

RESEARCH

Open Access



Rare variant analysis of 4241 pulmonary arterial hypertension cases from an international consortium implicates *FBLN2*, *PDGFD*, and rare de novo variants in PAH

Na Zhu^{1,2†}, Emilia M. Swietlik^{3†}, Carrie L. Welch^{1†}, Michael W. Pauciulo^{4,5†}, Jacob J. Hagen^{1,2}, Xueya Zhou^{1,2}, Yicheng Guo², Johannes Karten⁶, Divya Pandya³, Tobias Tilly³, Katie A. Lutz⁴, Jennifer M. Martin^{3,7}, Carmen M. Treacy³, Erika B. Rosenzweig¹, Usha Krishnan¹, Anna W. Coleman⁴, Claudia Gonzaga-Juaregui⁸, Allan Lawrie⁹, Richard C. Trembath¹⁰, Martin R. Wilkins¹¹, Regeneron Genetics Center⁸, PAH Biobank Enrolling Centers' Investigators¹², NIHR BioResource for Translational Research - Rare Diseases¹³, National Cohort Study of Idiopathic and Heritable PAH¹⁴, Nicholas W. Morrell^{3,7,15,16†}, Yufeng Shen^{2,17†}, Stefan Gräf^{3,18,7†}, William C. Nichols^{4,5†} and Wendy K. Chung^{1,19,20*†}

Abstract

Background: Pulmonary arterial hypertension (PAH) is a lethal vasculopathy characterized by pathogenic remodeling of pulmonary arterioles leading to increased pulmonary pressures, right ventricular hypertrophy, and heart failure. PAH can be associated with other diseases (APAH: connective tissue diseases, congenital heart disease, and others) but often the etiology is idiopathic (IPAH). Mutations in bone morphogenetic protein receptor 2 (*BMPR2*) are the cause of most heritable cases but the vast majority of other cases are genetically undefined.

(Continued on next page)

* Correspondence: wkc15@columbia.edu

[†]Na Zhu, Emilia M. Swietlik, Carrie L. Welch, Michael W. Pauciulo contributed equally to this work.

[†]Nicholas W. Morrell, Yufeng Shen, Stefan Gräf, William C. Nichols and Wendy K. Chung jointly supervised the work.

¹Department of Pediatrics, Columbia University Irving Medical Center, 1150 St. Nicholas Avenue, Room 620, New York, NY 10032, USA

¹⁹Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Methods: To identify new risk genes, we utilized an international consortium of 4241 PAH cases with exome or genome sequencing data from the National Biological Sample and Data Repository for PAH, Columbia University Irving Medical Center, and the UK NIHR BioResource – Rare Diseases Study. The strength of this combined cohort is a doubling of the number of IPAH cases compared to either national cohort alone. We identified protein-coding variants and performed rare variant association analyses in unrelated participants of European ancestry, including 1647 IPAH cases and 18,819 controls. We also analyzed de novo variants in 124 pediatric trios enriched for IPAH and APAH-CHD.

Results: Seven genes with rare deleterious variants were associated with IPAH with false discovery rate smaller than 0.1: three known genes (*BMP2*, *GDF2*, and *TBX4*), two recently identified candidate genes (*SOX17*, *KDR*), and two new candidate genes (fibulin 2, *FBLN2*; platelet-derived growth factor D, *PDGFD*). The new genes were identified based solely on rare deleterious missense variants, a variant type that could not be adequately assessed in either cohort alone. The candidate genes exhibit expression patterns in lung and heart similar to that of known PAH risk genes, and most variants occur in conserved protein domains. For pediatric PAH, predicted deleterious de novo variants exhibited a significant burden compared to the background mutation rate (2.45×, $p = 2.5e-5$). At least eight novel pediatric candidate genes carrying de novo variants have plausible roles in lung/heart development.

Conclusions: Rare variant analysis of a large international consortium identified two new candidate genes—*FBLN2* and *PDGFD*. The new genes have known functions in vasculogenesis and remodeling. Trio analysis predicted that ~15% of pediatric IPAH may be explained by de novo variants.

Keywords: Genetics, Pulmonary arterial hypertension, Exome sequencing, Genome sequencing, Case-control association testing, De novo variant analysis

Background

Pulmonary arterial hypertension (PAH) remains a progressive, lethal vasculopathy despite recent therapeutic advances. The disease is characterized by pulmonary vascular endothelial dysfunction and proliferative remodeling giving rise to increased pulmonary artery pressures and pulmonary vascular resistance. These pathological changes of the lung vasculature strain the right ventricle of the heart, leading to right ventricular hypertrophy, right heart failure, and high mortality [1–3]. Dysregulated vascular, inflammatory, and immune cells contribute to these pathological processes [3]. PAH can present at any age, but the ~3:1 female to male ratio in adult-onset disease is not observed in pediatric-onset disease, in which the disease incidence is similar for males and females. The estimated prevalence of PAH is 4.8–8.1 cases/million for pediatric-onset [4] and 5.6–25 cases/million for adult-onset disease [5]. Early genetic linkage and candidate gene studies indicated an autosomal dominant mode of inheritance for PAH risk. However, the known susceptibility variants are incompletely penetrant, many individuals who carry monogenic risk variants never develop PAH, and a subset of patients have deleterious variants in more than one risk gene. For example, bone morphogenetic protein receptor type 2 (*BMP2*) mutations are observed in 60–80% of familial (FPAH) cases, but data from population registries indicate that penetrance of the disease phenotype ranges from 14 to 42% [6]. These data suggest that

additional genetic, epigenetic, environmental factors, and gene × environment interactions contribute to disease.

Genetic analyses of larger cohorts using gene panels, exome sequencing (ES), or genome sequencing (GS) have further defined the frequency of individuals with deleterious variants in PAH risk genes and have identified novel candidate risk genes. *BMP2* mutations are observed in the majority of FPAH cases across genetic ancestries [7–11]. *BMP2* carriers have younger mean age-of-onset and are less responsive to vasodilators compared to non-carriers [7, 12, 13], with an enrichment of predicted deleterious missense (D-Mis) variants with younger age-of-onset [7, 14]. However, *BMP2* variants have been identified in only 10–20% of previously classified idiopathic PAH (IPAH) and rarely to PAH associated with other diseases (APAH: autoimmune connective tissue diseases, congenital heart disease (CHD), portopulmonary disease and others) or PAH induced by diet and toxins. Variants in two other genes in the transforming growth factor-beta (TGF- β) superfamily, activin A receptor type II-like 1 (*ACVRL1*), and endoglin (*ENG*) contribute to ~0.8% of PAH cases [7], especially PAH associated with hereditary hemorrhagic telangiectasia (APAH-HHT). Variants in growth differentiation factor 2 (*GDF2*), encoding the ligand of *BMP2/ACVRL1* (BMP9), contribute to ~1% of PAH (mostly IPAH) cases in European-enriched cohorts [7, 8] and more frequently in Chinese patients (~6.7%) [15]. Variants in mothers against decapentaplegic (*SMAD*)

genes, encoding downstream mediators of BMP signaling, contribute rarely.

A number of genes outside of the TGF- β signaling pathway have also been identified as PAH risk genes. Variants in developmental transcription factors, *TBX4* and *SOX17*, are enriched in pediatric patients [7, 16–18]. Each gene contributes to 7–8% of pediatric IPAH and ~5% (*TBX4*) or ~3.2% (*SOX17*) of pediatric APAH-CHD [19]. Originally described as a determinant of pattern formation including limb development [20], the association of *TBX4* with PAH, cardiac defects [21, 22], and a variety of developmental lung disorders [22, 23] indicate an expanding role for *TBX4* in embryonic development. Biallelic variants in eukaryotic initiation translation factor (*EIF2AK4*) cause pulmonary veno-occlusive disease (PVOD) and pulmonary capillary hemangiomatosis (PCH) [24, 25]. Loss of function variants in channelopathy genes potassium two pore domain channel (*KCNK3*) [26] and ATP-binding cassette subfamily member 8 (*ABCC8*) [27], as well as membrane reservoir gene caveolin-1 (*CAVI*) [28–30], are causative for PAH. Recent associations of variants in ATPase 13A3 (*ATP13A3*) and aquaporin 1 (*AQP1*) [8], as well as kallikrein 1 (*KLK1*) and gamma-glutamyl carboxylase (*GGCX*) [7], have been reported but require independent confirmation. Finally, a role for de novo variants in pediatric-onset PAH has been suggested based on a cohort of 34 child-parent trios [17].

Together, these data indicate that rare genetic variants underlie ~75–80% of FPAH [6], at least 10% of adult-onset idiopathic PAH (IPAH) [7, 8], and up to ~36% of pediatric-onset IPAH [31]. A substantial fraction of non-familial PAH cases remains genetically undefined. The low frequency of risk variants for each gene, except *BMPR2*, indicates that large numbers of individuals are required for further validation of rare risk genes and pathways, and to understand the natural history of each genetic subtype of PAH. Towards this end, we analyzed 4175 PAH cases from an international consortium with ES or GS. The National Biological Sample and Data Repository for PAH (aka PAH Biobank) was comprised of 2570 PAH cases (1110 IPAH and 1239 APAH) and the UK NIHR BioResource – Rare Diseases Study was comprised of 1144 cases, almost entirely IPAH. Thus, the increased power of the combined cohort was a 2-fold increase in the number of IPAH cases, and we focused our association analyses on this PAH subclass. The cohort size precluded testing of the oligogenicity hypothesis suggested by the incomplete penetrance of known PAH risk genes. Non-inherited de novo mutations could also contribute to genetically unexplained non-familial cases but require access to parental sequencing data. We previously showed that pediatric-onset PAH cases were enriched with damaging de novo variants. Here, we

expand the analysis to a cohort of 124 pediatric child-parent trios.

Methods

Patient cohorts and control datasets

A total of 4175 PAH cases from the National Biological Sample and Data Repository for PAH (PAH Biobank, $n = 2570$ exomes) [7], UK NIHR BioResource – Rare Diseases Study (UK NIHR BioResource, $n = 1144$ genomes) [8], and the Columbia University Irving Medical Center (CUIMC, $n = 461$ exomes) [17, 18, 27] were included in a combined analysis of rare inherited variants. The subset of 124 affected child-unaffected parents trios ($n = 111$ CUIMC, $n = 8$ UK NIHR BioResource, $n = 5$ PAH Biobank) were included in an analysis of de novo variants. An additional 65 *BMPR2* mutation-positive cases from CUIMC without exome sequencing data were previously reported [17, 18] and included in the overall cohort counts (total of 4241 cases). As previously described, cases were diagnosed by medical record review including right heart catheterization and all were classified as World Symposium on Pulmonary Hypertension (WSPH) Group I [32]. Written informed consent for publication was obtained at enrollment. The studies were approved by the institutional review boards at CCHMC, individual PAH Biobank Centers, the East of England Cambridge South national research ethics committee (REC, ref. 13/EE0325) or CUIMC.

The control group consisted of unaffected parents from the Simons Powering Autism Research for Knowledge (SPARK) study (exomes) [33] as well as gnomADv2.1.1 (gnomAD) individuals (genomes).

ES/GS data analysis

PAH Biobank, CUIMC, and SPARK cohort samples were all sequenced in collaboration with the Regeneron Genetics Center as previously described [7, 8, 17, 18, 27]; the UK NIHR BioResource sequence data were also previously described [8]. For case and SPARK control data, we used a previously established bioinformatics procedure [34] to process and analyze exome and genome sequence data. For the UK NIHR BioResource data, we extracted reads from GS data by the following procedure: (1) obtained all reads that were mapped to the human genome regions that overlapped with the target regions of xGEN exome capture intervals (Exome Research panel 1.0); (2) the mate pairs of these reads. We then processed the extracted GS data using the same pipeline as the ES data. Specifically, we used BWA-MEM [35] to map and align paired-end reads to the human reference genome (version GRCh38/hg38, accession GCA_000001405.15), Picard v1.93 MarkDuplicates to identify and flag PCR duplicates, and GATK v4.1 [36, 37] HaplotypeCaller in Reference Confidence Model

mode to generate individual-level gVCF files from the aligned sequence data. We then performed joint calling of variants from all three datasets using *GLnexus* [38]. We used the following inclusion rules to select variants for downstream analysis: AF < 0.05% in the cohort, < 0.01% in gnomAD exome_ALL (all ancestries); > 90% target region with dp \geq 10; mappability = 1; and allele balance \geq 0.25. We also ran DeepVariants [39, 40], a new tool based on machine learning, for all cases and SPARK controls. We used the ES mode for ES data and GS mode for GS data, and then filtered by “PASS” DeepVariants. Inclusion criteria for variants observed in multiple carriers was \geq 50% of all calls PASS DeepVariants. For gnomAD data, only variants located in xGen-captured protein-coding regions were used; filtering was based on GATK metrics obtained from gnomAD and only “PASS” variants were included. SNVs with VQSR < 20 and indels with VQSR < 5 were excluded. Variants used for downstream analyses were restricted to the subset called by both *GLnexus* and DeepVariants.

