Identification of *SPOCK2* As a Susceptibility Gene for Bronchopulmonary Dysplasia

Alice Hadchouel[1,2,3,4], Xavier Durrmeyer[5], Emmanuelle Bouzigon[6,7,8], Roberto Incitti[9], Johanna Huusko[13], Pierre-Henri Jarreau[3,4,10], Richard Lenclen[11†], Florence Demenais[6,7,8], Marie-Laure Franco-Montoya[1,4], Inès Layouni[5], Juliana Patkai[10], Jacques Bourbon[1,4], Mikko Hallman[13], Claude Danan[4,5,12], and Christophe Delacourt[1,2,3,4]

[1]Inserm, U955, Créteil, France; [2]AP-HP, Hôpital Necker-Enfants Malades, Service de Pneumologie Pédiatrique, Paris, France; [3]Université Paris-Descartes, Paris, France; [4]Fondation PremUp, Paris, France; [5]Service de Réanimation Néonatale, Centre Hospitalier Intercommunal, Créteil, France; [6]INSERM, U946, Paris, France; [7]Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, Paris, France; [8]Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain, Paris, France; [9]Institut Mondor de Recherche Biomédicale, INSERM, U955, Plateforme Microarray, Faculté de Médecine, Créteil, France; [10]Service de Réanimation Néonatale, CHU Cochin-Port Royal, Paris, France; [11]Service de Réanimation Néonatale, CH Poissy-Saint Germain, Poissy, France; [12]Unité Fonctionnelle de Recherche Clinique, Centre Hospitalier Intercommunal, Créteil, France; [13]Institute of Clinical Medicine, Department of Pediatrics, University of Oulu, Oulu, Finland

***Rationale*: Bronchopulmonary dysplasia is the most common chronic respiratory disease in premature infants. Genetic factors might contribute to bronchopulmonary dysplasia susceptibility.**

***Objectives*: To identify genetic variants involved in bronchopulmonary dysplasia through a genome-wide association study.**

***Methods*: We prospectively evaluated 418 premature neonates (gestational age <28 wk), of whom 22% developed bronchopulmonary dysplasia. Two discovery series were created, using a DNA pooling strategy in neonates from white and African ancestry. Polymorphisms associated with the disease were confirmed in an independent replication population. Genes were then explored by fine mapping and associations were replicated in an external Finnish population of 213 neonates. Validated genes expression patterns were studied in rat lung, after air or hyperoxia exposure.**

***Measurements and Main Results*: *SPOCK2* gene was identified by both discovery series. The most significant polymorphism (rs1245560; $P = 1.66 \times 10^{-7}$) was confirmed by individual genotyping, and in the replication population ($P = 0.002$). Fine mapping confirmed the association of rs1245560 with bronchopulmonary dysplasia in both white and African populations with adjusted odds ratios of 2.96 (95% confidence interval [CI], 1.37–6.40) and 4.87 (95% CI, 1.88–12.63), respectively. In white neonates, rs1049269 was also associated with the disease (odds ratio, 3.21; 95% CI, 1.51–6.82). These associations were replicated in the Finnish population. In newborn rat lungs, *SPOCK2* mRNA levels markedly increased during the alveolar stage of lung development. After rat exposure to hyperoxia, *SPOCK2* expression increased relative to air-exposed controls.**

***Conclusions*: We identified *SPOCK2* as a new possible candidate susceptibility gene for bronchopulmonary dysplasia. Its lung expression pattern points toward a potential role in alveolarization.**

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

Recent studies highlighted the contribution of genetic factors to bronchopulmonary dysplasia (BPD) susceptibility. Few candidate-gene association studies have attempted to identify BPD susceptibility genes.

What This Study Adds to the Field

By performing a genome-wide association study, we found an association between the SPOCK2 gene and BPD susceptibility. In rats, we observed that SPOCK2 was highly expressed during lung development. These data suggest SPOCK2 as a susceptibility gene for BPD. Its pattern of lung expression points to a potential role in alveolarization.

Keywords: infant; premature; DNA microarrays; lung; development

Bronchopulmonary dysplasia (BPD), defined as a requirement for oxygen supplementation at 36 weeks of postmenstrual age, is the most common chronic respiratory disease in premature infants, and its treatment places major demands on health services (1). Despite considerable obstetric and neonatal advances in the care of very-low-birth-weight infants, BPD continues to occur among 20–40% of survivors (2). BPD seems to result from multiple factors that can injure the immature lung and interrupt normal alveolar and distal vascular development (3). Besides the recognized detrimental effects of environmental factors, very-low-birth-weight twin concordance studies recently suggested a role of genetic factors (4–6). After controlling for covariates, genetic factors accounted for 53–82% of the variance in BPD. A few genetic association studies have attempted to identify candidate BPD susceptibility genes (7–11). Genome-wide association studies (GWAS) have the potential to identify genetic factors underlying complex traits and diseases. DNA pooling-based strategies have recently been used to screen for major genetic associations (12), providing the same efficiency and power as individual genotyping of cases and controls (13, 14). This strategy has been successfully used for initial prioritizing of single nucleotide polymorphisms (SNPs) for validation by individual genotyping in several complex diseases (15–18). Moreover, DNA pooling studies

(*Received in original form March 27, 2011; accepted in final form July 29, 2011*)

Supported by Program Hospitalier de Recherche Clinique AOR 07 018, Assistance Publique–Hôpitaux de Paris, and Agence Nationale de la Recherche ANR-09-GENO-037. A.H. was funded by INSERM.

