Sequencing Analysis at 8p23 Identifies Multiple Rare Variants in *DLC1* Associated with Sleep-Related Oxyhemoglobin Saturation Level

Jingjing Liang,[1] Brian E. Cade,[2,3,4] Karen Y. He,[1] Heming Wang,[2,3,4] Jiwon Lee,[2] Tamar Sofer,[2,3] Stephanie Williams,[2,3] Ruitong Li,[2,3] Han Chen,[5,6] Daniel J. Gottlieb,[2,3,7] Daniel S. Evans,[8] Xiuqing Guo,[9,10] Sina A. Gharib,[11] Lauren Hale,[12] David R. Hillman,[13] Pamela L. Lutsey,[14] Sutapa Mukherjee,[15,16] Heather M. Ochs-Balcom,[17] Lyle J. Palmer,[18] Jessica Rhodes,[4,19,20] Shaun Purcell,[2,3,4] Sanjay R. Patel,[21] Richa Saxena,[2,3,4,19,20] Katie L. Stone,[8] Weihong Tang,[22] Gregory J. Tranah,[8] Eric Boerwinkle,[5,23] Xihong Lin,[24] Yongmei Liu,[25] Bruce M. Psaty,[26,27]

(Author list continued on next page)

Average arterial oxyhemoglobin saturation during sleep ($AvSpO_2S$) is a clinically relevant measure of physiological stress associated with sleep-disordered breathing, and this measure predicts incident cardiovascular disease and mortality. Using high-depth whole-genome sequencing data from the National Heart, Lung, and Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) project and focusing on genes with linkage evidence on chromosome 8p23,[1,2] we observed that six coding and 51 noncoding variants in a gene that encodes the GTPase-activating protein (*DLC1*) are significantly associated with $AvSpO_2S$ and replicated in independent subjects. The combined *DLC1* association evidence of discovery and replication cohorts reaches genome-wide significance in European Americans ($p = 7.9 \times 10^{-7}$). A risk score for these variants, built on an independent dataset, explains 0.97% of the $AvSpO_2S$ variation and contributes to the linkage evidence. The 51 noncoding variants are enriched in regulatory features in a human lung fibroblast cell line and contribute to *DLC1* expression variation. Mendelian randomization analysis using these variants indicates a significant causal effect of *DLC1* expression in fibroblasts on $AvSpO_2S$. Multiple sources of information, including genetic variants, gene expression, and methylation, consistently suggest that *DLC1* is a gene associated with $AvSpO_2S$.

Arterial oxyhemoglobin saturation (SpO_2) reflects the adequacy of ventilation and oxygen transport, fundamental physiological properties that are tightly regulated at molecular and cellular levels to ensure delivery of oxygen to vital tissues. Reductions in oxyhemoglobin saturation lead to increased rates of mortality and cognitive decline.[3] Given its clinical relevance, oxygen saturation is commonly monitored in patients with pulmonary, cardiac, and sleep

[1]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA; [2]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA 02115, USA; [3]Division of Sleep Medicine, Harvard Medical School, Boston, MA, 02115, USA; [4]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA; [5]Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [6]Center for Precision Health, School of Public Health and School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [7]VA Boston Healthcare System, Boston, MA 02132, USA; [8]California Pacific Medical Center Research Institute, San Francisco, CA 94107, USA; [9]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor—UCLA Medical Center, Torrance, CA 90509, USA; [10]Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor—UCLA Medical Center, Torrance, CA 90509, USA; [11]Department of Medicine, Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, University of Washington, Seattle, WA 98195, USA; [12]Family, Population, and Preventive Medicine, Program in Public Health, Stony Brook University School of Medicine, Stony Brook, NY 11794, USA; [13]Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Perth, Western Australia 6009, Australia; [14]Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA; [15]Sleep Health Service, Respiratory and Sleep Service, Southern Adelaide Local Health Network, Adelaide, South Australia 5042, Australia; [16]Adelaide Institute for Sleep Health, Flinders University, Adelaide, South Australia 5042, Australia; [17]Department of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo, Buffalo, NY 14214, USA; [18]School of Public Health, University of Adelaide, South Australia 5000, Australia; [19]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA[19]Center for Genomic Medicine and Department of Anesthesia, Pain and Critical Care Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; [20]Department of Anesthesia, Pain and Critical Care Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; [21]Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA; [22]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55454, USA; [23]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [24]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA; [25]Department of Medicine, Division of Cardiology, Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC 27710, USA; [26]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA 98101, USA; [27]Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, USA; [28]Framingham Heart Study, Framingham, MA 01702, USA; [29]Section of Preventive Medicine and Epidemiology, Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA; [30]Section Cardiology, Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA; [31]Department of Epidemiology, Boston University School of Public Health, Boston, MA 02118, USA; [32]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA; [33]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard

(Affiliations continued on next page)

disorders in order to identify those at risk for adverse outcomes and to assess the success of therapy. Average SpO_2 during sleep ($AvSpO_2S$) is heritable (h2 = 0.41)[2] and can be reliably and relatively easily measured with a finger-placed pulse oximeter. Studying the genetic underpinnings of $AvSpO_2S$ can help elucidate the bases for variation in hypoxemia-related stresses and may ultimately explain differences in susceptibility to many sleep-disordered breathing (SDB)-related morbidities.[4,5] This information may also inform underlying susceptibility to hypoxemia in the setting of lung injury or disease.[6–9] Here we present an analytical approach based on a strategy that integrates linkage and whole-genome sequencing (WGS) analysis, complemented with gene expression and methylation data (Figure 1), and aims to increase statistical power to identify rare variants. Many disease variants have been identified through linkage analysis;[10–12] however, the utility of linkage information in combination with WGS data for complex traits has not been evaluated.

A prior linkage analysis of $AvSpO_2S$ conducted in 617 European American (EA) individuals from 132 families in the Cleveland Family Study (CFS) identified a significant linkage peak on chromosome 8p23.[1] These subjects were further genotyped using the Illumina Human OmniExpress+Exome chip. Because $AvSpO_2S$ was skewed distributed (Table S1), rank normal transformation was applied using the R package "RNOmni" for $AvSpO_2S$ in all analyses, which included linkage and association analyses. Linkage analysis based on the Illumina Human OmniExpress+Exome chip data showed persistence of linkage evidence, with LOD scores 2.56 and 3.28 with and without including body mass index (BMI) as a covariate, respectively (Figure 2A and Figure S2). In the linkage analysis, we always included gender, age, and age × age as covariates. To identify variants that are independent of BMI, a trait correlated with several SDB traits, we focused this analysis on $AvSpO_2S$ adjusted for BMI, gender, age, and age × age.