De novo variants were defined as a variant present in the offspring with homozygous reference genotypes in both parents. We used a series of filters to identify de novo variants: VQSR tranche \leq 99.7 for SNVs and \leq 99.0 for indels; GATK Fisher Strand \leq 25; quality by depth \geq 2. We required the candidate de novo variants in probands to have \geq 5 reads supporting the alternative allele, \geq 20% alternative allele fraction, Phred-scaled genotype likelihood \geq 60 (GQ), and population AF \leq 0.01% in ExAC and required both parents to have \geq 10 reference reads, < 5% alternative allele fraction, and GQ \geq 30.

We used Ensembl Variant Effect Predictor (VEP; Ensemble 93) [41] to annotate variant function and ANNOVAR [42] to aggregate variant population frequencies and in silico predictions of deleteriousness. Rare synonymous variants were further evaluated with SpliceAI [43] to identify cryptic splice site variants (score \geq 0.5). Rare variants were defined as AF \leq 0.01% in gnomAD exome_ALL (all ancestries). A total of 18,939 protein-coding genes were identified containing \geq 1 rare variant, excluding mucin and major histocompatibility complex genes due to low sequence complexity. Deleterious variants were defined as likely gene-disrupting (LGD, including premature stop-gain, frameshift indels, canonical splicing variants, cryptic splice site variants, and exon deletions) or predicted damaging missense (D-Mis) based on gene-specific REVEL score thresholds [18, 44] (see below). All rare inherited and de novo variants in candidate genes were manually inspected using Integrative Genome Viewer (IGV) [45]. Indels were confirmed independently by Sanger sequencing.

Statistical analysis

To identify novel risk genes for IPAH, we performed a rare variant association test in unrelated participants of European ancestry. Genetic ancestry and relatedness of cases and SPARK controls were checked using Peddy [46], and only unrelated cases ($n = 2789$) and controls (18,819: 11,101 SPARK parents and 7718 gnomAD individuals) were included in the association test. The gnomAD controls were confined to non-Finnish Europeans (NFE). We performed a gene-based case-control test comparing the frequency of rare deleterious variants in PAH cases with unaffected controls. To reduce batch effects in combined datasets from different sources [47], we limited the analysis to regions targeted by xGen and with at least 10 \times coverage in 90% of samples. We then tested for similarity of the rare synonymous variant rate among cases and controls, assuming that most rare synonymous variants do not have discernible effects on disease risk.

To identify PAH risk genes, we tested the burden of rare deleterious variants (AF \leq 0.01%, LGD or D-Mis) in each protein-coding gene in cases compared to controls using a variable threshold test [48]. Specifically, we used REVEL [44] scores to predict the deleteriousness of missense variants, searched for a gene-specific optimal REVEL score threshold that maximized the burden of rare deleterious variants in cases compared to controls, and then used permutations to calculate statistical significance as described previously [7] to control the type I error rate. We checked for inflation using a quantile-quantile (Q-Q) plot and calculated the genomic control factor, lambda, using QQperm (<https://cran.r-project.org/web/packages/qqperm/qqperm.pdf>). Lambda equal to 1 indicates no deviation from the expected distribution. We performed two association tests, one with LGD and D-Mis variants combined and the other with D-Mis variants alone. We defined the threshold for genome-wide significance by Bonferroni correction for multiple testing ($n = 40,000$, 18,939 protein-coding genes containing rare variants times two tests for each gene, yielding a threshold p value = $1.25e-6$). We used the Benjamini-Hochberg procedure to estimate false discovery rate (FDR) by p.adjust in R.

To test whether recurrent variants in individual genes represented independent mutational events or were due to founder events, we first tested for relatedness among samples using KING [49], in addition to Peddy [46]. None of the cases with recurrent variants had any evidence of relatedness. Second, we assessed shared haplotypes of recurrent variant carriers using SHAPEIT2 [50] and the HapMap genetic map [51]. Since all of the recurrent variant carriers were of European ancestry, we restricted the HapMap data to the European population.

To estimate the burden of de novo variants in cases, we calculated the background mutation rate using a previously published tri-nucleotide change table [52, 53] and calculated the rate in protein-coding regions that are uniquely mappable. We assumed that the number of de novo variants of various types (e.g., synonymous, missense, LGD) expected by chance in gene sets or all genes followed a Poisson distribution [52]. For a given type of de novo variant in a gene set, we set the observed number of cases to $m1$, the expected number to $m0$, estimated the enrichment rate by $(m1/m0)$, and tested for significance using an exact Poisson test (`poisson.test` in R) with $m0$ as the expectation.

Protein modeling

Homology structures of conserved protein domains in FBLN2 and PDGFD were built using EasyModeller 4.0 [54]. Template structures were downloaded from the protein database (PDB) for endothelial growth factor (EGF, PDB ID 5UK5) and CUB (PDB ID 3KQ4) domains. The template structure for platelet-derived growth factor (PDGF)/vascular EGF (VEGF) was downloaded directly from PrePPI [55, 56].

Gene expression

Single-cell RNA-seq data of aorta, lung, and heart tissues were obtained from *Tabula Muris*, a transcriptome compendium containing RNA-seq data from ~100,000 single cells from 20 adult-staged mouse organs [57]. We chose 14 tissue/cell types including endothelial, cardiac muscle, and stromal cells from the three tissues, restricting the analysis to tissues/cell types for which there was RNA-seq data from at least 70 individual cells (Additional file 1, Supplementary Figure 1). Relative gene expression was based on the fraction of cells with > 0 reads in each cell type. PCA of cell type-specific gene expression profiles was performed using a script available through GitHub [58].

Results

Cohort characteristics

Demographic data and mean hemodynamic parameters of the combined US/UK cohort are shown in Table 1. The cohort includes 4241 cases: 54.6% IPAH, 34.8% APAH, 5.9% FPAH, and 4.6% other PAH. Most of the APAH and other PAH cases came from the PAH Biobank and have been described previously [7]. The majority of cases were adult-onset (92.6%) with a mean age-of-diagnosis (by right heart catheterization) of 45.9 ± 20 years (mean \pm SD). As expected for adult-onset PAH cohorts [7, 8, 59], the majority of cases were female (75.1%). The genetically determined ancestries were European (74.5%), Hispanic (8.6%), African (8.7%), East Asian (2.5%), and South Asian (2.8%). Hemodynamic

data were collected at the time of PAH diagnosis. Diagnostic criteria for PAH is mean pulmonary arterial pressure (mPAP) > 20–25 mmHg [32]. The mPAP and mean pulmonary capillary wedge pressure (mPCWP) for the overall cohort were 51 ± 14 mmHg (mean \pm SD) and 10 ± 4 mmHg, respectively, compared to 58 ± 14 mmHg and 10 ± 4 mmHg for FPAH.

A comparison of the clinical characteristics and hemodynamic data for pediatric- versus adult-onset PAH cases is shown in Additional file 2 (Supplementary Table 1). Notably, the female:male ratio among pediatric-onset cases was significantly lower (1.65:1) compared to adult-onset cases (4:1, $p < 0.0001$ by Fisher's exact test), and children had higher mPAP and mPCWP, decreased cardiac output and increased pulmonary vascular resistance compared to adults at diagnosis (all differences $p < 0.0001$ by Student's t test).

Rare deleterious variants in *BMPR2* were identified in 7.7% of cases overall (209/2318, 9% of IPAH; 108/191, 56.6% of FPAH; and 13/1475, 0.88% of APAH). The variants include LGD and D-Mis variants as well as intra-genic or whole gene deletions as previously described [7, 8, 17, 18]. The percentage of *BMPR2* carriers in the US/UK international cohort is lower than previous reports [8, 12] due to the enrichment of APAH cases, rarely caused by *BMPR2* variants [7, 18].

Identification of novel risk genes: *FBLN2* and *PDGFD*

To perform a combined analysis of US and UK sequencing data, we reprocessed the UK data using our inhouse pipeline, including predictions of missense variant deleteriousness [7]. Quality control procedures included detection of cryptic relatedness among all PAH participants. We performed a gene-based case-control association analysis to identify novel PAH risk genes using only unrelated cases. To control for population stratification, we confined the association analysis to individuals of European ancestry (2789 cases, 18,819 controls) and then screened the whole cohort, including nonEuropeans, for rare deleterious variants in associated genes. As a quality control check for the filtering parameters employed, we compared the frequencies of rare synonymous variants, a variant class that is mostly neutral with respect to disease status, in European cases vs controls. We observed similar frequencies of synonymous variants in cases vs controls (enrichment rate = 1.0, p value = 0.28) (Additional file 2, Supplementary Table 2). Furthermore, a gene-level burden test revealed no enrichment of rare synonymous variants in cases (Additional file 1, Supplementary Figure 2). We then proceeded to test for gene-specific enrichment of rare deleterious variants (AF < 0.01%, LGD and D-Mis, or D-Mis only) in cases compared to controls. We note that to improve power, we empirically determined the

Table 1 Demographic data and mean hemodynamic parameters from the US/UK PAH cohort*

	All	IPAH	APAH**	FPAH	Other***
Total, n (%)	4241	2319 (54.6)	1479 (34.8)	252 (5.9)	191 (4.6)
Sex, n (%)					
F	3187 (75.1)	1721 (74.2)	1156 (78.2)	173 (68.6)	137 (71.7)
M	1054 (24.9)	598 (25.8)	323 (21.8)	79 (31.3)	54 (28.3)
F:M ratio	3:1	2.9:1	3.6:1	2.3:1	2.5:1
Age-of-onset, n (%)					
Adult (\geq 18 years)	3780 (89.2)	2126 (91.7)	1256 (85.2)	213 (84.5)	185 (96.9)
Child (< 18 years)	457 (10.8)	193 (8.3)	219 (14.8)	39 (15.4)	6 (3.1)
Mean \pm SD	45.9 \pm 20.0	47.0 \pm 19.5	45.2 \pm 21.3	36.8 \pm 16.8	47.7 \pm 15.0
Ancestry, n (%)					
EUR	3108 (74.5)	1798 (77.5)	988 (67.0)	166 (86.9)	156 (81.7)
HISP	359 (8.6)	145 (6.3)	186 (12.6)	13 (6.8)	15 (7.9)
AFR	365 (8.7)	166 (7.2)	189 (12.8)	2 (1.0)	8 (4.2)
EAS	104 (2.5)	42 (1.8)	57 (3.9)	1 (0.5)	4 (2.1)
SAS	116 (2.8)	82 (3.5)	28 (1.9)	4 (2.1)	2 (1.0)
Other/unknown	123 (2.9)	85 (3.7)	27 (1.8)	5 (2.6)	6 (3.1)
Hemodynamic parameters, mean \pm SD (n)					
MPAP, mmHg	51 \pm 14 (3594)	53 \pm 17 (2045)	48 \pm 14 (1235)	58 \pm 14 (158)	51 \pm 12 (156)
<i>BMPR2+</i>	59 \pm 12 (320)	60 \pm 12 (197)	56 \pm 16 (11)	58 \pm 12 (109)	53 \pm 5 (3)
<i>BMPR2-</i>	50 \pm 14 (3394)	52 \pm 14 (1928)	48 \pm 14 (1257)	57 \pm 17 (65)	51 \pm 13 (144)
MPCWP, mmHg	10 \pm 4 (3407)	10 \pm 24 (1912)	10 \pm 4 (1197)	10 \pm 4 (151)	11 \pm 4 (147)
<i>BMPR2+</i>	10 \pm 4 (0)	10 \pm 4 (0)	10 \pm 3 (0)	10 \pm 4 (0)	8 \pm 4 (0)
<i>BMPR2-</i>	10 \pm 4 (0)	10 \pm 4 (0)	10 \pm 4 (0)	10 \pm 4 (0)	11 \pm 4 (0)

*US/UK PAH cohort: 2572 PAH Biobank; 1134 UK NIHR BioResource; 534 CUIMC

**APAH: PAH associated with connective tissue diseases, congenital heart disease, HHT, HIV

***Other: diet- and toxin-induced PAH, non-familial pulmonary veno-occlusive disease/pulmonary capillary hemangiomatosis and one case of persistent pulmonary hypertension of the newborn

Abbreviations: *AFR*, African; *EAS*, East Asian; *EUR*, European; *HISP*, Hispanic; *SAS*, Southeast Asian; *MPAP*, mean pulmonary artery pressure; *MPCWP*, mean pulmonary capillary wedge pressure

optimal REVEL score threshold to define deleterious missense variants in a gene-specific manner using a variable threshold test [7]. To account for potential different modes of action for different risk genes, we tested the association twice for each gene: one with LGD and D-Mis variants and the other with D-Mis variants alone. In this approach, LGD and D-Mis together is optimized for complete or partial loss of function; D-Mis alone is optimized for gain of function or dominant negative variants. We set the total number of tests at twice the number of protein-coding genes for multiple test adjustment, a conservative approach considering that the data used in these two tests per gene are not independent. The Q-Q plot of *p* values from tests in all genes shows negligible genomic inflation (Additional file 1, Supplementary Figure 3). Rare deleterious variants in eleven genes were significantly associated (false discovery rate, FDR < 0.1) with PAH. Among these, seven are known or previously reported candidate PAH risk genes: *BMPR2*,

TBX4, *GDF2*, *ACVRL1*, *SOX17*, *AQPI*, *ATP13A3*, and *KDR*. Three are new candidate genes: *COL6A5* (collagen type VI alpha 5 chain), *JPT2* (Jupiter microtubule-associated homolog 2), and *FBLN2* (fibulin 2).

