intergenic	A/G	0.10/0.06	1.85	8.12×10^{-6}	1.36	1.03×10^{-1}	1.65 (1.33–2.04)	5.86×10^{-6}
	5q23.1	<i>PRR16, FTMT</i>	Intergenic	G/A	0.02/0.05	0.31	3.83×10^{-6}	0.61	1.09×10^{-1}	0.41 (0.28–0.60)	3.78×10^{-6}
	6q24.3	<i>STXBP5</i>	Intron	A/G	0.46/0.53	0.70	3.82×10^{-6}	0.82	6.01×10^{-2}	0.74 (0.66–0.84)	1.42×10^{-6}
	10q21.3	<i>CTNNA3</i>	Intron	G/A	0.21/0.26	0.67	6.77×10^{-6}	0.76	3.76×10^{-2}	0.70 (0.61–0.81)	1.54×10^{-6}
	5q35.3	<i>COL23A1</i>	Intron	A/G	0.06/0.02	2.67	1.30×10^{-6}	0.93	7.50×10^{-1}	1.76 (1.31–2.36)	1.82×10^{-4}
	5q35.3	<i>RNF130</i>	Intron	C/A	0.50/0.42	1.41	4.67×10^{-6}	1.04	6.99×10^{-1}	1.27 (1.13–1.43)	7.07×10^{-5}
	8q23.3	<i>CSMD3, TRPS1</i>	Intergenic	G/A	0.30/0.38	0.71	7.67×10^{-6}	0.94	5.94×10^{-1}	0.78 (0.69–0.88)	8.15×10^{-5}
	10q21.1	<i>BICC1</i>	Intron	A/G	0.06/0.02	2.92	1.24×10^{-6}	0.97	9.31×10^{-1}	2.03 (1.45–2.84)	3.31×10^{-5}
	10q22.1	<i>GHS3</i>	Intron	G/A	0.36/0.45	0.71	6.59×10^{-6}	0.96	7.34×10^{-1}	0.79 (0.70–0.88)	6.82×10^{-5}
	15q11.2	<i>NIPA1</i>	Intron	A/G	0.19/0.25	0.67	6.28×10^{-6}	0.96	7.35×10^{-1}	0.76 (0.66–0.88)	2.29×10^{-4}
II	15q23	<i>LOC145837, C15orf50</i>	Intergenic	G/A	0.42/0.49	0.72	8.51×10^{-6}	0.99	9.33×10^{-1}	0.79 (0.71–0.90)	1.75×10^{-4}
	15q26.3	<i>ARRDC4, FAM169B</i>	Intergenic	A/G	0.37/0.31	1.4	9.97×10^{-6}	0.94	5.58×10^{-1}	1.24 (1.09–1.40)	6.54×10^{-4}
	18q22.3	<i>LOC100289370, FBXO15</i>	Intergenic	G/A	0.41/0.34	1.44	8.64×10^{-7}	1.02	0.86×10^{-1}	1.28 (1.14–1.44)	4.28×10^{-5}
	2p24.3	<i>LOC729992, LOC100288243</i>	Intergenic	G/A	0.13/0.17	0.74	2.40×10^{-3}	0.59	4.80×10^{-4}	0.68 (0.58–0.80)	4.79×10^{-6}
	18q11.2	<i>GHS9</i>	Intron	G/A	0.31/0.37	0.77	5.99×10^{-4}	0.72	2.89×10^{-3}	0.75 (0.66–0.85)	3.47×10^{-6}
	1p34.2	<i>HIVEP3</i>	Intron	A/G	0.03/0.01 (0.99)	3.29	3.80×10^{-5}	1.92	3.50×10^{-2}	2.52 (1.67–3.79)	9.19×10^{-6}
	6q24.3	<i>STXBP5, YAP1P1</i>	Intergenic	G/A	0.48/0.53 (0.93)	0.71	9.80×10^{-6}	0.81	5.80×10^{-2}	0.75 (0.66–0.84)	2.11×10^{-6}
	10q21.3	<i>CTNNA3</i>	Intron	A/T	0.16/0.21 (0.99)	0.65	1.10×10^{-5}	0.74	2.20×10^{-2}	0.69 (0.59–0.80)	1.20×10^{-6}
	11q22.3	<i>CARD17, CASP1P1</i>	Intergenic	T/A	0.17/0.22 (0.90)	0.63	2.60×10^{-6}	0.84	2.00×10^{-1}	0.70 (0.60–0.81)	5.00×10^{-6}