We estimated family specific LOD scores (fsLOD) in the CFS families in the linkage analysis. We took the top 18 families with fsLOD $\geq$0.1 as those who potentially carry low-frequency or rare $AvSpO_2S$ variants. Our simulations suggested that using threshold fsLOD $\geq$0.1 did not inflate the type I error in association analyses (Table S1), and that this threshold has either comparable or better power than no threshold (Tables S2 and S3). This threshold is consistent with an estimated mixture model of two normal distributions (see Supplemental Data, Figure S5). 487 CFS EAs were sequenced through TOPMed, and their average sequencing depth was 38× (Table S5). We observed 212,282 variants that had a minor allele frequency (MAF) <0.05 and that passed quality control filters in this linkage region in the CFS EUs. We hypothesized that low-frequency and rare variants in protein coding genes are both more likely to have functional roles and to contribute to the observed linkage evidence, and thus are more likely to focus on the variants located in the 105 genes or their corresponding 5 kbps regions upstream and downstream (Figure 2A). Further, to search for variants that could potentially account for the observed linkage evidence, we filtered out variants that only presented at most once in any of the 18 selected families, thus reducing the number of variants to 20,168. We filtered out the genes with only one variant because of those genes' low statistical power. Among the 105 genes, 20 had at least two such variants that were also functional coding defined as missense, in-frame deletion or insertion, stop gained or lost, start gained or lost, splice acceptor or donor, or initiator or start codon (Table S6). For the remaining non-coding variants, we applied the CADD PHRED score,[13] which estimates the likely impact on encoded protein and variant deleterious metrics. We used a CADD score >10 as a threshold to filter the variants; this resulted in 709 variants distributed across 48 genes (Table S6). Both gene-based burden and SKAT tests[14,15] were performed in the CFS EA cohort for each gene; these tests analyzed functional coding and non-coding variants separately. We did not observe an inflated type I error rate when we performed linkage and association analyses in the same dataset (see Supplemental Data, Table S1), and this result is consistent with the independence of linkage and association information when there are no trait-associated genes in the linkage region.[16] To determine which genes to carry forward to Stage II in the Stage I analysis, we applied the empirical p = 0.05 threshold calculated based on the p values obtained through testing genes across the genome after the same filters were applied for the functional coding and non-coding variants. This empirical threshold can be conservative given so many

Medical School, Boston, MA 02115, USA; [34]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA; [35]Department of Public Health Sciences, Biostatistics Section, University of Virginia, Charlottesville, VA 22908, USA; [36]Department of Medicine, Columbia University Medical Center, New York, NY 10032, USA; [37]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA; [38]A list of members and affiliations appears in the Supplemental Data; [39]A list of members and affiliations appears in the Supplemental Data; [40]Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA
*Correspondence: sredline@bwh.harvard.edu (S.R.), xxz10@case.edu (X.Z.)
https://doi.org/10.1016/j.ajhg.2019.10.002.

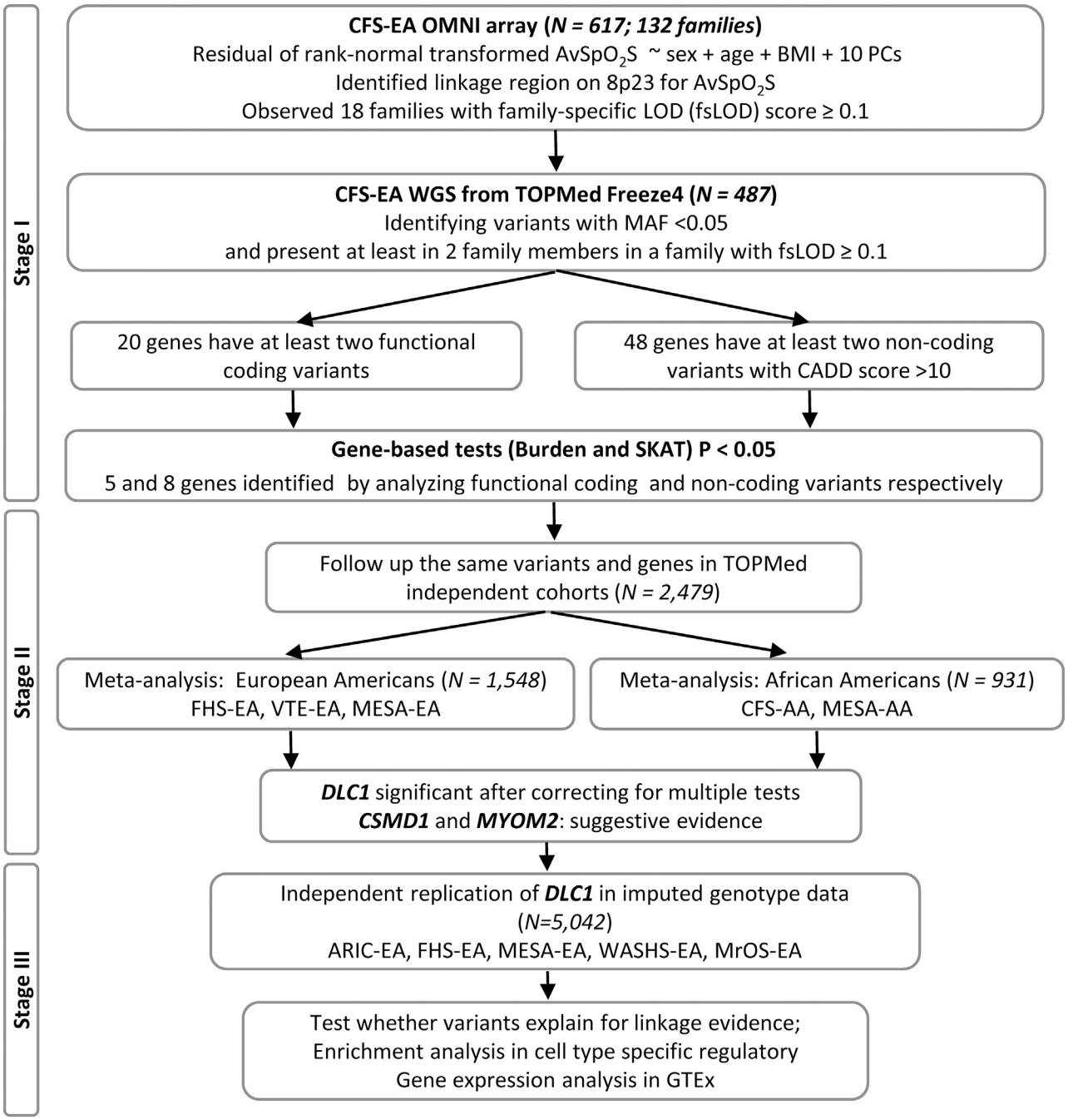

Figure 1. Analysis Flow Chart for Searching Low-Frequency and Rare Variants Associated with $AvSpO_2S$

genes likely contribute to $AvSpO_2S$ (see Supplemental Data). Five and eight genes in the functional coding and non-coding variant analyses, respectively, had empirical p values less than 0.05 (Table 1), whereas we would have expected two genes (for 20 genes and two tests) and five genes (for 48 genes and two tests) in the functional coding and non-coding variant analyses, respectively, by chance. Our results thus indicate an enrichment of genes associated with $AvSpO_2S$ under the linkage peak. The *DLC1* (MIM: 604258) gene was the only gene observed in both functional coding and non-coding variant analyses. When using different thresholds of fsLOD (0.05) and MAF (0.01), we observed no appreciable change in results (Table S4).