The increased power inherent to the combined cohort over the PAH Biobank or UK NIHR BioResource alone is due to a twofold increase in the number of IPAH cases, including the number of European cases used for association analysis. Power analyses indicated that the study had ample power to detect risk genes with large effect size and modest variant allele frequency, or large variant allele frequency and modest effect size, relative to IPAH risk genes identified in smaller cohorts (Additional file 1, Supplementary Figure 4). To take advantage of the increased number of European IPAH cases in the combined cohort, we then restricted the analysis to IPAH. Again, testing for association across all protein-coding genes for 1647 IPAH cases compared to 18,819 controls was generally consistent with expectation under

the null model (Fig. 1). Rare predicted deleterious variants in seven genes were significantly associated (FDR < 0.1) with IPAH, including three known genes (*BMPR2*, *GDF2*, and *TBX4*), two recently identified candidate genes (*SOX17* and *KDR*), and two new candidate genes (*FBLN2*, and *PDGFD*, platelet-derived growth factor D). More than 95% of samples for both cases and controls had at least 10× depth of sequence coverage across the target regions for *FBLN2* and *PDGFD* (Additional file 1, Supplementary Figure 5), excluding the possibility that the associations were driven by coverage differences between cases and controls. We also tested for gene-level associations restricting the analysis to European APAH cases ($n = 998$). The Q-Q plot of p values from all gene tests is shown in Additional file 1, Supplementary Figure 6. Known PAH gene *ACVRL1* showed association with APAH, consistent with its role in APAH-HHT, but no genes were significantly associated at FDR < 0.1.

KDR has recently been implicated as a causal gene for PAH based on a small familial study [60] and our population-based phenotype-driven (SKAT-O) analysis of the UK NIHR BioResource cohort with replication in the PAH Biobank [61]. Both of those analyses were based on protein-truncating variants. Herein, we provide additional statistical evidence based on a burden test including both LGD and D-Mis variants using our variable threshold method. Six cases (5 IPAH, 1 APAH-CHD) carry D-Mis variants with empirically determined REVEL $L > 0.86$; details of the variants are provided in Supplementary Table 3. All of the variants are located in the conserved tyrosine kinase domain of the encoded protein (www.uniprot.org). One of the variants, c.3439C>T is recurrent in three cases. There was no evidence of relatedness for these cases, and the relatively short shared haplotype length and common population frequency

(Additional file 2, Supplementary Table 4) indicate that the variant occurrences represent independent mutational events rather than being derived from a founder event. None of these cases have variants in other known PAH risk genes. The age-of-onset for the six cases is 57 ± 20 years (mean \pm SD, range 25-75 years) and all are of European ancestry. Statistically significant association following Bonferroni correction for multiple testing provides confirmation of the association of *KDR* with PAH using an alternative burden-based statistical method.

The associations of *FBLN2* and *PDGFD* were both driven by D-Mis variants. We next screened the entire combined cohort, including participants of non-European ancestry, for rare deleterious missense variants in *FBLN2* and *PDGFD*. In total, seven cases carry *FBLN2* variants (6 IPAH, 1 APAH) and ten cases carry *PDGFD* variants (9 IPAH, 1 PAH associated with diet and toxins) (Table 2). Most of the carriers are of European ancestry; one *FBLN2* carrier is of East Asian ancestry and one *PDGFD* carrier is of African ancestry. One *FBLN2* variant ((c.2944G>T; p.(Asp982Tyr)) and two *PDGFD* variants ((c.385G>A; p.(Glu129Lys) and c.961 T>A; p.(Tyr321Asn)) were recurrent in the cohort. Again, there was no evidence of relatedness among these cases, and the shared haplotype characteristics (Additional file 2, Supplementary Table 4) indicate that the variants occurred as independent mutational events. Locations of the predicted damaging missense amino acid residues are shown in Fig. 2. *FBLN2* contains multiple endothelial growth factor (EGF) domains, and *PDGFD* contains a conserved CUB domain and a platelet-derived growth factor (PDGF)/vascular EGF (VEGF) domain. All of the *FBLN2* and eight out of ten *PDGFD* D-Mis variants, occur in conserved protein domains. *FBLN2* p.(Gly880-Val) and p.(Gly889Asp) replace conserved reverse turn

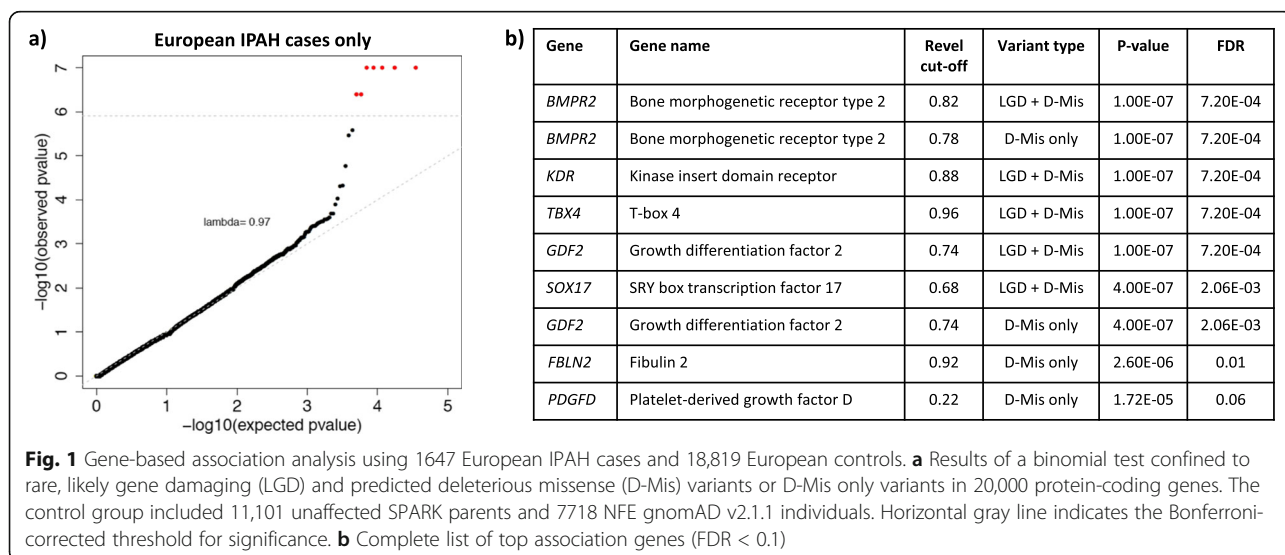


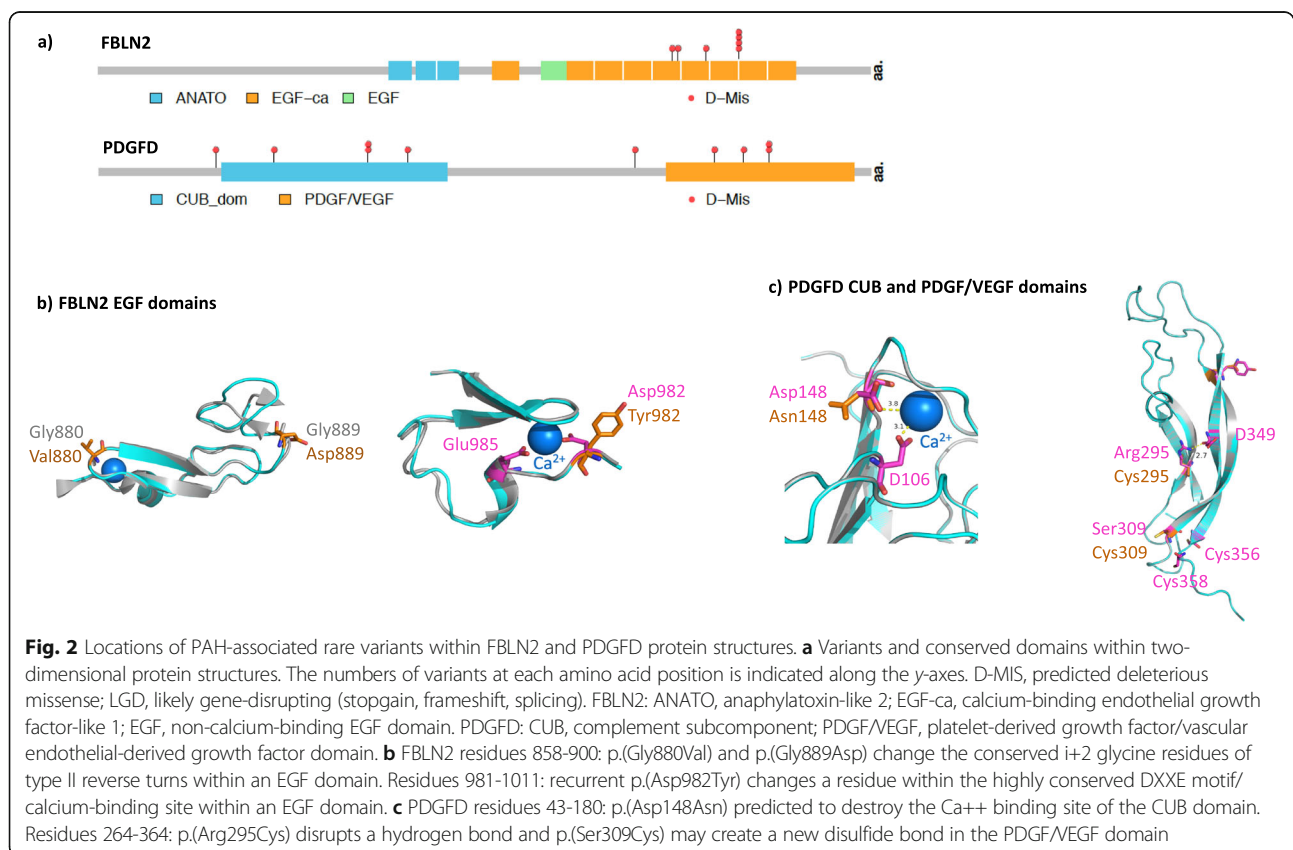
Table 2 Rare predicted deleterious *FBLN2* and *PDGFD* variants* among 4175 PAH cases**

Case ID	Sex	Age _{dx}	PAH subclass	Ancestry	Gene ***	Exon	Nucleotide change	Amino acid change	Variant type	MAF (gnomAD exomes)	CADD score	Revel
08-018	F	70	IPAH	EUR	<i>FBLN2</i>	12	c.2639G>T	p.(Gly880Val)	D-Mis	1.63E-05	27.1	0.94
17-035	F	41	APAH	EAS	<i>FBLN2</i>	12	c.2666G>A	p.(Gly889Asp)	D-Mis	–	27.6	0.94
12-207	F	44	IPAH	EUR	<i>FBLN2</i>	13	c.2794T>C	p.(Phe923Leu)	D-Mis	2.11E-05	29.6	0.92
23-001	M	66	IPAH	EUR	<i>FBLN2</i>	14	c.2944G>T	p.(Asp982Tyr)	D-Mis	1.88E-05	34.0	0.95
29-031	F	57	IPAH	EUR	<i>FBLN2</i>	14	c.2944G>T	p.(Asp982Tyr)	D-Mis	1.88E-05	34.0	0.95
34-005	M	69	IPAH	EUR	<i>FBLN2</i>	14	c.2944G>T	p.(Asp982Tyr)	D-Mis	1.88E-05	34.0	0.95
W000210	F	52	IPAH	EUR	<i>FBLN2</i>	14	c.2944G>T	p.(Asp982Tyr)	D-Mis	1.88E-05	34.0	0.95
W000073	F	40	IPAH	AFR	<i>PDGFD</i>	2	c.166G>A	p.(Gly56Ser)	D-Mis	1.99E-05	22.9	0.64
JM950	F	2	IPAH	EUR	<i>PDGFD</i>	2	c.250C>T	p.(Arg84Trp)	D-Mis	1.59E-05	16.4	0.51
E012465	F	55	IPAH	EUR	<i>PDGFD</i>	3	c.385G>A	p.(Glu129Lys)	D-Mis	–	25.2	0.262
E014342	F	40	IPAH	EUR	<i>PDGFD</i>	3	c.385G>A	p.(Glu129Lys)	D-Mis	–	25.2	0.262
E014400	F	43	IPAH	EUR	<i>PDGFD</i>	3	c.442G>A	p.(Asp148Asn)	D-Mis	7.97E-06	25.2	0.41
E000844	F	39	IPAH	EUR	<i>PDGFD</i>	6	c.770T>C	p.(Leu257Pro)	D-Mis	4.01E-06	31.0	0.62
13-037	M	43	DTOX	EUR	<i>PDGFD</i>	6	c.883C>T	p.(Arg295Cys)	D-Mis	4.00E-06	35.0	0.56
23-025	F	41	IPAH	EUR	<i>PDGFD</i>	6	c.926C>G	p.Ser309Cys	D-Mis	–	28.4	0.22
E000820	F	73	IPAH	EUR	<i>PDGFD</i>	6	c.961T>A	p.(Tyr321Asn)	D-Mis	1.21E-05	33.0	0.34
E010173	F	74	IPAH	EUR	<i>PDGFD</i>	6	c.961T>A	p.(Tyr321Asn)	D-Mis	1.21E-05	33.0	0.34

*Rare, deleterious variants defined as gnomAD_exome_ALL AF $\leq 1.00E-04$ and LGD or missense with variable REVEL cut-off (*FBLN2* 0.92 and *PDGFD* 0.22)

** Cases are heterozygous for the indicated variants

***Transcripts: *FBLN2* NM_001998.3 and *PDGFD* NM_033135.4



residues in an EGF domain which may change the conformation of the domain and impact protein function (Fig. 2b). Recurrent *FBLN2* p.(Asp982Tyr) disrupts a Ca^{++} binding site [62] in another EGF domain (Fig. 2b), which may reduce the affinity and frequency of Ca^{++} binding. *PDGFD* p.(Asp148Asn) disrupts a Ca^{++} binding site within the CUB domain [63] (Fig. 2c) and recurrent *PDGFD* p.(Tyr321Asn) is predicted to disrupt a hydrogen bond within the PDGF/VEGF domain (Fig. 2c). In addition, *PDGFD* p.(Arg295Cys) is located in close proximity to Cys356 and Cys358, potentially introducing new disulfide bonds within the PDGF/VEGF domain.