†Deceased.

Author Contributions: A.H. had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design, A.H., C.D., and C.D. Acquisition of data, A.H., X.D., P.-H.J., R.L., I.L., J.P., C.D., J.H., and M.H. Analysis and interpretation of data, A.H., R.I., E.B., M.-L.F.-M., F.D., J.B., and C.D. Drafting the article, A.H., C.D., R.I., E.B., and M.-L.F.-M. Critical revision of the manuscript for important intellectual content, A.H., X.D., E.B., P.-H.J., F.D., I.L., J.P., J.B., C.D., and C.D. Final approval of the version to be published, A.H., X.D., R.I., E.B., P.-H.J., F.D., M.-L.F.-M., I.L., J.P., J.B., C.D., and C.D.

Correspondence and requests for reprints should be addressed to Alice Hadchouel, M. D., Service de Pneumologie Pédiatrique, Hôpital Necker-Enfants Malades, 149 rue de Sèvres, 75743 Paris Cedex 15, France. E-mail: alice.hadchouel-duverge@nck.aphp.fr

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 184. pp 1164–1170, 2011
Originally Published in Press as DOI: 10.1164/rccm.201103-0548OC on August 11, 2011
Internet address: www.atsjournals.org

can accurately estimate allelic frequencies, even in pools of only about 50 individuals (19). This is of particular interest when studying a disorder, such as BPD, that affects preterm infants. To identify genetic variants influencing BPD susceptibility, we conducted a GWAS of a multicenter population, combining pooling-based genome-wide case-control analysis, fine-scale mapping of gene, and animal experiments.

METHODS

Detailed methods are available in the online supplement.

Study Design and Population

The study was approved by the local ethics committee (CPP Île-de-France IX). Written informed consent was obtained from the parents. We prospectively included 418 neonates, with a gestational age less than 28 weeks, from three French neonatal intensive care units between 2002 and 2009 (*see* Figure E1 in the online supplement). BPD definition was based on the physiologic test validated by Walsh and coworkers (20). Our population was divided into two discovery series and an independent replication series. We performed a GWAS based on a DNA pooling strategy in the two discovery series. SNPs associated with BPD were validated by individual genotyping in these two populations. Validated SNPs were then genotyped in two different replication populations: the internal replication population and an external population composed of 213 Finnish neonates born before 30 weeks of gestation and included between 1997 and 2010 from the neonatal intensive care unit of the Oulu University Hospital, Oulu, Finland. Selected genes were also investigated in neonatal rats, focusing on lung developmental gene expression patterns and on changes in gene expression during distal lung development in healthy controls and animals exposed to hyperoxia.

DNA Pooling

The first discovery series consisted of 22 white cases and 76 white controls, and the second of 21 black African cases and 86 black African controls. Pooled DNA samples were created in quadruplicate for case and control groups in each series. Genotyping was performed with the Infinium II Illumina HumanHap300 Genotyping BeadChip array for the white population (318,237 SNPs) and the Illumina HumanHap650Y (Illumina, San Diego, CA) array for the African population (660,918 SNPs). Raw and transformed data sets for each microarray are available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi (Accession Number GSE22284).

We used two methods to select SNPs for further individual genotyping. First, SNPs were ranked according to their allelic frequency differences between cases and controls, and the top-ranked 1% was selected. The second method was based on a combined Z-score analysis (15) and association signals reaching a threshold of P less than or equal to 7.2×10^{-8} in Africans (21) and a less stringent threshold of P less than or equal to 5×10^{-7} in whites (22), because this population was genotyped with less SNPs, was retained. We considered that the identification of at least two significant SNPs within 30 kb reflected a single locus (23). Finally, we selected for replication loci identified by both statistical methods in both series.

Fine Mapping of the SPOCK2 Region

Using HapMap data, we identified 70 SNPs with a minimum minor allele frequency (MAF) of 0.05 in a region spanning 20 kb upstream and downstream of the *SPOCK2* gene. We identified 29 tagging SNPs that allowed capturing the 70 polymorphisms with $r^2 = 1$ using the pairwise approach implemented in Haploview4.1. The genotyping of our case-control samples was performed by Integragen (Genopole, Evry, France) with Golden Gate Illumina technology (Illumina). SNPs showing deviation from Hardy-Weinberg equilibrium in controls ($P \leq 0.05$) or with genotyping success rates of less than 90% were excluded from analysis, as were individual DNA samples for which the genotyping success rate across all SNPs was less than 70%. Twenty-two SNPs passed this quality control and three DNA samples were excluded.

Animal Experiments

Rat lung tissues from controls and animals exposed to hyperoxia were collected to perform RNA extraction, cell isolation, and frozen tissue sections. mRNA expression profiles were determined by quantitative polymerase chain reaction and location of the SPOCK2 protein was studied by immunofluorescence.