The TOPMed WGS project included an additional six cohorts consisting of 2,772 EAs, 1,726 African Americans (AAs), and 2,795 Hispanic Americans (HAs) (Table S5) whose genomes were sequenced and who had $AvSpO_2S$ measured. To reduce the multiple comparison penalty, we performed the same burden and SKAT analyses in these samples, focusing on the same variants in the 12 genes identified in CFS Stage I analysis. We performed the burden

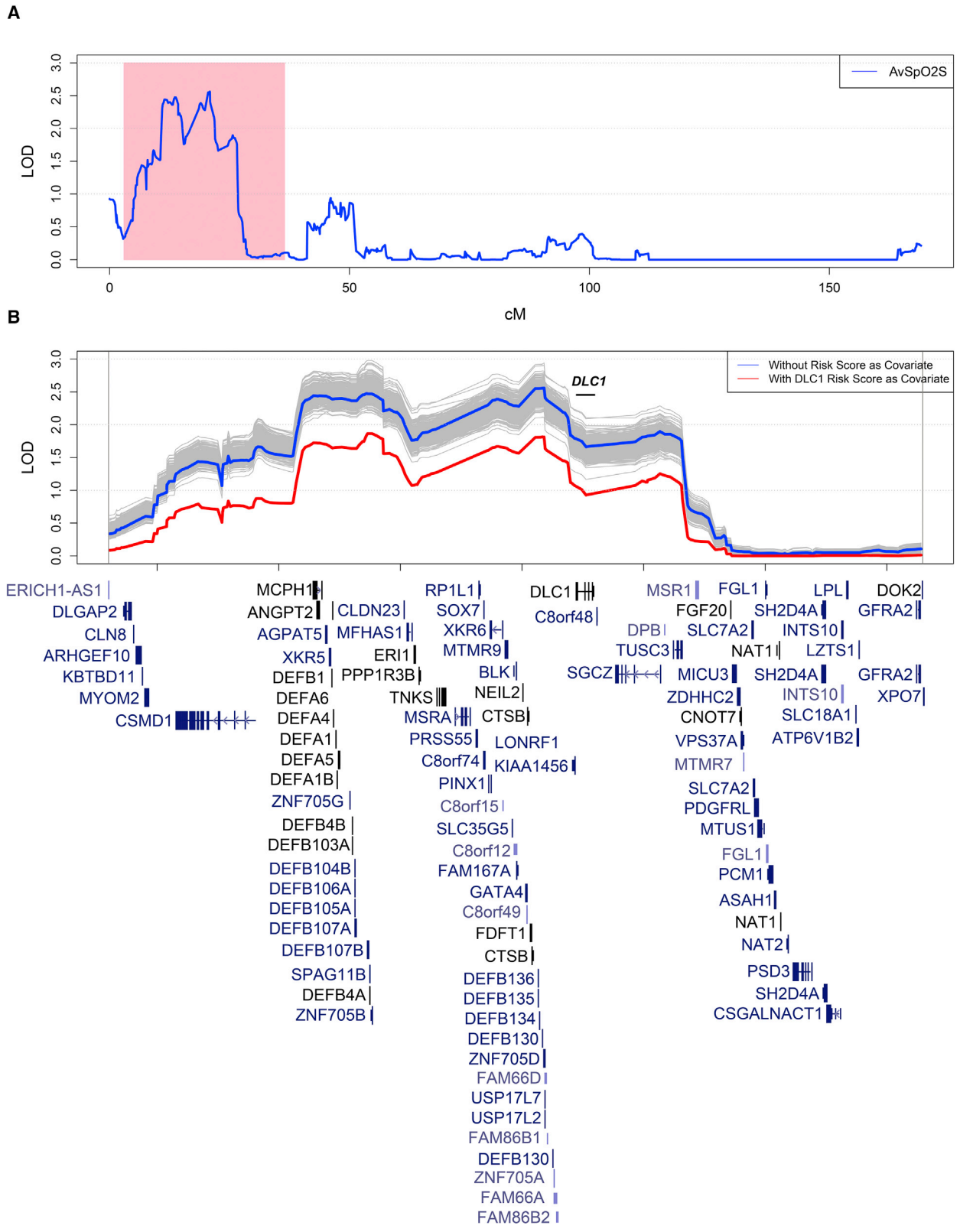

Figure 2. Linkage Evidence of AvSpO$_2$S on Chromosome 8 in Cleveland Family Study European Americans
(A) LOD score in 8p23 linked to AvSpO$_2$S. The pink region is the 20 Mb target region in the sequencing analysis and the protein coding genes are presented in the bottom.
(B) LOD score in 8p23 when the polygenic score (PS) of the 57 variants in *DLC1* was included in the linkage analysis. The linkage curves are plotted with (red curve) and without (blue curve) adjusting for the PS. The gray curves are the 1,000 linkage curves adjusted for PS defined by 57 randomly selected frequency-matched variants outside of the target region (chr8: 21,780,000–146,302,000bp for GRCh37/hg19) on chromosome 8. The location of *DLC1* is marked with a black bar.

Table 1. Stage I and II Gene-Based Association Tests with $AvSpO_2S$

	Stage I			Stage II[a]					
	Cleveland Family Study European Americans(n = 487)			European American(n = 2,772)		African American(n = 1,726)		Hispanic American(n = 2,795)	
Tests Using Functional Coding Variants									
Gene	**Variants Number**	**Burden p value**	**SKAT p value**	**Burden p value**	**SKAT p value**	**Burden p value**	**SKAT p value**	**Burden p value**	**SKAT P value**
ARHGEF10	4	0.031	0.023	0.360	0.091	0.600	0.620	0.660	0.200
FAM86B3P	2	0.140	0.005	0.480	0.700	0.610	0.780	0.350	0.280
DLC1	**6**	**0.027**	**0.051**	**0.036**	**0.072**	**0.035**	**0.440**	**0.200**	**0.450**
FGL1	4	0.220	0.038	0.670	0.990	0.200	0.530	0.590	0.680
ATP6V1B2	2	0.160	0.025	0.130	0.099	0.490	0.710	0.480	0.200
Tests Using Non-Coding Variants with CADD >10									
ERI1	4	0.003	0.031	0.440	0.360	0.170	0.380	0.520	0.850
ANGPT2	7	0.027	0.450	0.200	0.180	0.340	0.600	0.230	0.130
LONRF1	4	0.042	0.100	0.180	0.130	0.091	0.074	0.380	0.710
CSMD1	82	0.110	0.029	0.082	0.029	0.032	0.080	0.034	0.260
FDFT1	3	0.140	0.002	0.420	0.230	0.320	0.170	0.400	0.250
DLC1	**51**	**0.140**	**0.013**	**0.190**	**1.10×10^{-4}**	**0.250**	**0.280**	**0.061**	**0.380**
MTMR7	10	0.240	0.046	0.220	0.790	0.160	0.580	0.230	0.600
MYOM2	10	0.410	0.047	0.400	0.031	0.820	0.045	0.230	0.470

Only the genes with either burden or SKAT p value <0.05 in Stage I were reported. The results of *DLC1* are in bold.
[a]p values in Stage II were obtained by using weighted Fisher's method to combine p values from individual cohorts.

and SKAT analyses in each cohort separately, then combined association by the Fisher's method weighted by the sample sizes.[17] The test statistics was defined as

$$Q_{weighted-Fisher} = -2\sum_k N_k \log p_k$$

where N_k and p_k represent sample size and p value for the k^{th} cohort. The statistic $Q_{weight\text{-}Fisher}$ follows a mixture χ^2 distribution with mixture proportions by N_1, N_2, …., N_k. To calculate the p value, we applied the Satterthwaite's approximation.[18,19] Although the numbers of functional variants were small in each gene, we observed a nominal association with *DLC1* in EAs (p = 0.036), in AAs (p = 0.035), and in HAs (p = 0.2) in the burden test (Table 1). For the non-coding variants, we observed association in the SKAT analysis in EAs ($p = 1.1 \times 10^{-4}$) but not in AAs (p = 0.28) or in HAs (p = 0.38). This association evidence in EAs for non-coding variants was significant after accounting for multiple tests (12 genes × two analysis methods × two variant groups × three ethnic groups). The association evidence was further improved when combined with results for the Stage I CFS EA analysis ($p = 2.0 \times 10^{-5}$). We also observed nominal replication association evidence for *MYOM2* (MIM: 603509) in both EA and AA samples and for *CSMD1* (MIM: 608397) in EA samples in Stage II via SKAT analysis, although this evidence was not significant after correcting for multiple tests (Table 1).