Clinical phenotypes of *FBLN2* and *PDGFD* variant carriers

The clinical phenotypes of all *FBLN2* and *PDGFD* variant carriers are provided in Table 3. *FBLN2* variant carriers have a similar female:male ratio (2.5:1) compared to the overall cohort (3.1:1) or IPAH alone (2.9:1). *PDGFD* variant carriers are primarily female (9:1) but the distribution is not significantly different from the overall IPAH cohort ($p = 0.5$, Fisher's exact test). All of the *FBLN2* and *PDGFD* variant carriers have adult-onset disease, with the exception of one pediatric *PDGFD* variant carrier, with no statistically significant differences in mean age-of-onset (53 ± 11 and 45 ± 20 years, respectively) compared to that of the overall cohort (46 ± 20 years) or IPAH alone (47 ± 20 years), excluding *FBLN2* and *PDGFD* variant carriers. *FBLN2* variant carriers exhibit a trend towards increased mean pulmonary artery pressure (62 ± 17 , mmHg) and significantly increased mean pulmonary capillary wedge pressure (13 ± 2 mmHg) compared to the overall cohort (51 ± 14 , non-significant and 10 ± 4 mmHg, $p = 0.015$ respectively) or IPAH alone (53 ± 17 , non-significant and 10 ± 24 mmHg, $p = 0.01$, respectively). *PDGFD* variant carriers have similar pulmonary pressures compared to the overall cohort or IPAH alone. All of the *FBLN2* and *PDGFD* variant carriers were diagnosed with WHO PAH class II or III disease and have no history of lung transplantation. Most of the *FBLN2* and *PDGFD* variant carriers have comorbidities typical of adult IPAH patients [64, 65], including hypertension, hypothyroidism, other pulmonary diseases, and metabolic diseases. Five out of seven *FBLN2* carriers have a diagnosis of systemic hypertension.

Gene expression patterns of PAH candidate risk genes

We hypothesized that PAH risk genes are highly expressed in certain cell types relevant to the disease etiology and that joint analysis of cell type-specific expression data with genetic data could inform cell types associated with disease risk [66]. We obtained single-cell RNA-seq data of aorta, lung, and heart tissues available through the *Tabula Muris* project, a transcriptome

compendium containing RNA-seq data from adult-staged mouse organs [57]. We chose 14 tissue/cell types including endothelial, cardiac muscle, and stromal cells as a proxy for the cell types of the pulmonary artery (unavailable). A list of the tissues, cell types, and the number of cells sequenced per tissue/cell type is provided in Additional file 1 (Supplementary Figure 1a). We queried gene expression for twelve known PAH risk genes (*ACVRL1*, *BMPR2*, *CAV1*, *EIF2AK4*, *ENG*, *KCNK3*, *KDR*, *NOTCH1*, *SMAD4*, *SMAD9*, *SOX17*, *TBX4*) and the two new candidate risk genes (*FBLN2*, *PDGFD*). A heat map with hierarchical clustering of relative gene expression is shown in Fig. 3a. The majority of known risk genes (7/12) have relatively high expression in endothelial cells from the three tissues; most others have high expression in tissue-specific cardiac muscle, stromal cells, or fibroblasts. *PDGFD* is located in the same cluster as *BMPR2*, *SOX17*, and *KDR*; these genes are specifically and highly expressed in endothelial cell types. *FBLN2* is highly expressed in both endothelial and fibroblast cell types. We then randomly selected a set of 100 genes without reported associations with PAH and performed PCA of cell type-specific expression profiles of known risk genes and random genes. The second component (PC2) largely separates known risk genes and random genes (Fig. 3b, c). Consistent with hierarchical clustering, endothelial expression in all three tissues was positively correlated with PC2 (Additional file 1, Supplementary Figure 1b). Projecting all protein-coding genes onto PC2, seven of twelve known risk genes are ranked in top 5% among all genes (Fig. 3d) (binomial test: enrichment = 20, $p = 1.6E-05$). Two new candidate genes, *FBLN2* and *PDGFD*, are ranked in the top 1.8% of PC2.

Identification of novel candidate pediatric PAH risk genes by de novo variant analysis

We next focused on pediatric-onset disease, a subpopulation in which genetic factors likely play a larger causal role compared to adults. The study was underpowered to carry out a gene-based case-control association analysis due to the relatively small number of pediatric patients ($n = 442$); however, 124 pediatric-onset PAH probands with child-parent trio data were available for de novo variant analysis. The trio cohort consisted mostly of IPAH (55.6%, $n = 66$) and APAH-CHD (37.9%, $n = 45$) cases. We performed a burden test for enrichment of exonic de novo variants among all trio probands by comparing the number of variants observed vs expected based on the background mutation rate. Similar rates of de novo mutations were observed for synonymous, LGD alone, and total missense variants (Table 4). However, there was a significant burden of D-Mis and LGD + D-Mis variants among cases over that expected by chance (Table 4). Inclusion of all protein-

Table 3 Clinical phenotypes of *FBLN2* and *PDGFD* variant carriers. Sex ratios and mean \pm SD diagnostic age and hemodynamic values have been calculated separately for *FBLN2* and *PDGFD* variant carriers

Case ID	Sex	Age _{dx} (years)	PAH subclass	Ancestry	Gene	MPAP (mmHg)	MPCWP (mmHg)	WHO functional class	Lung tx	Other medical conditions
08-018	F	70	IPAH	EUR	<i>FBLN2</i>	58	14	III	No	HTN, kidney congenital anomaly, Paget's disease
17-035	F	41	APAH (MCTD)	EAS	<i>FBLN2</i>	43	16	III	No	
12-207*	F	44	IPAH	EUR	<i>FBLN2</i>	NA	NA	NA	No	HTN, hypothyroidism
23-001	M	66	IPAH	EUR	<i>FBLN2</i>	68	10	III	No	HTN, OLD (smoker)
29-031	F	57	IPAH	EUR	<i>FBLN2</i>	84	15	NA	No	HTN, mitral valve disease, hypothyroidism, OA, COPD
34-005*	M	69	IPAH	EUR	<i>FBLN2</i>	43	14	NA	No	HTN, CAD
W000210	F	52	IPAH	EUR	<i>FBLN2</i>	75	11	II	No	Hyposplenism
Mean \pm SD	2.5:	53 \pm 11				62 \pm 17	13 \pm 2			
W000073*	F	40	IPAH	AFR	<i>PDGFD</i>	73	NA	III	No	PFO, bilateral chronic subdural hematoma, hypothyroidism
JM930	F	2	IPAH	EUR	<i>PDGFD</i>	39	NA	NA	No	Bronchopulmonary dysplasia
E012465	F	55	IPAH	EUR	<i>PDGFD</i>	52	7	III	No	Hypothyroidism, IBS, major depression
E014342*	F	40	IPAH	EUR	<i>PDGFD</i>	57	7	III	No	Emphysema
E014400	F	43	IPAH	EUR	<i>PDGFD</i>	57	4	III	No	Obesity, T2DM
E000844	F	39	IPAH	EUR	<i>PDGFD</i>	51	9	III	No	GERD, asthma, bicornate uterus
13-037	M	43	DTOX	EUR	<i>PDGFD</i>	47	7	II	No	None
23-025*	F	41	IPAH	EUR	<i>PDGFD</i>	64	NA	III	No	Hypothyroidism
E000820	F	73	IPAH	EUR	<i>PDGFD</i>	48	9	II	No	Fatty liver, hypothyroidism, ductal carcinoma, gallstones, superior vena cava and azygos vein thrombosis related to port-a-cath
E010173	F	74	IPAH	EUR	<i>PDGFD</i>	32	9	III	No	MPVD, PVD, obesity, T2DM, HTN, chronic renal impairment, hypothyroidism, OA, hypouricemia, major depression
Mean \pm SD	9:	50 \pm 14				53 \pm 12	11 \pm 10			

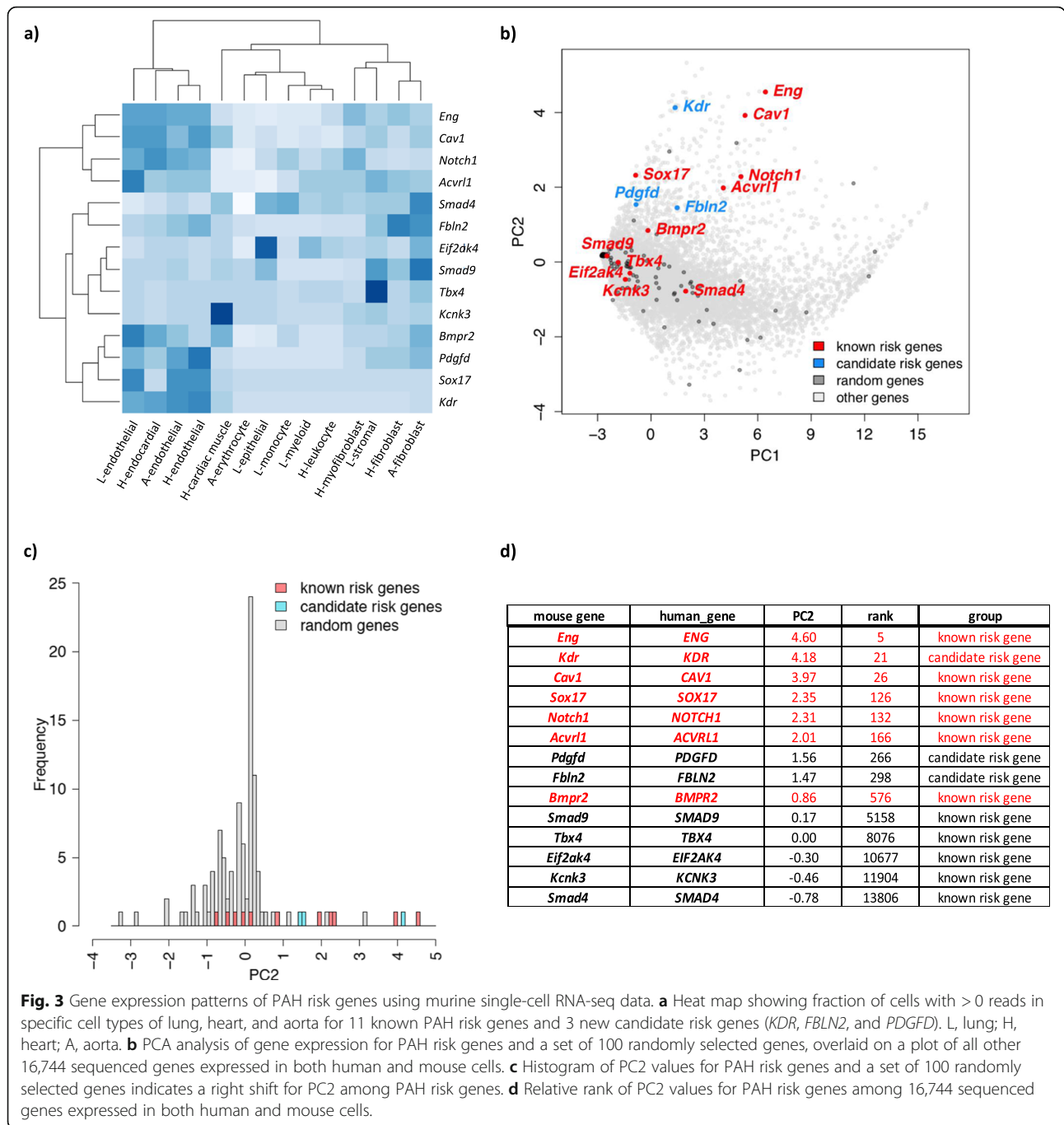
*Cases with risk variants in additional PAH risk genes: 12-207 (*ABCC8* and *GGCX*), 34-005 (*GGCX*), W000073 (*TBX4*), E014342 (*BMP2*), 23-025 (*ENG*)