Statistical Analysis

Individual genotyping. To assess the number of cases and controls required for the replication process, a power computation was applied and showed that a sample size of at least 120 controls and 30 cases allowed detection of a mean allelic frequency difference of at least 10% between cases and controls with a type I risk error of 5% and a power of 80%. Our internal replication set was composed of 49 cases and 163 controls (30 and 112 whites, 15 and 32 Africans, and 4 and 19 infants with one white parent and one black African parent). We performed separate analyses in neonates from white and African ancestry to avoid population stratification problems. Odds ratios (ORs) and 95% confidence intervals (CIs) for association between SNPs and BPD were estimated by logistic regression. Adjustment on relevant clinical factors associated to BPD in univariate analyses ($P \leq 0.05$) was also performed. The effect of a single SNP on disease was tested under additive and dominant genetic models for the minor allele. In the fine-scale mapping process, to correct for the effect of testing multiple SNPs ($n = 22$), we estimated the effective number of independent SNPs (Meff) using Li and Ji's method (24), which is based on the pairwise LD measure r^2 between SNPs. A Bonferroni correction was then applied using the Meff estimates leading to P value thresholds of 0.0056 and 0.0036 for white and African subjects, respectively. Additional analysis was performed to assess whether *SPOCK2* polymorphisms were also associated with mild disease, defined by the need for oxygen supply at 28 days but room air breathing at 36 weeks, using the same statistical models. In the Finnish population, the same statistical models were also used. All statistical analyses were performed using Stata10 (College Station, TX).

Animal studies. Differences between treatment groups were evaluated with the Kruskal-Wallis nonparametric test for multiple group comparisons and the Mann-Whitney U test for two-group comparisons; P less than or equal to 0.05 indicated statistical significance.

RESULTS

Study Populations

The 418 French neonates had a mean gestational age and birth weight of 26.4 ± 0.1 weeks (range 23.6–27.9) and 845 ± 9 g (range 460–1410), respectively. Ninety-one infants (22%) were diagnosed with BPD (22% in white and 23% in African samples). In the entire sample, the following five clinical factors were significantly associated with BPD: (*1*) birth weight (OR [95% CI], 0.996 [0.995–0.998] per gram birth weight; $P < 0.001$); (*2*) male sex (1.60 [0.99–2.56]; $P = 0.05$); (*3*) the need for a second dose of surfactant (2.60 (1.59–4.22]; $P < 0.001$); (*4*) persistent ductus arteriosus receiving surgical treatment (6.62 [3.19–13.74]; $P < 0.001$); and (*5*) postnatal sepsis (2.40 [1.49–3.98]; $P < 0.001$). Thus, these factors were used as adjustment covariates in genetic analyses. In the Finnish population mean gestational age was 27.8 ± 0.1 week (range 23–29.9) and mean birth weight was 1,017 ± 20 g (range 370–1755). Fifty five infants (26%) were diagnosed with BPD. Available adjustment covariates in this population were birth weight, sex, and the occurrence of neonatal respiratory distress syndrome.

DNA Pooling

The allele frequency difference method selected 583 SNPs belonging to 388 genes and 1,458 SNPs belonging to 933 genes in the white and African series of pools, respectively. Among these genes, 105 were identified in both populations. Identical

SNPs belonging to four genes were detected in the two populations (*IRF2*, *KIT*, *SPOCK2*, and *LRRC4C*) and 25 genes with SNPs less than 30 kb apart in the two populations were selected (*see* Table E1).

Using the combined Z-score analysis, we selected 51 SNPs in 45 genes and 134 SNPs in 108 genes in the white and African series of pools, respectively. Among these genes, four were detected in both populations (SPOCK2, AGBL1 DMD, and IL1RAPL1). *SPOCK2* was the single gene exhibiting association signals with SNPs less than 30 kb apart in the two populations (*see* Table E2).

SPOCK2 (Entrez Gene ID: 9,806) was the only gene identified in both populations using both selection methods. Comparison of MAF between each control pool and corresponding HapMap data showed high similarity for the four SNPs belonging to *SPOCK2* region identified by pooling analyses (*see* Table E3). Because rs1245560 was the only SNP showing similar MAF in white and African control pools (0.463 and 0.497, respectively), this genetic variant was chosen for validation by individual genotyping in the two discovery series and in an independent replication set including neonates from white or African ancestry.

Individual Genotyping of rs1245560 in the Discovery Series and in Independent Replication Sample

Individual genotyping of rs1245560 in DNA samples used for pooling studies confirmed the results obtained by pooling analyses. The C allele was significantly associated with the risk of BPD in both white ($P = 0.003$) and African ($P = 0.01$) series. C allele frequencies were 0.479 and 0.512 in the white and African controls, respectively, compared with 0.75 and 0.737 in the white and African BPD neonates, respectively. Results were also significant for risk genotype effects with ORs for having the CC genotype of 3.88 (95% CI, 1.39–10.94; $P = 0.01$) and of 5.22 (95% CI, 1.82–15.00; $P = 0.002$) in white and African series, respectively. In the replication sample composed of 49 cases and 163 controls, allele C of rs1245560 was also significantly associated with an increased risk of BPD (adjusted OR, 3.57; 95% CI, 1.19–10.76; $P = 0.02$) as was CC genotype (adjusted OR, 3.75; 95% CI, 1.61–8.76; $P = 0.002$).

Fine Mapping of SPOCK2 Region

Fine mapping was performed separately in the entire white sample (51 cases and 185 controls) (Figure 1) and in the entire African sample (36 cases and 118 controls) (*see* Figure E2).