We further obtained independent data from 4,449 EAs from four cohorts, the Osteopathic Fractures in Men Study (MrOS), the Framingham Heart Study (FHS), the Atherosclerosis Risk in Communities Study (ARIC), and the Western Australian Sleep Health Study (WASHS), including genotype data for Stage III replication and pooled analyses (see Supplemental Data, Table S5). Genotype imputation was performed using the Michigan Imputation Server.[20] The reference samples used were the subjects in TOPMed Freeze 5b who were whole genome sequenced.[21] Because the 57 variants in *DLC1* are either low-frequency or rare variants, we only kept the variants with imputation scores r^2 larger than 0.9. We were able to replicate the association evidence for coding variants only (p = 0.003), noncoding variants only (p = 0.0026), or combining coding and non-coding variants (p = 0.002), respectively. The association evidence of *DLC1* for analyses that combined data from stages II and III has a p value 2.9×10^{-6} and further increased to $p = 7.9 \times 10^{-7}$ (Table 2) when data from all stages (I, II, and III) were analyzed; this result is statistically significant after correcting for total 569 tests performed in this study, and it even reaches the genome-wide significance level $p = 2.5 \times 10^{-6}$ assuming there are 20,000 genes in the whole genome.

In the above analysis, we used the Fisher's method to combine p values from burden and SKAT tests. Fisher's method does not consider the effect directions of variants.

Table 2. Gene-Based Association Test for *DLC1* with AvSpO$_2$S in Stage I and II TOPMed Sequencing and Independent Replication Data with Stage III Imputed Genotypes

		Coding			Noncoding (CADD > 10)			ALL		
Study	**Sample Size**	**Variants Number**	**Burden p value**	**SKAT p value**	**Variants Number**	**Burden p value**	**SKAT p value**	**Variants Number**	**Burden p value**	**SKAT p value**
Stage I and II TOPMED WGS Data										
CFS	487	6	0.027	0.051	51	0.140	0.013	57	0.170	0.016
FHS	468	4	0.064	0.280	42	0.300	0.069	46	0.310	0.064
ARIC	1006	6	0.081	0.047	49	0.130	0.018	55	0.130	0.019
CHS	668	4	0.180	0.120	45	0.270	0.040	49	0.200	0.039
MESA	630	6	0.550	0.410	43	0.470	0.002	49	0.520	0.002
Meta-analysis[a]	3259	-	0.019	0.019	-	0.140	2.00×10^{-5}	-	0.130	2.20×10^{-5}
Stage III Imputed Genotype Data										
ARIC	583	4	0.620	0.390	36	0.500	0.210	40	0.510	0.220
FHS	181	3	0.380	0.670	39	0.400	0.280	42	0.400	0.210
MrOS	2178	4	0.005	0.017	36	0.180	0.006	40	0.180	0.005
WASHS	1507	5	0.110	0.540	39	0.140	0.110	44	0.260	0.130
Meta-analysis[a]	4449	-	0.003	0.035	-	0.140	0.003	-	0.210	0.002
Meta-analysis[a] **(all)**	7708	-	6.10×10^{-4}	0.006	-	0.097	9.20×10^{-7}	-	0.130	7.90×10^{-7}

CFS, Cleveland Family Study; FHS, Framingham Heart Study; ARIC, Atherosclerosis Risk in Communities Study; CHS, Cardiovascular Health Study; MESA, Multi-Ethnic Study of Atherosclerosis
[a]p values in meta-analysis were obtained by using weighted Fisher's method to combine p values from individual cohorts.

Therefore, we applied MetaSKAT, which can incorporate different levels of genetic heterogeneity across studies and can apply to population-bcased samples,[22] to obtain the combined DLC1 association evidence with AvSpO$_2$S. MetaSKAT was applied to the 57 variants in tage II and III data separately because Stage III data utilized genetic data imputed from chip assays while Stage II data utilized directly sequenced data. The p values for Stages II and III were 1.0×10^{-4} and 2.4×10^{-3}, respectively, which were consistent with the p values obtained via the Fisher's method (Table S12).

If the variants identified in *DLC1* are truly associated with AvSpO2S, conditioning on the effects of these variants should reduce the observed linkage evidence in the CFS. We performed single-variant analysis in each of the Stage II cohorts, followed by meta-analysis,[23] and obtained the effect sizes of the 57 *DLC1* variants (Table S7). Next, for each individual in the Stage I CFS EAs, we calculated a *DLC1* gene score defined by

$$PS = \sum_{i=1}^{57} \widehat{\beta}_i g_i$$

where g_i is the ith genotype value and $\widehat{\beta}_i$ is the corresponding effect size obtained from the Stage II data. We performed linkage analysis by including the *PS* score as a covariate, and we calculated the drop in the LOD score. To examine whether the LOD score drop is statistically significant, we randomly selected the same number of allele-frequency-matched variants outside of the linkage region. We estimated their effect sizes using Stage II cohorts, and we calculated the *PS* score again. We performed linkage analysis in CFS EAs 1,000 times and calculated the empirical distribution for the null hypothesis that the variants are not associated with the average AvSpO$_2$S. We then calculated the p value for the observed LOD score drop in each gene. The maximum LOD score dropped to 1.81 with the *DLC1* gene score included. The drop was statistically significant ($p < 0.001$, Figure 2B) suggesting that these variants contribute to the observed linkage evidence. Similar analyses showed that the variants in *CSMD1* led to a significant LOD score drop ($p = 0.004$) but variants in other genes did not (Figure S6).