**NS, Fisher's exact test

Abbreviations: *CAD*, coronary artery disease; *COPD*, chronic obstructive pulmonary disease; *GERD*, gastroesophageal reflux disease; *HTN*, systemic hypertension; *IBS*, irritable bowel syndrome; *MCTD*, mixed connective tissue disease; *MPVD*, mixed pulmonary valve disease; *OA*, osteoarthritis; *OLD*, obstructive lung disease; *PFO*, patent foramen ovale; *PVD*, peripheral vascular disease; *T2DM*, type 2 diabetes mellitus

coding genes ($n = 18,939$) in the burden test identified 44 rare variants, including 30 D-Mis and 14 LGD, in cases. Confining the test to a set of 5756 genes highly expressed in developing lung (murine E16.5 lung stromal cells) [67] or heart (murine E14.5 heart) [34] revealed a 2.45-fold enrichment of de novo variants among cases ($n = 19$ D-Mis, $n = 29$ LGD + D-Mis) over that expected by chance ($p = 2.0e-4$, $p = 2.5e-5$, respectively). We estimate that 17 of the variants are likely to be implicated in pediatric PAH based upon the enrichment over

controls or expected by chance. Among the variants, seven are in known PAH risk genes: four in *TBX4*, two in *BMP2*, and one in *ACVRL1*. Excluding these known risk genes, there are 22 LGD + D-Mis variants in genes highly expressed in developing heart and lung, still significantly more than expected (enrichment rate = 1.86, $p = 0.008$, 10 expected risk variants). We tested the burden of de novo variants among IPAH cases and observed enrichment of D-Mis and LGD + D-Mis variants similar to that of the overall trio cohort (Additional file 2,



Supplementary Table 5). The study was underpowered to detect a significant burden of de novo variants among APAH-CHD cases. The estimated fraction of pediatric IPAH and the overall pediatric cohort explained by de novo variants is 15.2% and 14.5%, respectively. A complete list of all rare, deleterious de novo variants carried by pediatric PAH cases is provided in Additional file 2 (Supplementary Table 6). Similar to other early-onset severe diseases, including CHD and bronchopulmonary dysplasia, the genes identified fit a general pattern for

developmental disorders—genes intolerant to loss of function variants ($pLI > 0.5$ for 40% of the genes) and with known functions as transcription factors, RNA-binding proteins, protein kinases, and chromatin modification. Three of the genes are known CHD risk genes (*NOTCH1*, *PTPN11*, and *RAF1*), and 37% of the genes are known causal genes for a variety of developmental syndromes. Case variant *PTPN11* p.(Asp61Gly) is a known causal variant for Noonan syndrome [68], and *RAF1* p.Pro261 is a hotspot for multiple gain-of-function

Table 4 Burden of de novo variants in pediatric-onset PAH ($n = 124$ child-parent trios)

	Variant type*	Observed	Expected by chance	Enrichment	p value	Estimated # of risk variants
All genes (18,939 genes)	SYN	42	38.3	1.1	0.51	N/A
	LGD	14	11.8	1.2	0.46	N/A
	MIS	93	84.7	1.1	0.36	N/A
	D-Mis	30	17.3	1.73	0.005	N/A
	LGD + D-Mis	44	28.9	1.52	0.009	15
HLE or HHE** (5756 genes)	SYN	18	14.01	1.28	0.28	N/A
	LGD	10	4.69	2.13	0.03	N/A
	MIS	40	31.68	1.26	0.15	N/A
	D-Mis	19	7.25	2.62	2.0e-4	12
	LGD + D-Mis	29	11.85	2.45	2.5e-5	17

* SYN, synonymous; LGD, likely gene-disrupting; MIS, missense; D-Mis, deleterious missense based on REVEL > 0.5

**HLE, high lung expression (murine E16.5 lung stromal cells); HHE, high heart expression (murine E14.5 heart)

mutations, including p.(Pro261Thr), causing Noonan syndrome [69].

Clinical phenotypes of pediatric de novo variant carriers

Among the 36 patients who carry LGD or D-Mis de novo variants (Additional file 2, Supplementary Table 7), there is a 1.8:1 ratio of females to males, a mean age-of-onset of 5.4 ± 4.6 years, 50% of the cases ($n = 18$) have a diagnosis of IPAH, 33.3% ($n = 12$) APAH-CHD and an overlapping but distinct 36.1% of cases have other congenital or growth and development anomalies. *NOTCH1* variant carrier, JM1357, has a diagnosis of APAH-CHD with tetralogy of Fallot, and a recent exome sequencing study of ~800 tetralogy of Fallot cases identified *NOTCH1* as the top association signal [70]. *PTPN11* variant carrier, JM155, has a diagnosis of APAH-CHD associated with Noonan syndrome and the c.182A>G variant is known to be pathogenic in Noonan syndrome. Variants in *PSMD12* cause Stankiewicz-Isidor syndrome, sometimes associated with congenital heart defects, and variant carrier 06-095 has a diagnosis of APAH-CHD. Hemodynamic data for the de novo variant carriers (Additional file 2, Supplementary Table 7) was similar to that of all pediatric cases in the cohort (Additional file 2, Supplementary Table 1).

Discussion

Combined analysis of a large US/UK cohort enriched in adult-onset IPAH cases enabled identification of five known and two new IPAH candidate risk genes with $FDR < 0.1$: *FBLN2* and *PDGFD* are the new genes. The association was based on a gene-level case-control analysis of 1647 unrelated European IPAH cases. The variants contributing to *FBLN2* and *PDGFD* associations are D-Mis variants predicted to alter highly conserved protein conformation, Ca^{++} binding sites, or intramolecular binding sites within conserved protein domains, likely

leading to important structural changes in critical domains. The non-founder nature of recurrent *FBLN2* p.(Asp982Tyr) ($n = 4$ cases), and two *PDGFD* variants recurrent in two unrelated cases each, adds further support for pathogenicity of these alleles. In addition, we confirmed the recent association of *KDR* with PAH [60, 61] based on an alternative statistical approach. We further show that all three of these candidate genes have high expression in lung and heart endothelial cell types, similar to other well established risk genes (*BMPR2* and *SOX17*), further supporting the plausibility of these genes contributing to PAH risk. De novo variant analysis of pediatric-onset PAH (124 trios) showed a 2.45× enrichment of rare deleterious exonic variants, indicating that de novo variants contribute to ~15% of pediatric cases across PAH subtypes. The de novo variants implicate new candidate risk genes likely unique to pediatric PAH, but some of the molecular pathways may inform both pediatric- and adult-onset PAH.

FBLN2 encodes an extracellular matrix protein important for elastic fiber formation and regulation of cell motility, proliferation, and angiogenesis. *FBLN2* is expressed in the lung vasculature but most studies have focused on gene expression in the heart vasculature. In mice, *Fbln2* is expressed in epithelial-mesenchymal transformation during embryonic heart development and is upregulated postnatally throughout coronary vasculogenesis and angiogenesis when transformed mesenchymal cells migrate to the extracellular matrix [71, 72]. *Fbln2*^{-/-} mice are viable, fertile, and have intact elastic fiber formation, attributable to compensation by the more widely expressed *Fbln1* gene [73, 74]. However, *Fbln2* expression is required for angiotensin II-induced TGFβ signaling and cardiac fibrosis [75]. In humans, *FBLN2* variants have been associated with atrioventricular septal defects [76] and intracranial aneurysm [77], providing additional support for a role in vascular

remodeling. We hypothesize that, in the pulmonary vasculature, gain of function variants may lead to increased TGF- β signaling, increased proliferation and medial hypertrophy. The *FBLN2* protein contains 10 consecutive EGF protein-protein interaction domains, nine of which are calcium-binding. All seven of the case variants are missense variants, two of which are predicted to alter the conformation of an EGF domain, and a recurrent variant carried by four cases is predicted to disrupt the Ca⁺⁺ binding site of another EGF domain. The carriers of *FBLN2* variants have adult-onset disease with mean age-of-onset similar to the overall cohort or IPAH alone. Five of seven carriers also have a diagnosis of systemic hypertension (HTN), and it is possible that gene damaging variants in *FBLN2* contribute to the development of HTN. However, given the frequency of HTN in the overall US/UK combined cohort (32% for adult-onset IPAH; similar to that reported in the REVEAL registry [65]), there may be other age-related genetic and environmental factors contributing to HTN. Finally, two of our cohort cases, 08-018 and 29-031, have additional diagnoses of renal or heart anomalies, and *FBLN2* has been identified as a key regulator of development in those tissues [78–80].

PDGFD is a member of the *PDGF* family that functions in recruiting cells of mesenchymal origin during development or to sites of injury [81]. *PDGFD* is widely expressed including arterial endothelial cells, adventitial pericytes and smooth muscle cells, lung endothelial cells, and smooth muscle cell progenitors of distal pulmonary arterioles. Secreted PDGFD specifically binds PDGFR β , a widely expressed protein that co-localizes with PDGFD in vascular smooth muscle cells. *Pdgfd* knockout mice are phenotypically normal with the exception of a modest increase in systemic blood pressure [82]. However, cardiac-specific *PDGFD* transgenic mice, overexpressing the active core domain of human PDGFD in the heart, exhibit vascular smooth muscle cell proliferation, vascular remodeling with wall thickening, severe cardiac fibrosis, heart failure, and premature death [83]. While effects of *Pdgfd* overexpression on the pulmonary vasculature have not been investigated, the cardiac vasculature data are consistent with a gain of function mechanism. Further evidence for the role of PDGFD as an effector molecule in cardiovascular diseases and cancer has been reviewed [81, 84, 85]. Full-length PDGFD contains two conserved protein domains, an autoinhibitory CUB domain and an enzymatic PDGF/VEGF domain; the protein undergoes proteolytic cleavage at Arg247 or Arg249 to produce an active growth factor promoting angiogenesis and vascular muscularization [86]. All ten of the case variants are missense variants; four reside in the CUB domain and five reside in the active processed protein. Variant p.(Asp148Asn), carried by two patients, is

predicted to disrupt the Ca⁺⁺ binding site of the CUB domain; variants p.(Arg295Cys) and p.(Ser309Cys), carried by one and two patients respectively, are predicted to alter the conformation of the PDGF/VEGF domain. All but one of the *PDGFD* variant carriers have adult-onset disease with mean age-of-onset similar to the overall cohort or IPAH alone. Four out of ten of the *PDGFD* variant carriers have additional diagnoses of other pulmonary fibrotic and/or vascular fibrotic diseases including bronchopulmonary dysplasia, emphysema, asthma, and one patient (E010173) with both mixed pulmonary valve disease and peripheral vascular disease (Table 3). Targeting the PDGF pathway with small molecule inhibitors of tyrosine kinase is an active area of investigation and several inhibitors are FDA-approved [87]. Notably, imatinib reduced cardiac fibroblast proliferation and PDGFD expression 15-fold [88]; data regarding effects on pulmonary arterial smooth muscle cells are warranted. A limitation of tyrosine kinase inhibitors is that they target multiple tyrosine kinases. Sequestering PDGFD with neutralizing antibodies or DNA/RNA aptamers, or preventing PDGFD-PDGFR β interaction via oligonucleotides, may provide more specific targeting.

To test the plausibility of the new candidate PAH genes identified by association analysis, we leveraged publicly available single-cell RNA-seq data. *PDGFD*, and recently identified *KDR*, have very similar expression patterns as *BMPR2* and *SOX17*, two established PAH genes. PCA indicated that the PAH risk genes can largely be separated from non-risk genes based on PC2. The majority of known PAH risk genes rank in the top 5% of PC2 among 16,744 genes queried, and the new genes—*FBLN2* and *PDGFD*—rank within the top 1.8%, providing support for their candidacy as PAH risk genes. Other risk genes, like *KCNK3* and *EIF2AK4*, exert important PAH-related functions in cell types other than endothelial cells, and *GDF2* is excreted from liver; thus, it will be important to consider expression patterns on a gene-specific basis. In addition, the dataset utilized in this study was based on adult-staged murine cells and is not well-suited for developmental genes such as *TBX4* and other genes likely to contribute to pediatric-onset disease. Thus, additional datasets from different time points are needed.

