

# Copy Number Variation in Bronchopulmonary Dysplasia

Thomas J. Hoffmann,<sup>1</sup> Gary M. Shaw,<sup>2</sup> David K. Stevenson,<sup>2</sup> Hui Wang,<sup>2</sup> Cecele C. Quaintance,<sup>2</sup> John Oehlert,<sup>2</sup> Laura L. Jelliffe-Pawlowski,<sup>3,4</sup> Jeffrey B. Gould,<sup>2,5</sup> John S. Witte,<sup>1</sup> and Hugh M. O'Brodovich<sup>2\*</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics and Institute for Human Genetics, University of California San Francisco, San Francisco, California

<sup>2</sup>Department of Pediatrics, Stanford University School of Medicine, Palo Alto, California

<sup>3</sup>California Genetic Disease Screening Program of the California Department of Public Health, Richmond, California

<sup>4</sup>Department of Epidemiology and Biostatistics, Division of Preventive Medicine and Public Health, University of California San Francisco School of Medicine, San Francisco, California

<sup>5</sup>California Perinatal Quality Care Collaborative, Stanford, California

Manuscript Received: 18 February 2014; Manuscript Accepted: 28 May 2014

## TO THE EDITOR:

Two twin studies [Bhandari et al., 2006; Lavoie et al., 2008] have found relatively high heritability (53–79%) of susceptibility to bronchopulmonary dysplasia (BPD), a severe disorder of the pulmonary and cardiovascular systems in very low birth weight (VLBW) infants. To identify genetic factors underlying BPD we carried out a California wide population-based case-control study ( $n = 1,726$ ) of  $>2$  million genome-wide markers. We recently reported findings from analyzing the association between individual single nucleotide polymorphisms (SNPs) and BPD in this study [Wang et al., 2013]. Here we evaluate the potential relationship between copy number variants (CNVs) and BPD, which is important since CNVs have been associated with other suspected heritable disorders (e.g., autism [Glessner et al., 2009]).

Our case-control study identified singleton VLBW infant births from the California Perinatal Quality Care Collaborative (CPQCC, <http://www.cpqcc.org/>) [Gould, 2010], which represents more than 90% of all NICU admissions in California. More detailed methodology is described in our previous publication [Wang et al., 2013]. In brief, inclusion criteria were gestational age (GA) 25<sup>0</sup>–29<sup>6/7</sup> weeks, birth weight (BW)  $<1,500$  grams, and  $\geq 3$  days mechanical ventilation during their hospitalization up to 36 weeks postmenstrual age (PMA). The  $\geq 3$  days mechanical ventilation was an inclusion criterion so that both cases and controls would be exposed to this “environmental” factor. NIH/NICHD criteria were used to diagnose mild (supplemental oxygen at 28 days after birth but not at 36 weeks PMA, moderate (supplemental oxygen  $<30\%$ ) and severe (supplemental oxygen  $>30\%$  or positive pressure support) BPD [Jobe and Bancalari, 2001; Walsh et al., 2004]. BPD cases were defined as infants requiring supplemental oxygen or positive pressure ventilator support at 36 weeks PMA whereas control infants were breathing room air at 36 weeks PMA. The practices of each NICU determined the need for supplemental oxygen and physiologic assessments [Walsh et al., 2004] were not routinely carried out.

### How to Cite this Article:

Hoffmann TJ, Shaw GM, Stevenson DK, Wang H, Quaintance CC, Oehlert J, Jelliffe-Pawlowski LL, Gould JB, Witte JS, O'Brodovich HM. 2014. Copy number variation in bronchopulmonary dysplasia. *Am J Med Genet Part A* 164A:2672–2675.

Exclusion criteria included multiple birth, major congenital abnormalities, major surgery (patent ductus arteriosus (PDA) ligation was not excluded), infant death or left hospital prior to 36 weeks PMA, or supplemental oxygen status at 36 weeks PMA not known.

Infants were linked to their newborn screening blood spot. Genomic DNA was extracted from bloodspots [St. Julien et al., 2013] and genotyped (Illumina HumanOmni2.5 beadchip, San Diego, CA). Non-amplified DNA was used and the genotype

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made. Conflicts of interest: The authors have no conflict of interest to declare. Grant sponsor: NIH/NHLBI; Grant number: RC2 HL101748; Grant sponsor: National Heart and Lung Institute and the March of Dimes Prematurity Research Center at Stanford University School of Medicine. \*Correspondence to:

Hugh O'Brodovich, M.D., Department of Pediatrics, Stanford University, 300 Pasteur Dr. (Room H310), Stanford CA 94305.