In univariate analysis, one SNP (rs1245560) was significantly associated with BPD in both populations with *P* value for the association between the risk of BPD and the CC genotype equal to 0.004 and 0.0007 in the white sample and in the African sample, respectively (Table 1; *see* Table E4 for detailed genotypes counts and frequencies). These associations remained statistically significant after correction for multiple testing. Carrying the CC genotype remained a significant risk factor for BPD after adjustment for perinatal factors with OR equal to 2.96 (95% CI, 1.37–6.40) for white subjects and 4.87 (95% CI, 1.88–12.63) for African subjects, respectively (Table 1). These associations remained also statistically significant after correction for multiple testing. Because the allele frequencies of rs1245560 did not differ among controls of the two ethnic groups (0.489 for white neonates and 0.479 for African neonates), a logistic regression analysis was applied to the entire population (91 cases and 322 controls). The overall ORs was equal to 3.71 (95% CI, 2.07–6.64; $P = 10^{-5}$) with subjects carrying CC genotype being at risk of BPD. Among the subjects from white ancestry, four supplementary SNPs were significantly associated with BPD (rs1245509, rs1245576, rs1049269, and rs1245558; $P = 0.001$–0.05). As shown in Figure 1, all significant SNPs were in strong linkage disequilibrium. After correction for multiple comparisons, a single association signal remained significant (rs1049269; OR = 3.21; 95% CI, 1.51–6.82 after adjustment for perinatal factors) (Table 1; *see* Table E4 for detailed genotypes counts and frequencies). Among African subjects, one SNP was also associated with the disease but this association did not remain significant after correction for multiple comparisons.

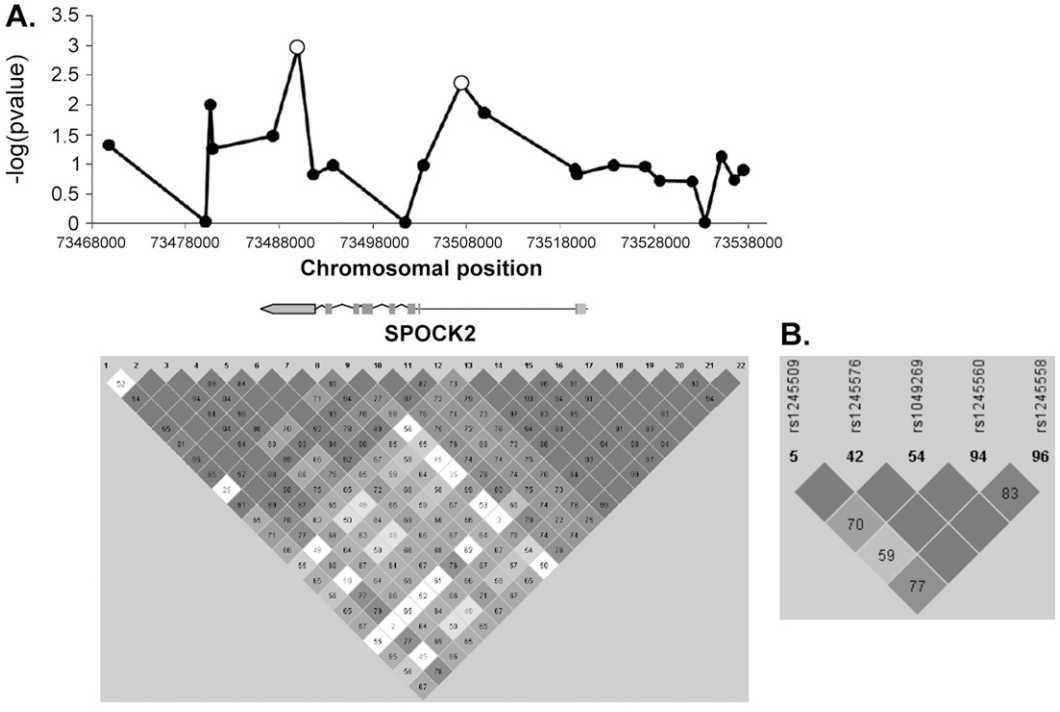

Figure 1. Fine-scale mapping of *SPOCK2* locus in neonates of white ancestry. Individual genotyping of tag– single nucleotide polymorphisms (SNPs) in the white population (51 cases and 185 controls) confirmed the results obtained in DNA pooling studies. Statistical analyses identified five SNPs as being significantly associated with bronchopulmonary dysplasia (*B*). Only rs1245560 and rs1049269 remained significant after correction for multiple testing and adjustment for major clinical risk factors (*A, white spots*). LD plots were constructed using Haploview version 4.1. *Dark grey squares* indicate regions of strong linkage disequilibrium (D′ given).

TABLE 1. ASSOCIATIONS OF THE *SPOCK2* LOCUS WITH BPD IN WHITE (51 CASES AND 185 CONTROLS) AND AFRICAN POPULATIONS (36 CASES AND 118 CONTROLS)*

	White Population												
	Risk Allele Effect							Risk Genotype Effect					
SNP	Risk Allele	AF in Cases	AF in Controls	OR (95% CI)	*P* Value	Adjusted OR† (95% CI)	*P* Value	Risk Genotype	Genotype Counts (%)‡ in Cases/Controls	OR (95% CI)	*P* Value	Adjusted OR† (95% CI)	*P* Value
rs1049269	G	0.7	0.541	1.76 (1.11–2.79)	0.015	1.79 (1.08–2.99)	0.025	GG	29 (58)/56 (31)	2.96 (1.55–5.68)	0.001	3.21 (1.51–6.82)	0.003
rs1245560	C	0.64	0.489	1.75 (1.11–2.78)	0.017	1.85 (1.10–3.11)	0.020	CC	23 (46)/43 (23)	2.62 (1.35–5.09)	0.004	2.96 (1.37–6.40)	0.005
	African Population												
	Risk Allele Effect							Risk Genotype Effect					
rs1245560	C	0.676	0.479	2.23 (1.26–3.94)	0.006	2.43 (1.28–4.62)	0.007	CC	18 (53)/26 (22)	4.01 (1.79–8.97)	0.0007	4.87 (1.88–12.63)	0.001

Definition of abbreviations: AF = allele frequency; BPD = bronchopulmonary dysplasia; CI = confidence interval; OR = odds ratio; SNP = single nucleotide polymorphism.