Rare deleterious variants in established PAH genes are clearly pathogenic based on segregation analyses, enrichment of rare deleterious variants in PAH cases compared to controls with replication over time, and demonstrated loss of function or aberrant function in vitro, in vivo (model organisms), or ex vivo [6]. However, none of the PAH genes are fully dominant and many carriers are never diagnosed with PAH. *BMPR2* variants exhibit variable penetrance with ~14%

penetrance in male carriers and 42% in females, suggesting sex as an important modifier of penetrance [6]. Other factors influencing penetrance likely include additional genetic factors, epigenetic factors [89], environmental factors, and gene \times environment interactions. Explicit testing of oligogenicity for rare diseases, or gene-environment interactions, require much larger cohorts than those currently available for PAH. However, as more putative risk genes are identified and more PAH cases are studied [7, 8, 15, 17], formal tests to assess the contributions of multiple genetic and environmental risks should be included. In the current study, five of the seventeen cases identified with rare deleterious variants in *FBLN2* or *PDGFD* also carry variants in one or two established or recently reported candidate risk genes. For example, participant 12-207 carries variants in *FBLN2* as well as *ABCC8* and *GGCX*, and participant W000073 carries variants in *PDGFD* and *TBX4*. We acknowledge the possibility that at least some of the variants identified to date may not be causal and that the relative contribution of individual variants requires further investigation. How multiple rare variants interact to affect PAH pathogenesis, penetrance, endophenotypes, or clinical outcomes will require much larger cohorts and will be one of the major aims of future large international consortia.

Our pediatric data indicate that children present with slightly higher mean pulmonary arterial pressure, decreased cardiac output, and increased pulmonary vascular resistance compared to adults at diagnosis. The early age-of-onset and increased severity of clinical phenotypes suggest that there may be differences in the genetic underpinnings. De novo mutations have emerged as an important class of genetic factors underlying rare diseases, especially early-onset severe conditions [34, 90], due to strong negative selection decreasing reproductive fitness [91]. Pediatric-onset PAH fits this category of diseases based on the high mortality during childhood [92–96]. Previously, we reported an enrichment of de novo variants in a cohort of 34 PAH probands with trio data [17]. We have now expanded this analysis to 124 trios with pediatric-onset PAH probands and confirmed the 2.45 \times enrichment of de novo variants in cases compared to the expected rate. Seven of the variant carriers have variants in known PAH risk genes (*TBX4*, *BMPR2*, *ACVRL1*), and three of the APAH-CHD variant carriers have variants in known CHD or CHD-associated risk genes (*NOTCH1*, *PTPN11*, *PSMD12*). We previously reported rare inherited LGD or D-Mis variants in CHD risk genes *NOTCH1* ($n = 5$), *PTPN11* ($n = 1$), and *RAF1* ($n = 2$) carried by APAH-CHD cases [18]. Specific inhibition of the protein encoded by *PTPN11* (SHP2) [97], and induction of mir-204 which negatively targets SHP2 [98], improved right ventricular function in the

monocrotaline rat model of PAH, suggesting a more general role of *PTPN11* in PAH.

At least eight of the other genes with case-derived de novo variants have plausible roles in lung/vascular development but have not been previously implicated in PAH: *AMOT* (angiominin), *CSNK2A2* (casein kinase 2 alpha 2), *HNRNPF* (heterogeneous nuclear ribonucleoprotein F), *HSPA4* (heat shock protein family A member 4), *KDM3B* (lysine demethylase 3B), *KEAP1* (kelch-like ECH-associated protein 1), *MECOM* (MDS1 and EVI1 complex locus), and *ZMYM2* (zinc finger MIM-type containing 2). A common single-nucleotide polymorphism in *MECOM* has been implicated in systemic blood pressure [99]. *KEAP1* encodes the principle negative regulator of transcription factor NF-E2 p45-related factor 2 (NRF2). The NRF2-KEAP1 partnership provides an evolutionarily conserved cytoprotective mechanism against oxidative stress. Under normal conditions, KEAP1 targets NRF2 for ubiquitin-dependent degradation and represses NRF2-dependent gene expression. KEAP1 is ubiquitously expressed and aberrant oxidative stress response in the pulmonary vasculature is a recognized mechanism underlying PAH. Together, our analysis indicates that 15% of PAH cases are attributable to de novo variants. A larger pediatric cohort will be necessary to confirm some of these genes via replication and identify additional new genes and pathways that will likely be unique to children and not identifiable through studies of adults with PAH.

Conclusions

We have identified *FBLN2* and *PDGFD* as new candidate risk genes for adult-onset IPAH, accounting for 0.26% and 0.35% of 2318 IPAH cases in the US/UK combined cohort, respectively. We note that five of seven *FBLN2* variant carriers also have a diagnosis of systemic hypertension. A few cases carry rare variants in more than one PAH risk gene, consistent with oligogenic nature of PAH in some individuals. Analysis of single-cell RNA-seq data shows that the new candidate genes have similar expression patterns to well-known PAH risk genes, providing orthogonal support for the new genes and providing more mechanistic insight. We estimate that \sim 15% of all pediatric cases are attributable to de novo variants and that many of these genes are likely to have important roles in developmental processes. Larger adult and pediatric cohorts are needed to better clinically characterize these rare genetic subtypes of PAH.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-021-00891-1>.

Additional file 1: Supplementary Figure 1. Selection of single-cell RNAseq data. **Supplementary Figure 2.** Gene-level burden test for rare

synonymous variants. **Supplementary Figure 3.** Gene-based association analysis for all PAH subclasses. **Supplementary Figure 4.** Power analysis. **Supplementary Figure 5.** Depth of coding sequence coverage for *FBLN2* and *PDGFD*. **Supplementary Figure 6.** Gene-based association analysis for APAH alone.

Additional file 2: Supplementary Table 1. Clinical characteristics and hemodynamic parameters of child- vs adult-onset PAH cases. **Supplementary Table 2.** Similar frequency of rare synonymous variants among cases and controls. **Supplementary Table 3.** Rare predicted deleterious *KDR* missense variants. **Supplementary Table 4.** Haplotype analysis of PAH cases with recurrent variants in new candidate genes. **Supplementary Table 5.** Burden of de novo variants in pediatric-onset IPAH. **Supplementary Table 6.** Rare de novo risk variants identified in pediatric-onset PAH. **Supplementary Table 7.** Clinical characteristics of pediatric PAH cases with rare de novo variants.

Acknowledgements

Samples and/or data from the National Biological Sample and Data Repository for PAH (aka PAH Biobank) were used in this study. We thank contributors, including the Pulmonary Hypertension Centers who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. We appreciate the contribution of the research coordinators across the clinical sites and Patricia Lanzano for coordinating the Columbia biorepository. Exome sequencing and genotyping data were generated by the Regeneron Genetics Center. We thank NIHR BioResource volunteers for their participation, and gratefully acknowledge NIHR BioResource centers, NHS Trusts, and staff for their contribution. We thank the National Institute for Health Research and NHS Blood and Transplant. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

We thank the research nurses and coordinators at the specialist pulmonary hypertension centers involved in this study. We acknowledge the support of the Imperial NIHR Clinical Research Facility, the Netherlands Cardiovascular Research Initiative, the Dutch Heart Foundation, Dutch Federation of University Medical Centres, the Netherlands Organization for Health Research and Development, and the Royal Netherlands Academy of Sciences. We thank all the patients and their families who contributed to this research and the Pulmonary Hypertension Association (UK) for their support.

PAH Biobank Enrolling Centers' Investigators: Russel Hirsch MD; R. James White MD, PhD; Marc Simon MD; David Badesch MD; Erika Rosenzweig MD; Charles Burger MD; Murali Chakinala MD; Thenappan Thenappan MD; Greg Elliott MD; Robert Simms MD; Harrison Farber, MD; Robert Frantz MD; Jean Elwing MD; Nicholas Hill MD; Dunbar Ivy MD; James Klinger MD; Steven Nathan MD; Ronald Oudiz MD; Ivan Robbins MD; Robert Schilz DO, PhD; Terry Fortin MD; Jeffrey Wilt MD; Delphine Yung MD; Eric Austin MD; Ferhaan Ahmad MD, PhD; Nitin Bhatt MD; Tim Lahm MD; Adaani Frost MD; Zeenat Safdar MD; Zia Rehman MD; Robert Walter MD; Fernando Torres MD; Sahil Bakshi DO; Stephen Archer MD; Rahul Argula MD; Christopher Barnett MD; Raymond Benza MD; Ankit Desai MD; Veeranna Maddipati MD.

NIHR BioResource – Rare Diseases and National Cohort Study of Idiopathic and Heritable PAH: Harm J. Bogaard, MD, PhD; Colin Church, PhD; Gery Coghlin, MD; Robin Condliffe, MD; Mélanie Eyries, PhD; Henning Gall, MD, PhD; Stefano Ghio, MD; Barbara Girerd, PhD; Simon Holden, PhD; Luke Howard, MD, PhD; Marc Humbert, MD, PhD; David G. Kiely, MD; Gabor Kovacs, MD; Jim Lordan, PhD; Rajiv D. Machado, PhD; Robert V. MacKenzie Ross, MB, BChir; Colm McCabe, PhD; Jennifer M. Martin, MSt; Shahin Moledina, MBChB; David Montani, MD, PhD; Horst Olschewski, MD; Christopher J. Penkett, PhD; Joanna Pepke-Zaba, PhD; Laura Price, PhD; Christopher J. Rhodes, PhD; Werner Seeger, MD; Florent Soubrier, MD, PhD; Laura Southgate, PhD; Jay Suntharalingam, MD; Andrew J. Swift, PhD; Mark R. Toshner, MD; Carmen M. Treacy, BSc; Anton Vonk Noordegraaf, MD; John Wharton, PhD; Jim Wild, PhD; Stephen John Wort, PhD.

Authors' contributions

WCN, WKC, MWP, and YS had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: WCN, WKC, SG, YS, NWM, and MWP. Acquisition, analysis, or interpretation of data: NZ, ES, MWP, CLW, JJH, XZ, YG, JK, DP, TT, KAL, JMM, CMT, ER, UK, AWC, CG-J, MRW, PAH Biobank, NIHR

BioResource – Rare Diseases Cohort Study of Idiopathic and Heritable PAH, YS, WKC, NWM, SG, and WCN. Drafting of the manuscript: NZ, CLW, MWP, YS, WKC, and WCN. Critical revision of the manuscript for important intellectual content: NZ, ES, MWP, CLW, JJH, XZ, YG, DP, TT, KAL, AWC, CG-J, MRW, PAH Biobank, Rare Diseases Cohort Study of Idiopathic and Heritable PAH, YS, WKC, NWM, SG, and WCN. Statistical analysis: NZ, ES, CLW, JJH, XZ, YS, SG, and WKC. Supervision: YS, WKC, NWM, SG, and WCN. The authors read and approved the final manuscript.

Funding

This study was funded in part by NIH grants R24HL105333 (WCN, MWP), RO1HL060056 (WKC), UO1HL125218 (WKC, ER), and RO1GM1200609 (YS). The UK National Cohort of Idiopathic and Heritable PAH is supported by the NIHR Cambridge Biomedical Research Centre (BRC) and Cambridge University Hospitals NHS Foundation Trust (CUH) (BRC 2012-2017), the British Heart Foundation (BHF) (SP/12/12/29836 and SP/18/10/33975), the BHF Cambridge Centre of Cardiovascular Research Excellence (CRE), the UK Medical Research Council (MR/K020919/1), and the Dinosaur Trust, BHF Programme grants to RCT (RG/08/006/25302) and NWM (RG/13/4/30107). NWM is a BHF Professor and NIHR Senior Investigator. AL is supported by a BHF Senior Basic Science Research Fellowship (FS/13/48/30453). All research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre.

Availability of data and materials

The datasets used and/or analyzed during the current study are available via contact with the senior authors. For PAH Biobank data, a Confidentiality Agreement with the collaborating Regeneron Sequencing Center grants to Dr. Nichols a nonexclusive, worldwide, irrevocable, perpetual, royalty-free sublicensable license to access and use the genomic data for any and all purposes. Therefore, while the PAH Biobank data are not uploaded to a publicly available database, direct access to the data are granted by the corresponding author on reasonable request who has full administrative access to all of the data. The data from the NIHRBR-RD study have been deposited in the European Genome-Phenome Archive [100]. Data from most of the affected participants in the US/UK combined cohort were included in previous publications from our group [7, 8, 17, 18, 27, 61, 100]. The following scripts are available: association test of rare variants with variable threshold of predicted functional scores [48] and principle component analysis of rare variants from Tabula Muris [58].

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Boards (IRBs) of the Cincinnati Children's Hospital Medical Center, the East of England Cambridge South national research ethics committee, Columbia University Irving Medical Center as well as the individual IRBs at each of the Enrolling Centers' institutions. All patients have signed consent forms which are on file at the individual Enrolling Centers. No protected health information (PHI) on any patients enrolled in the PAH Biobank, UK NIHR BioResource, or CUIMC has been forwarded to the data analyzing group. Only the individual Enrolling Centers have the ability to re-contact any of the patients enrolled in the study. All research using these patient samples conformed to the principles of the Helsinki Declaration.

Consent for publication

Written informed consent for publication was obtained from the patients/participants/legal guardians for minors at enrollment.

Competing interests

CG-J and the Regeneron Genetic Center collaborators are full-time employees of the Regeneron Genetics Center from Regeneron Pharmaceuticals Inc. and receive stock options as part of compensation. Johannes Karten is the full owner of 42Genetics BV. The remaining authors declare that they have no competing interests.