E-mail: [hugho@stanford.edu](mailto:hugho@stanford.edu)

Article first published online in Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com)): 26 June 2014

DOI 10.1002/ajmg.a.36659

calls were made using GenomeStudio software [Illumina, 2011] after quality control (QC) procedures [Wang et al., 2013]. After the above steps we successfully genotyped 899 BPD cases and 827 controls ( $n = 1,726$ ). The Institutional Review Board (IRB) 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.

For covariates and potential confounders, analyses of our data and findings from the literature indicate that both sex and BW are strong predictors of BPD. To address possible bias due to population stratification, we estimated genetic ancestry using a principal components (PCs) analysis [Wang et al., 2013]. We included in regression analyses of CNVs the first three PCs, self-reported ethnicity, sex, and BW to control for their potential confounding of associations with BPD.

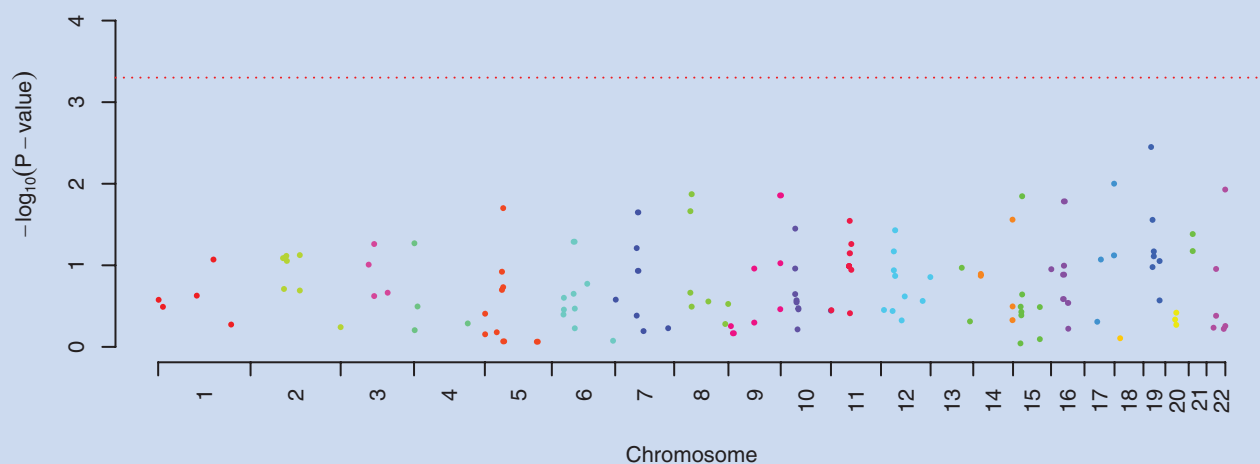
After removing individuals that had incorrectly called information from one data plate ( $n = 59$ ) and one additional individual who was missing substantial probe intensity information, we analyzed CNV data from 1,666 infants (866 BPD cases and 800 controls). We called CNVs via the software PennCNV [Wang et al., 2007] using GC-wave factor adjustment. We used the following established QC procedures for identifying CNVs. We first removed CNVs that had fewer than 10 SNPs or that were shorter than 50 kb in length, and merged large CNVs with a gap between them which was less than 20% of their length. We then removed poor quality samples that had standard deviation of normalized intensity (LRR)  $> 0.35$ , B Allele Frequency (BAF) drift  $> 0.01$ , number of CNVs  $> 80$ , or GC-wave factor (WF)  $> 0.05$ . These QC criteria are similar to those previously applied by others when using PennCNV [Glessner et al., 2009, 2010; Need et al., 2009; Davis et al., 2011].

After these steps a total of 21,399 CNVs were called for 1,631 individuals (848 BPD cases and 783 controls). Overall there was an average of 13.1 CNVs per infant, which was similar between the BPD cases (13.0) and controls (13.2). A formal test indicated no association between the logarithm of the number of CNVs and

BPD ( $P = 0.998$ , from logistic regression, controlling for ethnicity, sex, BW, and three PCs of ancestry). For BPD cases, CNVs were a median 95.4 kb length (range 50.0–11,308.7 kb) and median 41 SNPs (range 10–4,401 SNPs); for controls, CNVs were a median 96.4 kb length (range 50.0–37,391 kb) and median 41 SNPs (range 10–30,938 SNPs). The CNVs comprised four different copy number states: homozygous deletions ( $CN = 0$ ); hemizygous deletions ( $CN = 1$ ); and two possible duplication states ( $CN = 3$  or  $4$ ). The case/control ratios of the frequency of these copy number states, tested by dichotomizing at greater than the median, were: homozygous deletions, (27 observed in BPD cases/848 cases)/(18 in controls/783 controls) = 1.38 ( $P = 0.58$ ); hemizygous deletions, (5,376/848)/(5,340/783) = 0.93 ( $P = 0.54$ ); duplications ( $CN = 3$ ), (5,579/848)/(4,938/783) = 1.04 ( $P = 0.85$ ); and duplications ( $CN = 4$ ), (74/848)/(49/783) = 1.39 ( $P = 0.25$ ).