I, SNPs identified in both GWAS and joint analysis; II, SNPs identified only in joint analysis; III, SNPs identified only in GWAS; IV, imputed SNPs with stronger signals than its linked genotyped SNPs in joint analysis; Chr, chromosomal region; CI, confidence interval; Ctrls, controls; MAF, minor allele frequency.

^a The first letter denotes the minor allele; allele designation uses the top/bottom nomenclature.

^b MAFs were shown for GWAS samples; expected MAFs and information metrics (in parentheses) are reported for imputed SNPs.

^c Tested allele is the minor allele.

^d Index SNPs.

interactions among SNPs and nongenetic factors. Assessment of other models for heritability for BPD will require either different assay platforms or development of more advanced interrogation methods. Our study population does differ from the twin studies^{4,5} reporting substantive heritability underlying BPD. These twin studies did not report race/ethnicity of patients and their estimation of heritability did not investigate possible genetic heterogeneity among different populations. Based on the geographic location of the studies,^{4,5} we speculate that most of the patients were Caucasian, whereas our cases and controls were predominately of Mexican-Hispanic origin (Fig 1B). Genetic heterogeneity can affect power of detecting associations, especially in a highly heterogeneous population such as ours and, as discussed previously, there is convincing evidence that ancestral (genetic) determined race/ethnicity affects lung function in both children and adults. Finally, the eligibility criteria for both our cases and controls required a minimum of 3 days of IPPV. We chose this approach to better define the BPD phenotype and decrease the “environmental” differences between the groups in the hope that this would enhance our ability to detect genetic factors. However, BPD sometimes occurs in extremely premature infants who did not require IPPV and the findings of others may reflect, at least in part, the fact that they did not use this as one of the eligibility criteria. Moreover, unknown differences between the NICUs in California regarding their clinical approaches to these infants may have affected our ability to detect genetic effect. All these factors may have contributed to the lack of genome-wide significant findings of our study.

ACKNOWLEDGMENTS

The authors express their appreciation to Dr Richard Bland for his contributions in writing the grant to obtain funding for the research, to Drs Fred Lorey and Shabbir Ahmad for so aptly directing efforts to

make newborn blood specimens available for analyses, to Allan Santos for his de-

tailed efforts in finding and processing bloodspots, and to the many individuals

associated with the CPQCC for their efforts to create such an important database.