Conditional on allele frequencies, eight of the 57 variants in *DLC1* have statistically significant effect sizes that fall in the top 5% of allele-frequency-matched variants (Figure S7, $p = 0.020$). These observed effect sizes of low-frequency variants are not necessarily large, and this is consistent with prior literature (e.g., in regards to human height[24]). We further observed that four of six coding variants had consistent positive effect direction (higher oxygen saturation; a more favorable phenotype), but this pattern was not observed for non-coding variants (Figure S7). The 57 variants in *DLC1* together explained 0.97% of AvSpO$_2$S variation ($p = 0.0017$) in EAs. Based on this attributable variation, we calculated the power of the AA and HA sample sizes to be 21% and 28% at a 5% significance level, respectively. We also noted that the power

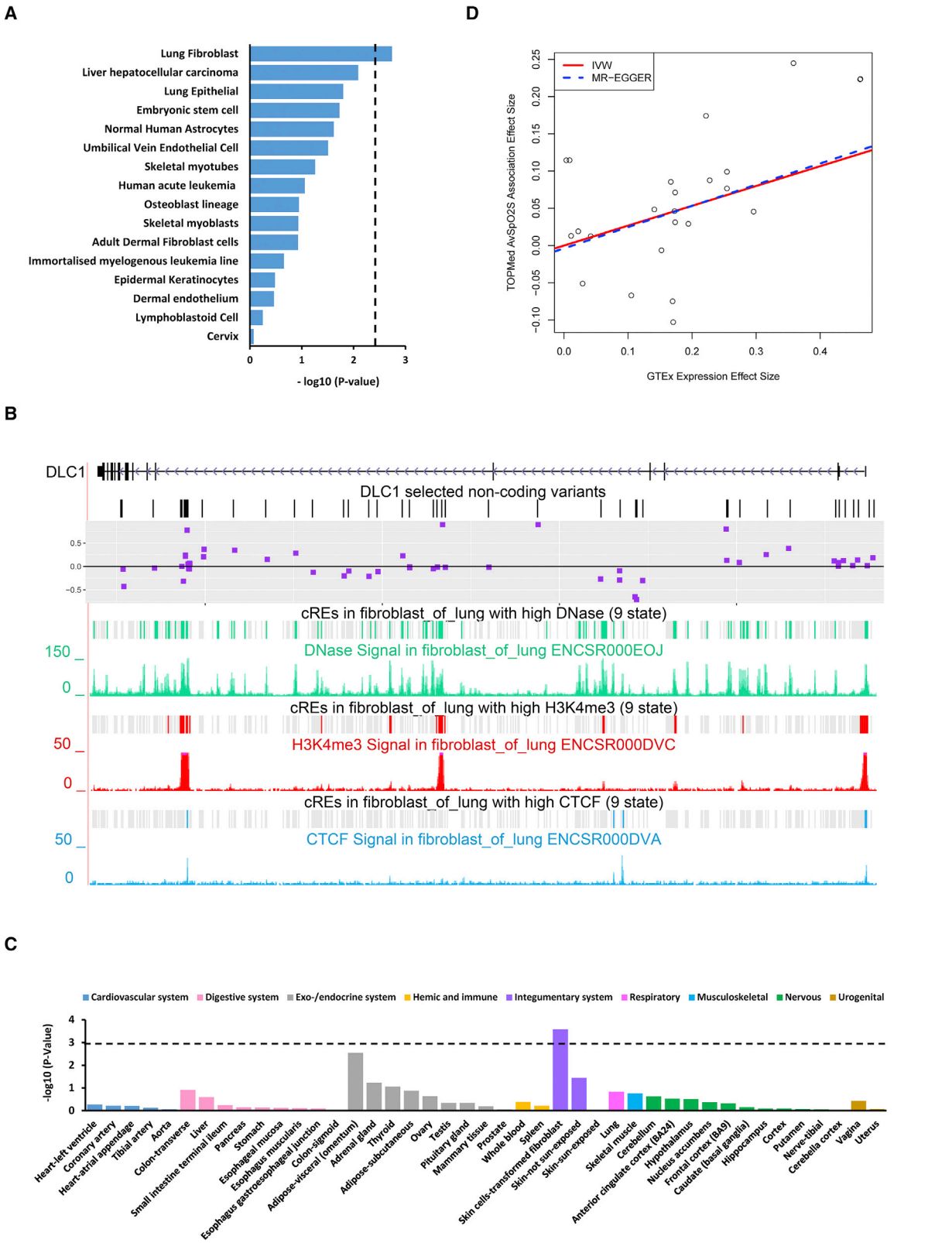

Figure 3. The 57 Variants in *DLC1*

(A) Cell-type-specific regulatory annotation enrichment tests for the 51 non-coding variants in *DLC1* in 16 cell lines defined in the Ensemble Regulatory Build. The vertical dotted line represents the significance level after adjusting for multiple tests.

(B) 51 non-coding variants and the corresponding effect sizes in *DLC1* genes plotted against physical locations. The corresponding DNase hypersensitive, H3K4me3, H3K27ac, and CTCF elements derived from lung fibroblasts in the Encyclopedia of DNA Elements (ENCODE) data were also presented.

(legend continued on next page)

should be further reduced because of different allele frequency in populations other than EAs. Therefore, our current sample sizes have low statistical power in AAs and HAs, likely explaining our failure to observe association evidence in AAs and HAs (Table 1).

We next examined whether the identified non-coding variants in *DLC1* are enriched in regulatory regions by comparing these with the remaining frequency- and CADD-score-matched variants (see Supplemental Data). We examined the regulatory-activity-predicted elements for the 16 cell lines defined in the Ensembl Regulatory Build,[25] which includes CTCF binding sites, enhancer, heterochromatin, promoter flank, and transcription start sites (TSS). The 51 non-coding variants in *DLC1* were significantly enriched with cis-regulatory elements (CREs) in the Human Lung Fibroblast cell line (NHLF) after correcting for multiple tests ($p = 0.003$, Figure 3A). Figure 3B demonstrates the genomic locations of the 51 variants with their corresponding CREs in the human lung fibroblast cell line. We observed significant aggregation of variants with similar effect direction in the genomic region ($p = 4.7 \times 10^{-4}$, see Supplemental Data). The noncoding variants in *MYOM2* were also marginally enriched in skeletal myotubes and skeletal myoblasts (Figure S8).

We further investigated whether the low-frequency non-coding variants have gene regulatory roles by conducting eQTL analysis of their corresponding RNaseq data across the 44 tissues from the Genotype-Tissue Expression (GTEx) program.[26] After correcting for 44 tests ($p = 2.6 \times 10^{-4}$, Figure 3C), we found that the 51 non-coding variants in *DLC1* significantly contributed to *DLC1* expression level in human-skin-cell-transformed fibroblasts; this result is consistent with our observation that these variants are enriched in regulatory features in the human lung fibroblast cell line (Figure 3A). We observed a significant correlation between the $AvSpO_2S$ effect sizes of the 24 available variants in GTEx and *DLC1* expression effect sizes ($p = 5.6 \times 10^{-5}$). The additive score we found by using *DLC1* expression effect sizes of the 24 *DLC1* variants explained 0.41% of $AvSpO_2S$ variation ($p = 6.8 \times 10^{-5}$).

We next performed Mendelian randomization analysis to test and estimate the causal effect of *DLC1* expression in human-skin-cell-transformed fibroblasts by using an inverse-variance weighted (IVW) estimate, MR-Egger regression,[27] and MR-presso.[28] We constructed instrumental variables using the 24 available *DLC1* variants in GTEx. The *DLC1* expression level was treated as an exposure and $AvSpO_2S$ was treated as an outcome. Mendelian randomization analysis using the *DLC1* variants demonstrated a significant causal effect of *DLC1* expression on $AvSpO_2S$ ($p = 4.56 \times 10^{-4}$, Figure 3D, Table S8), suggesting that *DLC1* variants contribute to $AvSpO_2S$ variation through *DLC1* expression. The non-coding variants in *MYOM2* were significantly associated with *MYOM2* expression level in the brain cortex (burden test $p = 2.1 \times 10^{-4}$) but not in *CSMD1* (Figure S9A and S9B).