Focusing on specific CNVs, we evaluated the association between BPD and each SNP, evaluating separately for each SNP as a deletion ( $CN = 0$ ,  $CN = 1$ ,  $CN = \text{other}$ ) and a duplication ( $CN = 4$ ,  $CN = 3$ ,  $CN = \text{other}$ ) as in Glessner et al. [2013] requiring a count of at least 20 in the minor homozygote plus heterozygote. For each SNP, we evaluated the deletion/duplication frequency between BPD cases and controls with a logistic regression of BPD adjusted for the deletion/duplication, GA, sex, BW,  $PC_1$ ,  $PC_2$ , and  $PC_3$ . Then, because of boundary truncation problems, we defined CNV regions (CNVRs) which collapsed SNPs within a 1 MB distance of each other if they were of comparable significance ( $\pm \log_{10} P\text{-value}$ ) as in Glessner et al. [2013], forming 74 deletions and 57 duplications. The CNVR results are shown in the Manhattan plot in Figure 1, and the top 5 CNVR results are shown in Table I. No SNP reached the suggested multiple-testing correction for CNVRs of  $5 \times 10^{-4}$  [Glessner et al., 2013].

Thus, we found no evidence that BPD cases have a larger number of CNVs than controls. We also did not observe particular CNVs that were significantly associated with an increased risk of BPD. The top CNVRs were in different regions of the genome than the top associated SNPs we found in our previous GWAS results. These



**FIG. 1.** Manhattan plot of the association of bronchopulmonary dysplasia (BPD) and each copy number variant region (CNVR). The dotted line at 0.0005 indicates the genome-wide CNVR significance level.

TABLE I. Top Five Copy Number Variant Region (CNVR) Associated with BPD

CNVR	Gene	Type	Case Freq.	Cont Freq.	OR	P
chr19: 20601335–20715233	<i>ZNF826P</i>	Del	0.019	0.029	1.71 [1.06, 2.78]	0.0035
chr17: 79925150–79986974	<i>ASPSCR1, LRRC45, STRA13</i>	Dup	0.011	0.0026	0.23 [0.08, 0.70]	0.01
chr22: 50302618–50458513	<i>CRELD2, ALG12, PIM3, IL17REL</i>	Del	0.0077	0.0051	0.67 [0.26, 1.69]	0.012
chr8: 46842124–47467663	<i>POTEA—LINC00293</i>	Dup	0.025	0.022	0.89 [0.55, 1.45]	0.013
chr9: 139243790–139273402	<i>CARD9, GPSM1, DNLZ, SNAPC4</i>	Del	0.0029	0.01	3.37 [1.18, 9.66]	0.014

Del, deletion; Dup, duplication.  
None reach genome-wide CNVR significance (0.0005). Genes are given when contained in the CNVR region, and a range between two genes if located in between genes (e.g., POTEA—LINC00293).

results, similar to our previous GWAS results, do not point to particular genomic loci as the explanation for the previously described heritability for BPD. This may be due to several reasons, as has been discussed [Wang et al., 2013]. First our study population differs from the twin studies that reported high heritability [Bhandari et al., 2006; Lavoie et al., 2008]. These twin studies did not report the race/ethnicity of the patients, though we speculate that they were Caucasian given the geographic location [Bhandari et al., 2006; Lavoie et al., 2008], in contrast to our cases and controls of predominantly Mexican-Hispanic origin [Wang et al., 2013]. This genetic heterogeneity may reduce power, as different race/ethnicity groups have different prevalence of BPD. Thirdly, our eligibility of cases and controls required  $\geq 3$  days mechanical ventilation, which not all studies have used. This was chosen to better define the BPD phenotype and decrease the “environmental” differences between cases and controls, as we hoped it would improve our ability to detect genetic factors. However, extremely premature infants who did not require mechanical ventilation sometimes have BPD. Lastly, there may have been unknown differences between the NICUs.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Dr. Richard Bland for his contributions in writing the grant to obtaining 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 detailed 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

Bhandari V, Bizzarro MJ, Shetty A, Zhong X, Page GP, Zhang H, Ment LR, Gruen JR. 2006. Familial and genetic susceptibility to major neonatal morbidities in preterm twins. *Pediatrics* 117:1901–1906.