REFERENCES

- Walsh MC, Szeffler S, Davis J, et al. Summary proceedings from the bronchopulmonary dysplasia group. *Pediatrics*. 2006;117(3 pt 2):S52–S56
- Baraldi E, Filippone M. Chronic lung disease after premature birth. *N Engl J Med*. 2007;357(19):1946–1955
- Bhandari A, Bhandari V. Pitfalls, problems, and progress in bronchopulmonary dysplasia. *Pediatrics*. 2009;123(6):1562–1573
- Bhandari V, Bizzarro MJ, Shetty A, et al; Neonatal Genetics Study Group. Familial and genetic susceptibility to major neonatal morbidities in preterm twins. *Pediatrics*. 2006;117(6):1901–1906
- Lavoie PM, Pham C, Jang KL. Heritability of bronchopulmonary dysplasia, defined according to the consensus statement of the national institutes of health. *Pediatrics*. 2008;122(3):479–485
- Lavoie PM, Dubé MP. Genetics of bronchopulmonary dysplasia in the age of genomics. *Curr Opin Pediatr*. 2010;22(2):134–138
- Shaw GM, O'Brodovich HM. Progress in understanding the genetics of bronchopulmonary dysplasia. *Semin Perinatol*. 2013;37(2):85–93
- Gould JB. The role of regional collaboratives: the California Perinatal Quality Care Collaborative model. *Clin Perinatol*. 2010;37(1):71–86
- Jobe AH, Bancalari E. Bronchopulmonary dysplasia. *Am J Respir Crit Care Med*. 2001;163(7):1723–1729
- Walsh MC, Yao Q, Gettner P, et al; National Institute of Child Health and Human Development Neonatal Research Network. Impact of a physiologic definition on bronchopulmonary dysplasia rates. *Pediatrics*. 2004;114(5):1305–1311
- St Julien KR, Jelliffe-Pawlowski LL, Shaw GM, Stevenson DK, O'Brodovich HM, Krasnow MA. High quality genome-wide genotyping from archived dried blood spots without DNA amplification. *PLoS One*. 2013;8(5):e64710
- Laughon MM, Langer JC, Bose CL, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Prediction of bronchopulmonary dysplasia by postnatal age in extremely premature infants. *Am J Respir Crit Care Med*. 2011;183(12):1715–1722
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–909
- Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55(4):997–1004
- Panagiotou OA, Ioannidis JP; Genome-Wide Significance Project. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol*. 2012;41(1):273–286
- Visel A, Thaller C, Eichele G. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res*. 2004;32(Database issue):D552–D556
- Hadchouel A, Durrmeyer X, Bouzigon E, et al. Identification of SPOCK2 as a susceptibility gene for bronchopulmonary dysplasia. *Am J Respir Crit Care Med*. 2011;184(10):1164–1170
- Kumar R, Seibold MA, Aldrich MC, et al. Genetic ancestry in lung-function predictions. *N Engl J Med*. 2010;363(4):321–330
- Brehm JM, Acosta-Pérez E, Klei L, et al. African ancestry and lung function in Puerto Rican children. *J Allergy Clin Immunol*. 2012;129(6):1484–1490.e6
- Floros J, Londono D, Gordon D, et al. IL-18R1 and IL-18RAP SNPs may be associated with bronchopulmonary dysplasia in African-American infants. *Pediatr Res*. 2012;71(1):107–114
- Chen X, Wu X, Zhao Y, et al. A novel binding protein of single immunoglobulin IL-1 receptor-related molecule: Paralemmin-3. *Biochem Biophys Res Commun*. 2011;404(4):1029–1033
- Bouaziz M, Ambrose C, Guedj M. Accounting for population stratification in practice: a comparison of the main strategies dedicated to genome-wide association studies. *PLoS One*. 2011;6(12):e28845

(Continued from first page)

Dr Wang and Ms St Julien contributed equally to this work.

Drs Shaw and O'Brodovich had access to all of the data and take responsibility for the integrity of the data and accuracy of the data analysis. Drs Gould, Hoffmann, Jelliffe-Pawlowski, Krasnow, O'Brodovich, Shaw, Stevenson, and Witte and Ms Quaintance contributed to study concept and design. Dr Lazzaroni contributed to the statistical design and analysis plan in the grant application. Drs Gould, Hoffmann, Jelliffe-Pawlowski, Shaw, and Wang, and Mr Oehlert and Ms St Julien contributed to acquisition of data. All authors contributed to analysis and interpretation of data and critical revision of the manuscript for important intellectual content. Drs Gould, Hoffmann, Jelliffe-Pawlowski, Shaw, Witte, and Wang, and Mr Oehlert performed the statistical analyses. Dr O'Brodovich, as principal investigator, and Drs Gould, Krasnow, Lazzaroni, Shaw, and Stevenson, and Ms Quaintance obtained funding. Dr Jelliffe-Pawlowski, Ms Quaintance, and Mr Oehlert provided administrative, technical, or material support. Drs O'Brodovich and Shaw supervised the study.

www.pediatrics.org/cgi/doi/10.1542/peds.2013-0533

doi:10.1542/peds.2013-0533

Accepted for publication May 13, 2013

Address correspondence to Hugh O'Brodovich, MD, Department of Pediatrics, Stanford University, 300 Pasteur Dr (Room H310), Stanford, CA, 94305. E-mail: hugh.obrodovich@stanford.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2013 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Funded by National Institutes of Health/National Heart, Lung, and Blood Institute grant RC2 HL101748. Funded by the National Institutes of Health (NIH).