*This table shows SNPs significantly associated with BPD in whites and Africans before and after adjustment for clinical covariates. Each SNP was tested under an additive model (risk allele effect) and a dominant model for the minor allele (risk genotype effect).

† Adjusted ORs were computed with a logistic regression model including clinical significant risk factors for BPD (i.e., birth weight, sex, persistent ductus arteriosus, postnatal sepsis, and the need for a second dose of surfactant).

‡ For rs1049269, genotyping failed for two white samples (one case and one control) and for five African samples (one case and four controls). For rs1245560, genotyping failed for three white samples (one case and two controls) and for three African samples (two cases and one control).

Regarding mild disease, 50% of included children had mild BPD. No significant association was found with the CC genotype of rs1245560 ($P = 0.63$) for mild BPD. C allele frequency was 47.9%, 49.2%, and 65.9% in infants with no BPD, mild BPD, and moderate-to-severe BPD, respectively, leading to *P* values of 0.75 for mild BPD versus no BPD and of 0.0003 for moderate-to-severe BPD versus no BPD for the risk allele effect. Those results strengthened the finding that rs1245560 was only associated with moderate-to-severe BPD.

Genotyping of rs1245560 and rs1049269 in the Finnish Population

In the Finnish replication population, rs1245560 CC and rs1049269 GG genotypes were significantly associated with an increased risk of BPD in univariate analysis (Table 2). After adjustment for available covariates, results remained significant with OR = 2.38 (95% CI, 1.15–4.98) for rs1245560 and OR = 2.10 (95% CI, 1.03–4.28) for rs1049269 (Table 2). Detailed genotype counts and frequencies are given in Table E4.

SPOCK2 Expression in Rat Lung

mRNA expression studies showed that *SPOCK2* is expressed in the developing lung. *SPOCK2* mRNA in whole-lung tissue increased gradually after birth, reaching levels about 10-fold higher on Postnatal Day 18, remaining high until Postnatal Day 28, and returning to the neonatal level in adulthood (Figure 2A). *SPOCK2* was expressed in both fibroblasts and airway epithelial cells (AECs). In isolated fibroblasts, mRNA levels were low until Postnatal Day 7 and then increased sharply from Days 14–21, following a pattern similar to that observed in the whole lung (Figures 2A and 2C). No major change in *SPOCK2* expression was seen in AECs (Figure 2D). The SPOCK2 protein was strongly expressed throughout the extracellular matrix (Figure 3). When rat pups were exposed to hyperoxia for 5 days (P5–P10), *SPOCK2* gene expression increased by about 60% relative to untreated controls (Figure 2B).

DISCUSSION

Recent studies highlighted the contribution of genetic factors to BPD susceptibility (4–6). We conducted a GWAS based on DNA pooling in two different ethnic populations. *SPOCK2* was the only BPD-associated gene that emerged in all analyses. Fine-scale mapping of the *SPOCK2* region validated its association with BPD. These results were replicated in an independent external population for the two associated SNPs belonging to *SPOCK2*, with genotype effects in the same direction and of the same order of magnitude as in the discovery panels. In

TABLE 2. INDIVIDUAL GENOTYPING OF RS1245560 AND RS1049269 IN THE FINNISH POPULATION (55 CASES AND 158 CONTROLS)*

	Risk Allele Effect							Risk Genotype Effect					
SNP	Risk Allele	AF in Cases	AF in Controls	OR (95% CI)	*P* Value	Adjusted OR† (95% CI)	*P* Value	Risk Genotype	Genotype Counts (%) in Cases/Controls	OR (95% CI)	*P* Value	Adjusted OR† (95% CI)	*P* Value
rs1049269	G	0.564	0.494	1.77 (0.73–4.30)	0.2	1.84 (0.77–4.68)	0.2	GG	20 (36)/35 (22)	2.01 (1.03–3.91)	0.040	2.10 (1.03–4.28)	0.040
rs1245560	C	0.546	0.471	1.87 (0.76–4.59)	0.17	1.99 (0.78–5.09)	0.15	CC	19 (35)/30 (19)‡	2.30 (1.16–4.56)	0.017	2.38 (1.15–4.98)	0.019

Definition of abbreviations: AF = allele frequency; CI = confidence interval; OR = odds ratio; SNP = single nucleotide polymorphism.

*Each SNP was tested under an additive model (risk allele effect) and a dominant model for the minor allele (risk genotype effect).

† OR adjusted for birth weight, sex, and neonatal respiratory distress syndrome.

‡ For rs1245560, genotyping failed for one case and one control.