We investigated the association between DNA methylation in peripheral monocytes in *DLC1* and in $AvSpO_2S$. We tested 77 DNA methylation sites in *DLC1* and observed one associated site, cg08148801 at chr8:12992570 ($p = 0.001$, FDR adjusted $p = 0.078$), using the 623 subjects from the Multi-Ethnic Study of Atherosclerosis (MESA)(Table S9). We observed weak associations between *DLC1* gene expression from peripheral white blood cells and $AvSpO_2S$ ($p = 0.15$) and apnea hypopnea index (AHI) ($p = 0.06$), a measure of SDB severity that correlates with sleep-associated hypoxemia ($r = 0.63$ in CFS EAs), in 517 subjects from Framingham Heart Study (FHS) (Table S10).

We further adjusted for AHI, the most common metric for SDB in clinical assessments, and found that the association of both coding and non-coding variants in *DLC1* remained significant, with their effects almost unchanged (Table S11, Figure S10); this result suggests that the association was not mediated by frequency of apneas. Two common variants in *DLC1*, SNPs rs74834049 and rs7520069, have been found via computed tomography to be associated with two emphysema-related phenotypes,[29] a pulmonary disease trait that is associated with dyspnea,[30] reduced activity levels,[31] and exercise tolerance.[32] We found that SNPs rs74834049 and rs7520069 were marginally associated with $AvSpO_2S$, with p values of 0.053 and 0.050, respectively, in Stage I and II EA samples; this finding suggests that the common variants in *DLC1* may also contribute to $AvSpO_2S$. We further examined the association of *DLC1* while adjusting for lung function (predicted forced expiratory volume in one second [FEV_1] and forced vital capacity [FVC]). The association came to a slightly reduced significance level but was still remained highly correlated (Pearson correlation 0.9) with and without adjusting for FEV_1 or FVC (Figure S10). This likely reflects the sample size reduction due to missing lung function data (Table 3). We also observed association evidence with FEV_1 and FVC in the burden tests (p values = 0.014 and 0.037 respectively, Table 3). In aggregate, these associations suggest a potential pleiotropic effect between *DLC1*, $AvSpO_2S$, lung function, and SDB traits at a gene level (Table 3).

The implication of *DLC1* in sleep-related oxygenation is of interest given that *DLC1* is highly expressed in lung tissue, where it functions as an inhibitor of small GTPases,

(C). Association of the 57 variants in *DLC1* with *DLC1* expression level in 44 tissues from GTEx. The horizontal dotted line represents the significance level after adjustment for multiple tests.
(D) Mendelian randomization analysis using the 24 *DLC1* variants as instrument variables. *DLC1* expression level in skin-cell-transformed-fibroblasts in GTEx is treated as exposure and $AvSpO_2S$ is treated as outcome. The solid red and blue dotted lines represent the causal effects estimated by the inverse-variance-weighted method and MR-Egger regression (see Supplemental Data).

Table 3. Stage I and II Gene-Based Association Tests for *DLC1* with $AvSpO_2S$ or FEV_1/FVC by Adjusting FEV_1/FVC as Covariates in Subsample of Stage I and II Cohorts

		Stage I			Stage II[a]							
		CFS EAs(N = *402*)			Stage II EA cohorts (Total N = 1,200)		Stage II AA cohorts (Total N = *653*)		Stage II EA and AA cohorts (Total N = 1,853)		Stage I and II EAs(Total N = 1,602)	
Phenotype	Covariate	Variants Number	Burden P value	SKAT P value	Burden P value	SKAT P value	Burden P value	SKAT P value	Burden P value	SKAT P value	Burden P value	SKAT P value
Functional Coding Variants												
$AvSpO_2S$	-	6	0.054	0.080	0.303	0.460	0.045	0.671	0.076	0.625	0.110	0.217
$AvSpO_2S$	FEV_1	6	0.018	0.024	0.318	0.443	0.038	0.675	0.070	0.613	0.014	0.210
$AvSpO_2S$	FVC	6	0.018	0.021	0.217	0.385	0.045	0.702	0.054	0.577	0.038	0.079
FEV_1	-	6	0.387	0.475	0.590	0.828	0.262	0.052	0.358	0.272	0.590	0.828
FVC	-	6	0.742	0.937	0.602	0.926	0.327	0.299	0.358	0.272	0.602	0.926
Non-Coding Variants with CADD PHRED > 10												
$AvSpO_2S$	-	51	0.470	0.025	0.469	0.006	0.135	0.254	0.246	0.009	0.525	0.001
$AvSpO_2S$	FEV_1	51	0.214	0.023	0.099	0.038	0.159	0.222	0.069	0.040	0.061	0.009
$AvSpO_2S$	FVC	51	0.130	0.017	0.145	0.019	0.156	0.233	0.094	0.023	0.092	0.003
FEV_1	-	51	0.090	0.768	0.027	0.401	0.926	0.195	0.126	0.268	0.014	0.566
FVC	-	51	0.080	0.924	0.077	0.510	0.579	0.487	0.162	0.560	0.037	0.712

CFS, Cleveland Family Study; EA, Eastern European; AA, African American
[a]p values in Stage II were obtained by using the Fisher's method to combine p values from individual cohorts.

influencing cell proliferation, migration, apoptosis, and angiogenesis.[22,23] *DLC1* also functions as an activator of *PLCD1*, a repressor of airway smooth muscle hypertrophy.[33] Thus, by modulating endothelial cell function and smooth muscle contractility, it may influence cardiovascular and pulmonary traits.[34,35] A common variant near *DLC1*, but distinct from variants associated with $AvSpO_2S$, was associated with quantitative lung imaging markers of emphysema.[29] *DLC1* may specifically influence oxygen saturation levels in SDB by modulating the effects of SDB-related mechanical (i.e., via episodic airway obstruction) or oxidative stressors on subclinical lung parenchymal disease.[36] Notably, *DLC1* function is modulated by reactive oxygen radicals,[37] which are commonly elevated in SDB.[38] *DLC1* is also a PPARG target critical for adipocyte differentiation,[34] and thus it may influence SDB-related hypoxemia through effects on body fat distribution and its influences on ventilation.

From a clinical perspective, it is recognized that oxygen saturation levels are reduced in the presence of lung disease, and such reductions in oxygen saturation would be expected to be more pronounced during sleep, when respiratory drive declines and gravitational effects of the supine position may reduce lung volume. In the presence of SDB, one would expect to see the largest decreases in average oxygen saturation due to recurrent apneas and hypopneas. To further explore the inter-connections among oxygen saturation, lung function, and lung disease, we also conducted regression analysis of $AvSpO_2S$ with lung function (FEV_1, FVC), adjusting for covariates in the CFS. As expected, both FEV_1 and FVC correlated with $AvSpO_2S$. However, the correlation disappeared after we adjusted for gender, age, age × age, AHI, and smoking (Table S13). Chronic Obstructive Pulmonary Disease (COPD) was also, as expected, negatively correlated ($r = -0.176$, $p = 1.77 \times 10^{-9}$) with $AvSpO_2S$, but asthma was not ($r = -0.035$, $p = 1.0$). It is important to note that individuals in this sample were not selected on the basis of lung disease, and most had lung function in the normal range. Therefore, additional research that includes a broader spectrum of lung disease, in addition to measuring traits associated with SDB, will be useful for further understanding how variants in *DLC1* affect both lung function and oxygenation during sleep.