Author details

¹Department of Pediatrics, Columbia University Irving Medical Center, 1150 St. Nicholas Avenue, Room 620, New York, NY 10032, USA. ²Department of

Systems Biology, Columbia University, New York, NY, USA. ³Department of Medicine, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK. ⁴Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ⁵Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. ⁶42Genetics, Belfast, Ireland. ⁷NIHR BioResource for Translational Research, Cambridge Biomedical Campus, Cambridge, UK. ⁸Regeneron Pharmaceuticals, New York, NY, USA. ⁹Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK. ¹⁰Department of Medical and Molecular Genetics, King's College London, London, UK. ¹¹National Heart & Lung Institute, Imperial College London, London, UK. ¹²www.pahbiobank.org, Cincinnati, OH, USA. ¹³University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, UK. ¹⁴www.ipahcohort.com, Cambridge, UK. ¹⁵Addenbrooke's Hospital NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, UK. ¹⁶Royal Papworth Hospital NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, UK. ¹⁷Department of Biomedical Informatics, Columbia University, New York, NY, USA. ¹⁸Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK. ¹⁹Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA. ²⁰Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA.

Received: 18 June 2020 Accepted: 19 April 2021

Published online: 10 May 2021

References

- Vonk-Noordegraaf A, Haddad F, Chin KM, Forfia PR, Kawut SM, Lumens J, et al. Right heart adaptation to pulmonary arterial hypertension: physiology and pathobiology. *J Am Coll Cardiol*. 2013;62(25 Suppl):D22–33. <https://doi.org/10.1016/j.jacc.2013.10.027>.
- Ryan JJ, Archer SL. The right ventricle in pulmonary arterial hypertension: disorders of metabolism, angiogenesis and adrenergic signaling in right ventricular failure. *Circ Res*. 2014;115(1):176–88. <https://doi.org/10.1161/CIRCRESAHA.113.301129>.
- Humbert M, Guignabert C, Bonnet S, Dorfmüller P, Klinger JR, Nicolls MR, et al. Pathology and pathobiology of pulmonary hypertension: state of the art and research perspectives. *Eur Respir J*. 2019;53(1).
- Li L, Jick S, Breitenstein S, Hernandez G, Michel A, Vizcaya D. Pulmonary arterial hypertension in the USA: an epidemiological study in a large insured pediatric population. *Pulm Circ*. 2017;7(1):126–36. <https://doi.org/10.1086/690007>.
- Swinnen K, Quarck R, Godinas L, Belge C, Delcroix M. Learning from registries in pulmonary arterial hypertension: pitfalls and recommendations. *Eur Respir Rev*. 2019;28(154).
- Morrell NW, Aldred MA, Chung WK, Elliott CG, Nichols WC, Soubrier F, et al. Genetics and genomics of pulmonary arterial hypertension. *Eur Respir J*. 2019;53(1).
- Zhu N, Pauciuolo MW, Welch CL, Lutz KA, Coleman AW, Gonzaga-Jauregui C, et al. Novel risk genes and mechanisms implicated by exome sequencing of 2572 individuals with pulmonary arterial hypertension. *Genome Med*. 2019; 11(1):69. <https://doi.org/10.1186/s13073-019-0685-z>.
- Graf S, Haimel M, Bleda M, Hadinnapola C, Southgate L, Li W, et al. Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nat Commun*. 2018;9(1):1416. <https://doi.org/10.1038/s41467-018-03672-4>.
- Kabata H, Satoh T, Kataoka M, Tamura Y, Ono T, Yamamoto M, et al. Bone morphogenetic protein receptor type 2 mutations, clinical phenotypes and outcomes of Japanese patients with sporadic or familial pulmonary hypertension. *Respirology*. 2013;18(7):1076–82. <https://doi.org/10.1111/resp.12117>.
- Navas P, Tenorio J, Quezada CA, Barrios E, Gordo G, Arias P, et al. Molecular analysis of BMPR2, TBX4, and KCNK3 and genotype-phenotype correlations in Spanish patients and families with idiopathic and hereditary pulmonary arterial hypertension. *Rev Esp Cardiol (Engl Ed)*. 2016;69(11):1011–9. <https://doi.org/10.1016/j.recesp.2016.03.031>.
- Abou Hassan OK, Haidar W, Nemer G, Skouri H, Haddad F, BouAkl I. Clinical and genetic characteristics of pulmonary arterial hypertension in Lebanon. *BMC Med Genet*. 2018;19(1):89. <https://doi.org/10.1186/s12881-018-0608-7>.
- Evans JD, Girerd B, Montani D, Wang XJ, Galie N, Austin ED, et al. BMPR2 mutations and survival in pulmonary arterial hypertension: an individual participant data meta-analysis. *Lancet Respir Med*. 2016;4(2):129–37. [https://doi.org/10.1016/S2213-2600\(15\)00544-5](https://doi.org/10.1016/S2213-2600(15)00544-5).
- Yang H, Zeng Q, Ma Y, Liu B, Chen Q, Li W, et al. Genetic analyses in a cohort of 191 pulmonary arterial hypertension patients. *Respir Res*. 2018; 19(1):87. <https://doi.org/10.1186/s12931-018-0789-9>.
- Austin ED, Phillips JA, Cogan JD, Hamid R, Yu C, Stanton KC, et al. Truncating and missense BMPR2 mutations differentially affect the severity of heritable pulmonary arterial hypertension. *Respir Res*. 2009;10(1):87. <https://doi.org/10.1186/1465-9921-10-87>.
- Wang XJ, Lian TY, Jiang X, Liu SF, Li SQ, Jiang R, et al. Germline BMP9 mutation causes idiopathic pulmonary arterial hypertension. *Eur Respir J*. 2019;53(3).
- Kerstjens-Frederikse WS, Bongers EMHF, Roofthoof MTR, Leter EM, Douwes JM, Van Dijk A, et al. TBX4 mutations (small patella syndrome) are associated with childhood-onset pulmonary arterial hypertension. *J Med Genet*. 2013; 50(8):500–6. <https://doi.org/10.1136/jmedgenet-2012-101152>.
- Zhu N, Gonzaga-Jauregui C, Welch CL, Ma L, King AK, et al. Exome sequencing in children with pulmonary arterial hypertension demonstrates differences compared with adults. *Circ Genom Precis Med*. 2018;11(4): e001887.
- Zhu N, Welch CL, Wang J, Allen PM, Gonzaga-Jauregui C, Ma L, et al. Rare variants in SOX17 are associated with pulmonary arterial hypertension with congenital heart disease. *Genome Med*. 2018;10(1):56. <https://doi.org/10.1186/s13073-018-0566-x>.
- Welch CL, Chung WK. Genetics and genomics of pediatric pulmonary arterial hypertension. *Genes (Basel)*. 2020;11(10):1213–28.
- Sheeba CJ, Logan MP. The roles of T-Box genes in vertebrate limb development. *Curr Top Dev Biol*. 2017;122:355–81. <https://doi.org/10.1016/bs.ctdb.2016.08.009>.
- Krause A, Zacharias W, Camarata T, Linkhart B, Law E, Lischke A, et al. Tbx5 and Tbx4 transcription factors interact with a new chicken PDZ-LIM protein in limb and heart development. *Dev Biol*. 2004;273(1):106–20. <https://doi.org/10.1016/j.ydbio.2004.05.024>.
- Galambos C, Mullen MP, Shieh JT, Schwerk N, Kiehl MJ, Ullmann N, et al. Phenotype characterisation of TBX4 mutation and deletion carriers with neonatal and pediatric pulmonary hypertension. *Eur Respir J*. 2019;54(2): 1801965. <https://doi.org/10.1183/13993003.01965-2018>.
- Karolak JA, Vincent M, Deutsch G, Gambin T, Cogne B, Pichon O, et al. Complex Compound Inheritance of Lethal Lung Developmental Disorders due to Disruption of the TBX-FGF Pathway. *Am J Hum Genet*. 2019;104(2): 213–28. <https://doi.org/10.1016/j.ajhg.2018.12.010>.
- Best DH, Sumner KL, Austin ED, Chung WK, Brown LM, Borczuk AC, et al. EIF2AK4 mutations in pulmonary capillary hemangiomatosis. *Chest*. 2014; 145(2):231–6. <https://doi.org/10.1378/chest.13-2366>.
- Eyries M, Montani D, Girerd B, Perret C, Leroy A, Lonjou C, et al. EIF2AK4 mutations cause pulmonary veno-occlusive disease, a recessive form of pulmonary hypertension. *Nat Genet*. 2014;46(1):65–9. <https://doi.org/10.1038/ng.2844>.
- Ma L, Roman-Campos D, Austin ED, Eyries M, Sampson KS, Soubrier F, et al. A novel channelopathy in pulmonary arterial hypertension. *N Engl J Med*. 2013;369(4):351–61. <https://doi.org/10.1056/NEJMoa1211097>.
- Bohnen MS, Ma L, Zhu N, Qi H, McClenaghan C, Gonzaga-Jauregui C, et al. Loss-of-function ABCC8 mutations in pulmonary arterial hypertension. *Circ Genom Precis Med*. 2018;11(10):e002087. <https://doi.org/10.1161/CIRCGEN.118.002087>.
- Austin ED, Ma L, LeDuc C, Berman Rosenzweig E, Borczuk A, Phillips JA 3rd, et al. Whole exome sequencing to identify a novel gene (caveolin-1) associated with human pulmonary arterial hypertension. *Circ Cardiovasc Genet*. 2012;5(3):336–43. <https://doi.org/10.1161/CIRCGENETICS.111.961888>.
- Han B, Copeland CA, Kawano Y, Rosenzweig EB, Austin ED, Shahmirzadi L, et al. Characterization of a caveolin-1 mutation associated with both pulmonary arterial hypertension and congenital generalized lipodystrophy. *Traffic*. 2016;17(12):1297–312. <https://doi.org/10.1111/tra.12452>.
- Copeland CA, Han B, Tiwari A, Austin ED, Loyd JE, West JD, et al. A disease-associated frameshift mutation in caveolin-1 disrupts caveolae formation and function through introduction of a de novo ER retention signal. *Mol Biol Cell*. 2017;28(22):3095–111. <https://doi.org/10.1091/mbc.e17-06-0421>.
- Welch CL, Chung WK. Genetics and other omics in pediatric pulmonary arterial hypertension. *Chest*. 2020;155(5):1287–95.
- Simonneau G, Montani D, Celermajer DS, Denton CP, Gatzoulis MA, Krowka M, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J*. 2018.