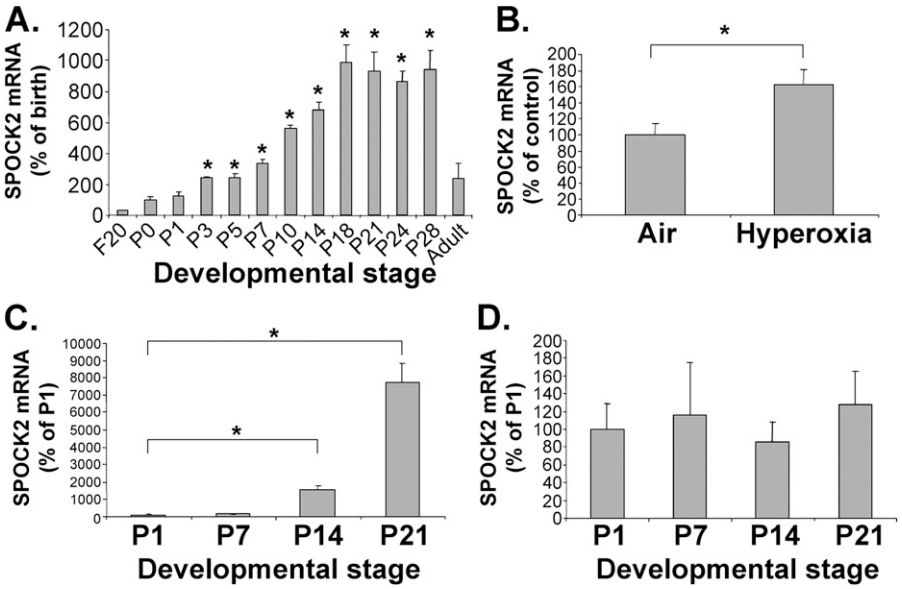

Figure 2. *SPOCK2* mRNA expression patterns. Developmental *SPOCK2* mRNA expression patterns in rat whole-lung tissue (*A*), lung fibroblasts (*C*), and airway epithelial cells (*D*). Expression was quantified by real-time polymerase chain reaction from fetal life to adulthood, in three to six individual lung samples per stage. The birth level in whole-lung tissue and the Day 1 level in isolated cells were arbitrarily attributed a value of 100. Values are mean ± SEM. * $P < 0.05$ versus reference level. (*B*) Changes in lung *SPOCK2* mRNA expression in newborn rats exposed to hyperoxia from Days 5–10. Values are mean ± SEM and results are expressed as a percentage of the control value. * $P < 0.05$.

addition, *SPOCK2* expression increased during normal rat lung alveolarization.

GWAS have identified a large number of robust associations between specific loci and complex human diseases. However, regarding genetic susceptibility to BPD, differences in clinical practice across neonatal intensive care units may hamper analyses, and environmental stressors must be tightly controlled to detect genetic associations in this setting. The relatively low rate of moderate to severe BPD (22%) in our high-risk study population suggests good control of these external factors. Furthermore, all potential risk factors for BPD were systematically collected in our study population, which allowed accurate multivariate analyses. We found that birth weight, male sex, persistent ductus arteriosus, and postnatal sepsis were independent risk factors for BPD in our population. These factors are fully concordant with those recently identified in an epidemiologic study including 3,629 subjects (25), arguing for the well representativeness of our population. The need for supplemental oxygen at 36 weeks postmenstrual age was chosen as the main outcome for analysis, and was determined with the same standardized test (20) in all the participating centers. Indeed, the relative contributions of genetic and environmental effects were demonstrated to depend on the severity of BPD (5). Variations in 28-day oxygen need-based BPD was previously shown to be attributable completely to environmental effects, whereas dependence on supplemental oxygen at 36 weeks seems to better reflect underlying genetic susceptibility (5). In agreement with these previous findings, the genetic factors identified in the present study were associated with moderate or severe BPD, but not with mild BPD. The Finnish population used for external validation included less premature neonates (gestational age <30 wk). However, BPD phenotype was also determined by the standardized oxygen reduction test, and the BPD rate (26%) was very similar as in our population. Although confounding factors were not all systematically collected, we were able to adjust results on birth weight, gestational age, and respiratory distress syndrome in this population.

SPOCK2 (SPARC/osteonectin, CWCV, and Kazal-like domains proteoglycan 2), also known as Testican-2, is a member of the testican group of extracellular chondroitin and heparan sulfate proteoglycans. *SPOCK2* was originally cloned from a human cDNA library (26). Its role has mainly been explored in the central nervous system (26, 27). It has been shown in mice that *SPOCK2* is expressed in lung (26). In humans, available data concern analyses of expressed sequence tag databases, which detect human SPOCK2 ESTs in brain, ovary, testis, retina, lung, prostate, and kidney (26). More recent studies evidenced hypermethylation of SPOCK2 regions in cell lines derived from human prostate, breast, and colon cancer (28, 29).