Davis LK, Meyer KJ, Schindler EI, Beck JS, Rudd DS, Grundstad AJ, Scheetz TE, Braun TA, Fingert JH, Alward WLM, Kwon YH, Folk JC, Russell SR, Wassink TH, Sheffield VC, Stone EM. 2011. Copy number variations and primary open-angle glaucoma. *Invest Ophthalmol Vis Sci* 52:7122–7133.

Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, Imielinski M, Frackelton EC, Reichert J, Crawford EL, Munson J, Sleiman PMA, Chiavacci R, Annaiah K, Thomas K, Hou C, Glaberson W, Flory J, Otieno F, Garriss M, Soorya L, Klei L, Piven J, Meyer KJ, Anagnostou E, Sakurai T, Game RM, Rudd DS, Zurawiecki D, McDougale CJ, Davis LK, Miller J, Posey DJ, Michaels S, Kolevzon A, Silverman JM, Bernier R, Levy SE, Schultz RT, Dawson G, Owley T, McMahon WM, Wassink TH, Sweeney JA, Nurnberger JL, Coon H, Sutcliffe JS, Minshew NJ, Grant SFA, Bucan M, Cook EH, Buxbaum JD, Devlin B, Schellenberg GD, Hakonarson H. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459:569–573.

Glessner JT, Reilly MP, Kim CE, Takahashi N, Albano A, Hou C, Bradfield JP, Zhang H, Sleiman PMA, Flory JH, Imielinski M, Frackelton EC, Chiavacci R, Thomas KA, Garriss M, Otieno FG, Davidson M, Weiser M, Reichenberg A, Davis KL, Friedman JL, Cappola TP, Margulies KB, Rader DJ, Grant SFA, Buxbaum JD, Gur RE, Hakonarson H. 2010. Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc Natl Acad Sci USA* 107:10584–10589.

Glessner JT, Li J, Hakonarson H. 2013. ParseCNV integrative copy number variation association software with quality tracking. *Nucleic Acids Res* 41:e64–e64.

Gould JB. 2010. The role of regional collaboratives: The California Perinatal Quality Care Collaborative Model. *Clin Perinatol* 37:71–86.

Illumina. 2011. GenomeStudio Data Analysis Software [Whitepaper]. Retrieved from [http://res.illumina.com/documents/products/data-sheets/datasheet\\_genomestudio\\_software.pdf](http://res.illumina.com/documents/products/data-sheets/datasheet_genomestudio_software.pdf).

Jobe AH, Bancalari E. 2001. Bronchopulmonary dysplasia. *Am J Respir Crit Care Med* 163:1723–1729.

Lavoie PM, Pham C, Jang KL. 2008. Heritability of bronchopulmonary dysplasia, defined according to the consensus statement of the national institutes of health. *Pediatrics* 122:479–485.

Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen EL, Shianna KV, Yoon W, Kasperavičiūtė D, Gennarelli M, Strittmatter WJ, Bonvicini C, Rossi G, Jayatilake K, Cola PA, McEvoy JP, Keefe RSE, Fisher EMC, St. Jean PL, Giegling I, Hartmann AM, Möller H-J, Ruppert A, Fraser G, Crombie C, Middleton LT, St. Clair D, Roses AD, Muglia P, Francks C, Rujescu D, Meltzer HY, Goldstein DB. 2009. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet* 5:e1000373.

St. Julien KR, Jelliffe-Pawlowski LL, Shaw GM, Stevenson DK, O’Brodivich HM, Krasnow MA, the Stanford BPD Study Group. 2013. High quality genome-wide genotyping from archived dried blood spots without DNA amplification. *PLoS ONE* 8:e64710.

Walsh MC, Yao Q, Gettner P, Hale E, Collins M, Hensman A, Everette R, Peters N, Miller N, Muran G, Auten K, Newman N, Rowan G, Grisby C,

- Arnell K, Miller L, Ball B, McDavid G. 2004. Impact of a physiologic definition on bronchopulmonary dysplasia rates. *Pediatrics* 114:1305–1311.
- Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, Hakonarson H, Bucan M. 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17:1665–1674.
- Wang H, St. Julien KR, Stevenson DK, Hoffmann TJ, Witte JS, Lazzeroni LC, Krasnow MA, Quaintance CC, Oehlert JW, Jelliffe-Pawlowski LL, Gould JB, Shaw GM, O’Brodoovich HM. 2013. A genome-wide association study (GWAS) for bronchopulmonary dysplasia. *Pediatrics* 132:290–297.