33. Feliciano P, Zhou X, Astrovskaya I, Turner TN, Wang T, Brueggeman L, et al. Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. *NPJ Genom Med*. 2019;4(1):19. <https://doi.org/10.1038/s41525-019-0093-8>.
34. Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*. 2015;350(6265):1262–6. <https://doi.org/10.1126/science.aac9396>.
35. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research*. 2008;18(11):1851–8. <https://doi.org/10.1101/gr.078212.108>.
36. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491–8. <https://doi.org/10.1038/ng.806>.
37. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11 0 1–33.
38. Lin MF, Rodeh O, Penn J, Bai X, Reid JG, Krasheninina O, et al. GLnexus: joint variant calling for large cohort sequencing. *BioRxiv*. 2018:343970. <https://doi.org/10.1101/343970>.
39. Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol*. 2018;36(10):983–7. <https://doi.org/10.1038/nbt.4235>.
40. Supernat A, Vidarsson OV, Steen VM, Stokowy T. Comparison of three variant callers for human whole genome sequencing. *Sci Rep*. 2018;8(1):17851. <https://doi.org/10.1038/s41598-018-36177-7>.
41. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122. <https://doi.org/10.1186/s13059-016-0974-4>.
42. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. <https://doi.org/10.1093/nar/gkq603>.
43. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176(3):535–48 e24. <https://doi.org/10.1016/j.cell.2018.12.015>.
44. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877–85. <https://doi.org/10.1016/j.ajhg.2016.08.016>.
45. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–92. <https://doi.org/10.1093/bib/bbs017>.
46. Pedersen BS, Quinlan AR. Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with peddy. *Am J Hum Genet*. 2017;100(3):406–13. <https://doi.org/10.1016/j.ajhg.2017.01.017>.
47. Tom JA, Reeder J, Forrest WF, Graham RR, Hunkapiller J, Behrens TW, et al. Identifying and mitigating batch effects in whole genome sequencing data. *BMC Bioinformatics*. 2017;18(1):351. <https://doi.org/10.1186/s12859-017-1756-z>.
48. Zhu NaS Y. Association test of rare variants with variable threshold of predicted functional scores. *GitHub*; 2019. <https://github.com/ShenLab/VariationThresholdTest/doi:10.5281/zenodo.4659344>.
49. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867–73. <https://doi.org/10.1093/bioinformatics/btq559>.
50. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011;9(2):179–81. <https://doi.org/10.1038/nmeth.1785>.
51. International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449(7164):851–61. <https://doi.org/10.1038/nature06258>.
52. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet*. 2014;46(9):944–50. <https://doi.org/10.1038/ng.3050>.
53. Ware JS, Samocha KE, Homsy J, Daly MJ. Interpreting de novo variation in human disease using denovolyzeR. *Curr Protoc Hum Genet*. 2015;87:7 25 1–15 editorial board, Jonathan L Haines [et al].
54. Kuntal BK, Aparoy P, Reddanna P. EasyModeller: a graphical interface to MODELLER. *BMC Res Notes*. 2010;3(1):226. <https://doi.org/10.1186/1756-0500-3-226>.
55. Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, et al. Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature*. 2012;490(7421):556–60. <https://doi.org/10.1038/nature11503>.
56. Zhang QC, Petrey D, Garzon JJ, Deng L, Honig B. PrePPI: a structure-informed database of protein-protein interactions. *Nucleic Acids Res*. 2013;41(Database issue):D828–33. <https://doi.org/10.1093/nar/gks1231>.
57. Tabula Muris C, Overall C, Logistical C, Organ C, Processing, Library P, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. 2018;562(7727):367–72. <https://doi.org/10.1038/s41586-018-0590-4>.
58. Zhu NaS Y. Principle component analysis of rare variants from Tabula Muris/ The Tabula Muris Consortium et al. 2018. *GitHub*. <https://doi.org/10.5281/zenodo.4660098>.
59. Batton KA, Austin CO, Bruno KA, Burger CD, Shapiro BP, Fairweather D. Sex differences in pulmonary arterial hypertension: role of infection and autoimmunity in the pathogenesis of disease. *Biol Sex Differ*. 2018;9(1):15. <https://doi.org/10.1186/s13293-018-0176-8>.
60. Eyries M, Montani D, Girerd B, Favrolt N, Riou M, Favre L, et al. Familial pulmonary arterial hypertension by KDR heterozygous loss of function. *Eur Respir J*. 2020;55(4).
61. Swietlik EM, Greene D, Zhu N, Megy K, Cogliano M, Rajaram S, et al. Reduced transfer coefficient of carbon monoxide in pulmonary arterial hypertension implicates rare protein-truncating variants in KDR. *BioRxiv*. 2019.
62. Cao C, Wang S, Cui T, Su XC, Chou JJ. Ion and inhibitor binding of the double-ring ion selectivity filter of the mitochondrial calcium uniporter. *Proc Natl Acad Sci U S A*. 2017;114(14):E2846–E51. <https://doi.org/10.1073/pnas.1620316114>.
63. Gaboriaud C, Gregory-Pauron L, Teillet F, Thielens NM, Bally I, Arlaud GJ. Structure and properties of the Ca(2+)-binding CUB domain, a widespread ligand-recognition unit involved in major biological functions. *Biochem J*. 2011;439(2):185–93. <https://doi.org/10.1042/BJ20111027>.
64. Lang IM, Palazzini M. The burden of comorbidities in pulmonary arterial hypertension. *Eur Heart J Suppl*. 2019;21(Suppl K):K21–K8.
65. Badesch DB, Raskob GE, Elliott CG, Krichman AM, Farber HW, Frost AE, et al. Pulmonary arterial hypertension: baseline characteristics from the REVEAL Registry. *Chest*. 2010;137(2):376–87. <https://doi.org/10.1378/chest.09-1140>.
66. Zhang C, Shen Y. A cell type-specific expression signature predicts haploinsufficient autism-susceptibility genes. *Hum Mutat*. 2017;38(2):204–15. <https://doi.org/10.1002/humu.23147>.
67. Galvis LA, Holik AZ, Short KM, Pasquet J, Lun AT, Blewitt ME, et al. Repression of Igf1 expression by Ezh2 prevents basal cell differentiation in the developing lung. *Development*. 2015;142(8):1458–69. <https://doi.org/10.1242/dev.122077>.
68. Tartaglia M, Martinelli S, Stella L, Bocchinfuso G, Flex E, Cordeddu V, et al. Diversity and functional consequences of germline and somatic PTPN11 mutations in human disease. *Am J Hum Genet*. 2006;78(2):279–90. <https://doi.org/10.1086/499925>.
69. Razzaque MA, Nishizawa T, Komoike Y, Yagi H, Furutani M, Amo R, et al. Germline gain-of-function mutations in RAF1 cause Noonan syndrome. *Nat Genet*. 2007;39(8):1013–7. <https://doi.org/10.1038/ng2078>.
70. Page DJ, Miossec MJ, Williams SG, Monaghan RM, Fotiou E, Cordell HJ, et al. Whole exome sequencing reveals the major genetic contributors to nonsyndromic tetralogy of fallot. *Circ Res*. 2019;124(4):553–63. <https://doi.org/10.1161/CIRCRESAHA.118.313250>.
71. Fassler R, Sasaki T, Timpl R, Chu ML, Werner S. Differential regulation of fibulin, tenascin-C, and nidogen expression during wound healing of normal and glucocorticoid-treated mice. *Exp Cell Res*. 1996;222(1):111–6. <https://doi.org/10.1006/excr.1996.0014>.
72. Tsuda T, Wang H, Timpl R, Chu ML. Fibulin-2 expression marks transformed mesenchymal cells in developing cardiac valves, aortic arch vessels, and coronary vessels. *Dev Dyn*. 2001;222(1):89–100. <https://doi.org/10.1002/dvdy.1172>.
73. Sicot FX, Tsuda T, Markova D, Klement JF, Arita M, Zhang RZ, et al. Fibulin-2 is dispensable for mouse development and elastic fiber formation. *Mol Cell Biol*. 2008;28(3):1061–7. <https://doi.org/10.1128/MCB.01876-07>.
74. Olijnyk D, Ibrahim AM, Ferrier RK, Tsuda T, Chu ML, Gusterson BA, et al. Fibulin-2 is involved in early extracellular matrix development of the

- outgrowing mouse mammary epithelium. *Cell Mol Life Sci.* 2014;71(19):3811–28. <https://doi.org/10.1007/s00018-014-1577-4>.
75. Khan SA, Dong H, Joyce J, Sasaki T, Chu ML, Tsuda T. Fibulin-2 is essential for angiotensin II-induced myocardial fibrosis mediated by transforming growth factor (TGF)-beta. *Lab Invest.* 2016;96(7):773–83. <https://doi.org/10.1038/labinvest.2016.52>.
 76. Ackerman C, Locke AE, Feingold E, Reshey B, Espana K, Thusberg J, et al. An excess of deleterious variants in VEGF-A pathway genes in Down-syndrome-associated atrioventricular septal defects. *Am J Hum Genet.* 2012;91(4):646–59. <https://doi.org/10.1016/j.ajhg.2012.08.017>.
 77. van't Hof FNG, Lai D, van Setten J, Bots ML, Vaartjes I, Broderick J, et al. Exome-chip association analysis of intracranial aneurysms. *Neurology.* 2020;94(5):e481–e8.
 78. Kim AD, Lake BB, Chen S, Wu Y, Guo J, Parvez RK, et al. Cellular recruitment by podocyte-derived pro-migratory factors in assembly of the human renal filter. *iScience.* 2019;20:402–14. <https://doi.org/10.1016/j.isci.2019.09.029>.
 79. Torregrosa-Carrion R, Luna-Zurita L, Garcia-Marques F, D'Amato G, Pineiro-Sabaris R, Bonzon-Kulichenko E, et al. NOTCH activation promotes valve formation by regulating the endocardial secretome. *Mol Cell Proteomics.* 2019;18(9):1782–95. <https://doi.org/10.1074/mcp.RA119.001492>.
 80. Zhang HY, Chu ML, Pan TC, Sasaki T, Timpl R, Ekblom P. Extracellular matrix protein fibulin-2 is expressed in the embryonic endocardial cushion tissue and is a prominent component of valves in adult heart. *Dev Biol.* 1995;167(1):18–26. <https://doi.org/10.1006/dbio.1995.1003>.
 81. Lee C, Li X. Platelet-derived growth factor-C and -D in the cardiovascular system and diseases. *Mol Aspects Med.* 2018;62:12–21. <https://doi.org/10.1016/j.mam.2017.09.005>.
 82. Gladh H, Folestad EB, Muhl L, Ehnman M, Tannenberg P, Lawrence AL, et al. Mice lacking platelet-derived growth factor d display a mild vascular phenotype. *PLoS One.* 2016;11(3):e0152276. <https://doi.org/10.1371/journal.pone.0152276>.
 83. Ponten A, Folestad EB, Pietras K, Eriksson U. Platelet-derived growth factor D induces cardiac fibrosis and proliferation of vascular smooth muscle cells in heart-specific transgenic mice. *Circ Res.* 2005;97(10):1036–45. <https://doi.org/10.1161/01.RES.0000190590.31545.d4>.
 84. Folestad E, Kunath A, Wagsater D. PDGF-C and PDGF-D signaling in vascular diseases and animal models. *Mol Aspects Med.* 2018;62:1–11. <https://doi.org/10.1016/j.mam.2018.01.005>.
 85. Wu Q, Hou X, Xia J, Qian X, Miele L, Sarkar FH, et al. Emerging roles of PDGF-D in EMT progression during tumorigenesis. *Cancer Treat Rev.* 2013;39(6):640–6. <https://doi.org/10.1016/j.ctrv.2012.11.006>.
 86. Huang W, Kim HR. Dynamic regulation of platelet-derived growth factor D (PDGF-D) activity and extracellular spatial distribution by matriptase-mediated proteolysis. *J Biol Chem.* 2015;290(14):9162–70. <https://doi.org/10.1074/jbc.M114.610865>.
 87. Papadopoulos N, Lennartsson J. The PDGF/PDGF-R pathway as a drug target. *Mol Aspects Med.* 2018;62:75–88. <https://doi.org/10.1016/j.mam.2017.11.007>.
 88. Burke MJ, Walmsley R, Munsey TS, Smith AJ. Receptor tyrosine kinase inhibitors cause dysfunction in adult rat cardiac fibroblasts in vitro. *Toxicol In Vitro.* 2019;58:178–86. <https://doi.org/10.1016/j.tiv.2019.03.026>.
 89. Reyes-Palomares A, Gu M, Grubert F, Berest I, Sa S, Kasowski M, et al. Remodeling of active endothelial enhancers is associated with aberrant gene-regulatory networks in pulmonary arterial hypertension. *Nat Commun.* 2020;11(1):1673. <https://doi.org/10.1038/s41467-020-15463-x>.
 90. Qi H, Yu L, Zhou X, Wynn J, Zhao H, Guo Y, et al. De novo variants in congenital diaphragmatic hernia identify MYRF as a new syndrome and reveal genetic overlaps with other developmental disorders. *PLoS Genet.* 2018;14(12):e1007822. <https://doi.org/10.1371/journal.pgen.1007822>.
 91. Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet.* 2012;13(8):565–75. <https://doi.org/10.1038/nrg3241>.
 92. D'Alonzo GE, Barst RJ, Ayres SM, Bergofsky EH, Brundage BH, Detre KM, et al. Survival in patients with primary pulmonary hypertension. Results from a national prospective registry. *Ann Intern Med.* 1991;115(5):343–9. <https://doi.org/10.7326/0003-4819-115-5-343>.
 93. Barst RJ, McGoon MD, Elliott CG, Foreman AJ, Miller DP, Ivy DD. Survival in childhood pulmonary arterial hypertension: insights from the registry to evaluate early and long-term pulmonary arterial hypertension disease management. *Circulation.* 2012;125(1):113–22. <https://doi.org/10.1161/CIRCULATIONAHA.111.026591>.
 94. Ivy D. Pulmonary Hypertension in Children. *Cardiol Clin.* 2016;34(3):451–72. <https://doi.org/10.1016/j.ccl.2016.04.005>.
 95. Rosenzweig EB, Abman SH, Adatia I, Beghetti M, Bonnet D, Haworth S, et al. Paediatric pulmonary arterial hypertension: updates on definition, classification, diagnostics and management. *Eur Respir J.* 2019;53(1).
 96. Steurer MA, Baer RJ, Oltman S, Ryckman KK, Feuer SK, Rogers E, et al. Morbidity of persistent pulmonary hypertension of the newborn in the first year of life. *J Pediatr.* 2019;213:58–65 e4. <https://doi.org/10.1016/j.jpeds.2019.06.053>.
 97. Cheng Y, Yu M, Xu J, He M, Wang H, Kong H, et al. Inhibition of Shp2 ameliorates monocrotaline-induced pulmonary arterial hypertension in rats. *BMC Pulm Med.* 2018;18(1):130. <https://doi.org/10.1186/s12890-018-0700-y>.
 98. Courboulin A, Paulin R, Giguere NJ, Saksouk N, Perreault T, Meloche J, et al. Role for miR-204 in human pulmonary arterial hypertension. *J Exp Med.* 2011;208(3):535–48. <https://doi.org/10.1084/jem.20101812>.
 99. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011;478(7367):103–9. <https://doi.org/10.1038/nature10405>.
 100. Turro E, Astle WJ, Megy K, Graf S, Greene D, Shamardina O, et al. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature.* 2020;583(7814):96–102. <https://doi.org/10.1038/s41586-020-2434-2>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