We noted that *SPOCK2* was expressed in the developing rat lung and that mRNA levels changed with stages of lung development. Alveolarization takes place between Days 4 and 21 in rats. We found that *SPOCK2* expression was low during the canalicular and saccular stages of rat lung development and started to increase significantly very close to the beginning of alveolarization. The highest expression was found on Day 18 (i.e., at termination of alveolarization) and remained elevated until Day 28, before falling to the neonatal level in adulthood. This pattern of expression may be consistent with the involvement of SPOCK2 in the progress of septation or in its termination,

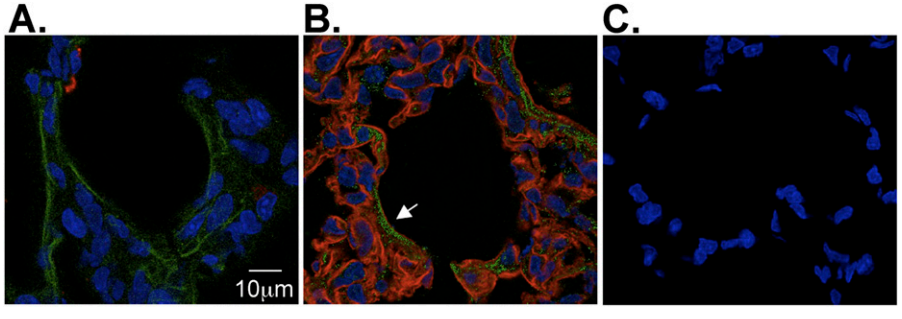

Figure 3. SPOCK2 immunofluorescence studies in rat lung tissue on postnatal Day 14. Immunocolocalization of SPOCK2 (*green labeling*) and SP-B (*red labeling* (*A*) or collagen IV (*red labeling*) (*B*) showed the presence of SPOCK2 throughout the extracellular matrix, including the basement membrane, as shown by superimposing the two fluorescent signals in some areas (*B, arrow*). (*C*) Negative control lung incubated with mouse and rabbit polyclonal IgG showing fluorescence restricted to the nucleus (original magnification ×126).

rather than in its triggering. Immunoflurorescent experiments confirmed the expression of SPOCK2 during lung development and the protein was found to be expressed throughout the extracellular matrix. Both fibroblasts and AECs may contribute to the expression of *SPOCK2* during lung development. We also found that *SPOCK2* mRNA expression increased in rat pups exposed to hyperoxia, which induces alveolar growth disorders in newborn rats (30). According to the hypothetical roles of SPOCK2 during alveolarization discussed previously, the hyperoxia-induced increase of SPOCK2 could be interpreted either as a deleterious effect triggering a premature termination of septation, or as a beneficial response attempting to counteract the effects of environmental injuries on lung development. Further work is required to determine the role of SPOCK2 in lung development.

SPOCK2 is known to interact with matrix metalloproteinase (MMP)-14 (31), whose key role in lung development has been previously established (32, 33). SPOCK2 interacts also with MMP-16, and we recently brought arguments for the role of this protease in BPD (7). In particular, we found a pattern of expression very similar to that observed here with *SPOCK2* (7).

The fine-scale mapping identified several SNPs associated with BPD, and in close linkage disequilibrium with rs1245560. Among these, the association of rs1049269 with BPD was replicated in the Finnish population. rs1049269 is located in the 3′UTR region of *SPOCK2* and could thus modify micro-RNA (miRNA) binding. MiRNAs are small noncoding RNAs known to regulate gene expression at the post-transcriptional level (34, 35). Computational analysis of *SPOCK2* 3′UTR region identified three miRNAs (mir-194*, mir-939, and mir-449b) for which binding to their target sequence could be altered by the polymorphic site rs1049269. Further experiments are required to explore the interaction between miRNAs and this polymorphism, and to determine the potential consequence on *SPOCK2* expression.

Other susceptibility genes have been previously suggested in candidate gene studies in BPD (36). Among these, two were identified in our genome-wide DNA pooling studies: *MMP16* and *L Selectin* were selected by the allele frequency difference method in the white series. However, none were confirmed by the combined Z-score analysis, or were selected in the African series.

Thus, a genome-wide association study of a multicenter population of preterm neonates identified *SPOCK2* as a new candidate susceptibility locus for BPD. Its pattern of expression during lung development points to a potential role in alveolarization.

Author Disclosure: None of the authors has a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