We observed that the two SNPs previously reported to be associated with emphysema-related traits, rs74834049 and rs7520069, were marginally associated with $AvSpO_2S$ (p values 0.053 and 0.050 respectively) in combined Stage I and II EA samples. A search of the genome-wide association study (GWAS) database identified a number of additional associations of variants in *DLC1* with several traits, including associations with oxygen carrying capacity and inflammation (mean corpuscular hemoglobin, white and red blood cell count[39]), lung inflammation (childhood fractional exhaled nitric oxide[40]), and cardiovascular risk factors (high density lipoprotein cholesterol,[41] venous thromboembolism[42]), as well as traits that are correlated with increased mortality (male pattern baldness,[43] height[39]), heel bone mineral density,[44] and intraocular pressure.[45] However, these variants from GWAS do not

overlap with the rare variants reported in our study. Future work is warranted in order to understand the potentially pleiotropic effects of *DLC1* and their influence on lung function and SDB, as well as on other conditions, such as cardiovascular disease and premature mortality.

In summary, our analyses identified an association between oxyhemoglobin saturation levels during sleep, a clinically important but understudied phenotype, and *DLC1*, a gene having pleiotropic functions most studied in relationship to tumor activity but also relevant to lung function and, and as shown here, oxygenation. Although our total sample size was small compared to the sample sizes of most large low-frequency and rare variant association studies,[24] we show consistent association evidence of low-frequency and rare variants in *DLC1* and $AvSpO_2S$ in multiple omics data, strongly suggesting that the association is real and that there is improved statistical power in using family cohorts in rare variant studies.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.10.002.

Acknowledgments

This work was supported by grants HL113338, HL046389, and HL135818 (to S.R.) from the National Heart, Lung, and Blood Institute (NHLBI) and HG003054 (to X.Z.) from the National Human Genome Research Institute (NHGRI). J.L. was supported by T32HL007567 from the NHLBI. Whole-genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the NHLBI. Detailed funding information can be found in the Supplemental Data. We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. The contributions of the investigators of the NHLBI TOPMed Consortium are gratefully acknowledged. A full TOPMed authorship list is available in the Supplemental Data.

Declaration of Interests

The authors declare no competing interests.

Received: May 5, 2019
Accepted: October 2, 2019
Published: October 24, 2019

Web Resources

GTEx project, https://gtexportal.org/home/
GWAS database, https://www.ebi.ac.uk/gwas/
Michigan Imputation Server, http://imputationserver.sph.umich.edu/
Online Mendelian Inheritance in Man, https://www.omim.org
TOPMed consortium, https://www.nhlbiwgs.org/

References

1. Wang, H., Cade, B.E., Chen, H., Gleason, K.J., Saxena, R., Feng, T., Larkin, E.K., Vasan, R.S., Lin, H., Patel, S.R., et al. (2016). Variants in angiopoietin-2 (ANGPT2) contribute to variation in nocturnal oxyhaemoglobin saturation level. Hum. Mol. Genet. *25*, 5244–5253.
2. Liang, J., Cade, B.E., Wang, H., Chen, H., Gleason, K.J., Larkin, E.K., Saxena, R., Lin, X., Redline, S., and Zhu, X. (2016). Comparison of heritability estimation and linkage analysis for multiple traits using principal component analyses. Genet. Epidemiol. *40*, 222–232.
3. Antonelli Incalzi, R., Marra, C., Giordano, A., Calcagni, M.L., Cappa, A., Basso, S., Pagliari, G., and Fuso, L. (2003). Cognitive impairment in chronic obstructive pulmonary disease–a neuropsychological and spect study. J. Neurol. *250*, 325–332.
4. Nieto, F.J., Peppard, P.E., Young, T., Finn, L., Hla, K.M., and Farré, R. (2012). Sleep-disordered breathing and cancer mortality: Results from the Wisconsin Sleep Cohort Study. Am. J. Respir. Crit. Care Med. *186*, 190–194.
5. Yaffe, K., Laffan, A.M., Harrison, S.L., Redline, S., Spira, A.P., Ensrud, K.E., Ancoli-Israel, S., and Stone, K.L. (2011). Sleep-disordered breathing, hypoxia, and risk of mild cognitive impairment and dementia in older women. JAMA *306*, 613–619.
6. Kolilekas, L., Manali, E., Vlami, K.A., Lyberopoulos, P., Triantafillidou, C., Kagouridis, K., Baou, K., Gyftopoulos, S., Vougas, K.N., Karakatsani, A., et al. (2013). Sleep oxygen desaturation predicts survival in idiopathic pulmonary fibrosis. J. Clin. Sleep Med. *9*, 593–601.
7. Corte, T.J., Wort, S.J., Talbot, S., Macdonald, P.M., Hansel, D.M., Polkey, M., Renzoni, E., Maher, T.M., Nicholson, A.G., and Wells, A.U. (2012). Elevated nocturnal desaturation index predicts mortality in interstitial lung disease. Sarcoidosis Vasc. Diffuse Lung Dis. *29*, 41–50, Sarcoidosis Vasc Diffuse Lung Dis.
8. Ross, K.R., Storfer-Isser, A., Hart, M.A., Kibler, A.M., Rueschman, M., Rosen, C.L., Kercsmar, C.M., and Redline, S. (2012). Sleep-disordered breathing is associated with asthma severity in children. J. Pediatr. *160*, 736–742.
9. Lacasse, Y., Sériès, F., Vujovic-Zotovic, N., Goldstein, R., Bourbeau, J., Lecours, R., Aaron, S.D., and Maltais, F. (2011). Evaluating nocturnal oxygen desaturation in COPD—revised. Respir. Med. *105*, 1331–1337.
10. Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., et al.; The Breast Cancer Linkage Consortium (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. Am. J. Hum. Genet. *62*, 676–689.
11. He, K.Y., Wang, H., Cade, B.E., Nandakumar, P., Giri, A., Ware, E.B., Haessler, J., Liang, J., Smith, J.A., Franceschini, N., et al. (2017). Rare variants in fox-1 homolog A (RBFOX1) are associated with lower blood pressure. PLoS Genet. *13*, e1006678.
12. Wang, H., Nandakumar, P., Tekola-Ayele, F., Tayo, B.O., Ware, E.B., Gu, C.C., Lu, Y., Yao, J., Zhao, W., Smith, J.A., et al. (2019). Combined linkage and association analysis identifies rare and low frequency variants for blood pressure at 1q31. Eur. J. Hum. Genet. *27*, 269–277.
13. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
14. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. *89*, 82–93.

15. Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., Christiani, D.C., Wurfel, M.M., Lin, X.; and NHLBI GO Exome Sequencing Project—ESP Lung Project Team (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am. J. Hum. Genet. *91*, 224–237.
16. Chung, R.H., Hauser, E.R., and Martin, E.R. (2007). Interpretation of simultaneous linkage and family-based association tests in genome screens. Genet. Epidemiol. *31*, 134–142.
17. Liu, D.J., Peloso, G.M., Zhan, X., Holmen, O.L., Zawistowski, M., Feng, S., Nikpay, M., Auer, P.L., Goel, A., Zhang, H., et al. (2014). Meta-analysis of gene-level tests for rare variant association. Nat. Genet. *46*, 200–204.
18. Davtes, R.B. (1973). Numerical inversion of a characteristic function. Biometrika *60*, 415–417.
19. Kuonen, D. (1999). Saddlepoint approximations for distributions of quadratic forms in normal variables. Biometrika *86*, 929–935.
20. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.
21. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Gagliano Taliun, S.A., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2019). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. bioRxiv.
22. Lee, S., Teslovich, T.M., Boehnke, M., and Lin, X. (2013). General framework for meta-analysis of rare variants in sequencing association studies. Am. J. Hum. Genet. *93*, 42–53.
23. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
24. Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., et al.; EPIC-InterAct Consortium; CHD Exome+ Consortium; ExomeBP Consortium; T2D-Genes Consortium; GoT2D Genes Consortium; Global Lipids Genetics Consortium; ReproGen Consortium; and MAGIC Investigators (2017). Rare and low-frequency coding variants alter human adult height. Nature *542*, 186–190.
25. Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T., and Flicek, P.R. (2015). The ensembl regulatory build. Genome Biol. *16*, 56.
26. Consortium, G.T.; and GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.
27. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. Int. J. Epidemiol. *44*, 512–525.
28. Verbanck, M., Chen, C.Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat. Genet. *50*, 693–698.
29. Cho, M.H., Castaldi, P.J., Hersh, C.P., Hobbs, B.D., Barr, R.G., Tal-Singer, R., Bakke, P., Gulsvik, A., San José Estépar, R., Van Beek, E.J., et al.; NETT Genetics, ECLIPSE, and COPDGene Investigators (2015). A genome-wide association study of emphysema and airway quantitative imaging phenotypes. Am. J. Respir. Crit. Care Med. *192*, 559–569.
30. Oelsner, E.C., Lima, J.A., Kawut, S.M., Burkart, K.M., Enright, P.L., Ahmed, F.S., and Barr, R.G. (2015). Noninvasive tests for the diagnostic evaluation of dyspnea among outpatients: The Multi-Ethnic Study of Atherosclerosis lung study. Am. J. Med. *128*, 171–180.e5.
31. Lo Cascio, C.M., Quante, M., Hoffman, E.A., Bertoni, A.G., Aaron, C.P., Schwartz, J.E., Avdalovic, M.V., Fan, V.S., Lovasi, G.S., Kawut, S.M., et al. (2017). Percent emphysema and daily motor activity levels in the general population: Multi-Ethnic Study of Atherosclerosis. Chest *151*, 1039–1050.
32. Smith, B.M., Prince, M.R., Hoffman, E.A., Bluemke, D.A., Liu, C.Y., Rabinowitz, D., Hueper, K., Parikh, M.A., Gomes, A.S., Michos, E.D., et al. (2013). Impaired left ventricular filling in COPD and emphysema: Is it the heart or the lungs? The Multi-Ethnic Study of Atherosclerosis COPD Study. Chest *144*, 1143–1151.
33. Sasse, S.K., Kadiyala, V., Danhorn, T., Panettieri, R.A., Jr., Phang, T.L., and Gerber, A.N. (2017). Glucocorticoid receptor ChIP-seq identifies PLCD1 as a KLF15 target that represses airway smooth muscle hypertrophy. Am. J. Respir. Cell Mol. Biol. *57*, 226–237.
34. Sim, C.K., Kim, S.Y., Brunmeir, R., Zhang, Q., Li, H., Dharmasegaran, D., Leong, C., Lim, Y.Y., Han, W., and Xu, F. (2017). Regulation of white and brown adipocyte differentiation by RhoGAP DLC1. PLoS ONE *12*, e0174761.
35. Lu, Q., Longo, F.M., Zhou, H., Massa, S.M., and Chen, Y.H. (2009). Signaling through Rho GTPase pathway as viable drug target. Curr. Med. Chem. *16*, 1355–1365.
36. Kim, J.S., Podolanczuk, A.J., Borker, P., Kawut, S.M., Raghu, G., Kaufman, J.D., Stukovsky, K.D.H., Hoffman, E.A., Barr, R.G., Gottlieb, D.J., et al. (2017). Obstructive sleep apnea and subclinical interstitial lung disease in the Multi-Ethnic Study of Atherosclerosis (MESA). Ann. Am. Thorac. Soc. *14*, 1786–1795.
37. Yang, B., Zhu, W., Zheng, Z., Chai, R., Ji, S., Ren, G., Liu, T., Liu, Z., Song, T., Li, F., et al. (2017). Fluctuation of ROS regulates proliferation and mediates inhibition of migration by reducing the interaction between DLC1 and CAV-1 in breast cancer cells. In Vitro Cell. Dev. Biol. Anim. *53*, 354–362.
38. Lira, A.B., and de Sousa Rodrigues, C.F. (2016). Evaluation of oxidative stress markers in obstructive sleep apnea syndrome and additional antioxidant therapy: a review article. Sleep Breath. *20*, 1155–1160.
39. Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging polygenic functional enrichment to improve GWAS power. Am. J. Hum. Genet. *104*, 65–75.
40. van der Valk, R.J., Duijts, L., Timpson, N.J., Salam, M.T., Standl, M., Curtin, J.A., Genuneit, J., Kerhof, M., Kreiner-Møller, E., Cáceres, A., et al.; EArly Genetics & Lifecourse Epidemiology (EAGLE) Consortium (2014). Fraction of exhaled nitric oxide values in childhood are associated with 17q11.2-q12 and 17q12-q21 variants. J. Allergy Clin. Immunol. *134*, 46–55.
41. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514–518.
42. Greliche, N., Germain, M., Lambert, J.C., Cohen, W., Bertrand, M., Dupuis, A.M., Letenneur, L., Lathrop, M., Amouyel, P., Morange, P.E., and Trégouët, D.A. (2013). A genome-wide search for common SNP x SNP interactions on the risk of venous thrombosis. BMC Med. Genet. *14*, 36.

43. Yap, C.X., Sidorenko, J., Wu, Y., Kemper, K.E., Yang, J., Wray, N.R., Robinson, M.R., and Visscher, P.M. (2018). Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. Nat. Commun. *9*, 5407.
44. Morris, J.A., Kemp, J.P., Youlten, S.E., Laurent, L., Logan, J.G., Chai, R.C., Vulpescu, N.A., Forgetta, V., Kleinman, A., Mohanty, S.T., et al.; 23andMe Research Team (2019). An atlas of genetic influences on osteoporosis in humans and mice. Nat. Genet. *51*, 258–266.
45. MacGregor, S., Ong, J.S., An, J., Han, X., Zhou, T., Siggs, O.M., Law, M.H., Souzeau, E., Sharma, S., Lynn, D.J., et al. (2018). Genome-wide association study of intraocular pressure uncovers new pathways to glaucoma. Nat. Genet. *50*, 1067–1071.