References

1. Baraldi E, Filippone M. Chronic lung disease after premature birth. *N Engl J Med* 2007;357:1946–1955.
2. Fanaroff AA, Stoll BJ, Wright LL, Carlo WA, Ehrenkranz RA, Stark AR, Bauer CR, Donovan EF, Korones SB, Laptook AR, *et al*. Trends in neonatal morbidity and mortality for very low birthweight infants. *Am J Obstet Gynecol* 2007;196:147, e141–148.
3. Coalson JJ. Pathology of bronchopulmonary dysplasia. *Semin Perinatol* 2006;30:179–184.
4. Bhandari V, Bizzarro MJ, Shetty A, Zhong X, Page GP, Zhang H, Ment LR, Gruen JR. Familial and genetic susceptibility to major neonatal morbidities in preterm twins. *Pediatrics* 2006;117:1901–1906.
5. Lavoie PM, Pham C, Jang KL. Heritability of bronchopulmonary dysplasia, defined according to the consensus statement of the National Institutes of Health. *Pediatrics* 2008;122:479–485.
6. Parker RA, Lindstrom DP, Cotton RB. Evidence from twin study implies possible genetic susceptibility to bronchopulmonary dysplasia. *Semin Perinatol* 1996;20:206–209.
7. Hadchouel A, Decobert F, Franco-Montoya ML, Halphen I, Jarreau PH, Boucherat O, Martin E, Benachi A, Amselem S, Bourbon J, *et al*. Matrix metalloproteinase gene polymorphisms and bronchopulmonary dysplasia: identification of MMP as a new player in lung development. *PLoS ONE* 2008;3:e3188.
8. Kazzi SN, Kim UO, Quasney MW, Buhimschi I. Polymorphism of tumor necrosis factor-alpha and risk and severity of bronchopulmonary dysplasia among very low birth weight infants. *Pediatrics* 2004;114: e243–e248.
9. Kwinta P, Bik-Multanowski M, Mitkowska Z, Tomasik T, Legutko M, Pietrzyk JJ. Genetic risk factors of bronchopulmonary dysplasia. *Pediatr Res* 2008;64:682–688.
10. Manar MH, Brown MR, Gauthier TW, Brown LA. Association of glutathione-s-transferase-p1 (GST-P1) polymorphisms with bronchopulmonary dysplasia. *J Perinatol* 2004;24:30–35.
11. Rova M, Haataja R, Marttila R, Ollikainen V, Tammela O, Hallman M. Data mining and multiparameter analysis of lung surfactant protein genes in bronchopulmonary dysplasia. *Hum Mol Genet* 2004;13:1095–1104.
12. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, Brun M, Szelinger S, Coon KD, Zismann VL, *et al*. Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. *Am J Hum Genet* 2007;80:126–139.
13. Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;2:91–99.
14. Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000;405:847–856.
15. Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O'Donovan MC, *et al*. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC Med Genomics* 2008;1:44.
16. Dunckley T, Huentelman MJ, Craig DW, Pearson JV, Szelinger S, Joshipura K, Halperin RF, Stamper C, Jensen KR, Letizia D, *et al*. Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N Engl J Med* 2007;357:775–788.
17. Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, *et al*. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet* 2008;4:e28.
18. Steer S, Abkevich V, Gutin A, Cordell HJ, Gendall KL, Merriman ME, Rodger RA, Rowley KA, Chapman P, Gow P, *et al*. Genomic DNA pooling for whole-genome association scans in complex disease: empirical demonstration of efficacy in rheumatoid arthritis. *Genes Immun* 2007;8:57–68.
19. Craig DW, Huentelman MJ, Hu-Lince D, Zismann VL, Kruer MC, Lee AM, Puffenberger EG, Pearson JM, Stephan DA. Identification of disease causing loci using an array-based genotyping approach on pooled DNA. *BMC Genomics* 2005;6:138.
20. Walsh MC, Yao Q, Gettner P, Hale E, Collins M, Hensman A, Everette R, Peters N, Miller N, Muran G, *et al*. Impact of a physiologic definition on bronchopulmonary dysplasia rates. *Pediatrics* 2004;114: 1305–1311.
21. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 2008;32:227–234.
22. Welcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678.
23. Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med* 2009;360:1759–1768.
24. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 2005;95:221–227.
25. Laughon MM, Langer JC, Bose CL, Smith PB, Ambalavanan N, Kennedy KA, Stoll BJ, Buchter S, Laptook AR, Ehrenkranz RA, *et al*. Prediction of bronchopulmonary dysplasia by postnatal age in extremely premature infants. *Am J Respir Crit Care Med* 2011;183: 1715–1722.
26. Vannahme C, Schubel S, Herud M, Gosling S, Hulsmann H, Paulsson M, Hartmann U, Maurer P. Molecular cloning of testican-2: defining a novel calcium-binding proteoglycan family expressed in brain. *J Neurochem* 1999;73:12–20.

27. Schnepp A, Komp Lindgren P, Hulsmann H, Kroger S, Paulsson M, Hartmann U. Mouse testican-2. Expression, glycosylation, and effects on neurite outgrowth. *J Biol Chem* 2005;280:11274–11280.
28. Chung W, Kwabi-Addo B, Ittmann M, Jelinek J, Shen L, Yu Y, Issa JP. Identification of novel tumor markers in prostate, colon and breast cancer by unbiased methylation profiling. *PLoS ONE* 2008;3:e2079.
29. Nordgard SH, Johansen FE, Alnaes GI, Bucher E, Syvanen AC, Naume B, Borresen-Dale AL, Kristensen VN. Genome-wide analysis identifies 16q deletion associated with survival, molecular subtypes, mRNA expression, and germline haplotypes in breast cancer patients. *Genes Chromosomes Cancer* 2008;47:680–696.
30. Boucherat O, Franco-Montoya ML, Thibault C, Incitti R, Chailley-Heu B, Delacourt C, Bourbon JR. Gene expression profiling in lung fibroblasts reveals new players in alveolarization. *Physiol Genomics* 2007;32:128–141.
31. Nakada M, Yamada A, Takino T, Miyamori H, Takahashi T, Yamashita J, Sato H. Suppression of membrane-type 1 matrix metalloproteinase (MMP)-mediated MMP-2 activation and tumor invasion by testican 3 and its splicing variant gene product, n-tes. *Cancer Res* 2001;61:8896–8902.
32. Atkinson JJ, Holmbeck K, Yamada S, Birkedal-Hansen H, Parks WC, Senior RM. Membrane-type 1 matrix metalloproteinase is required for normal alveolar development. *Dev Dyn* 2005;232:1079–1090.
33. Boucherat O, Bourbon JR, Barlier-Mur AM, Chailley-Heu B, D'Ortho MP, Delacourt C. Differential expression of matrix metalloproteinases and inhibitors in developing rat lung mesenchymal and epithelial cells. *Pediatr Res* 2007;62:20–25.
34. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by micrornas: are the answers in sight? *Nat Rev Genet* 2008;9:102–114.
35. Rana TM. Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol* 2007;8:23–36.
36. Lavoie PM, Dube MP. Genetics of bronchopulmonary dysplasia in the age of genomics. *Curr Opin Pediatr* 2010;22:134–138.