Reduced transfer coefficient of carbon monoxide in pulmonary arterial hypertension implicates rare protein-truncating variants in *KDR*

Genotype-phenotype inference reveals novel PAH risk genes

Emilia M. Swietlik, MD[1], Daniel Greene, PhD[2,3], Na Zhu, PhD[4,5], Karyn Megy, PhD[2,3], Marcella Cogliano, MSc[6], Smitha Rajaram, MD[7], Divya Pandya, MSc[1], Tobias Tilly, MSc[1], Katie A. Lutz, BS[8], Carrie C. L. Welch, PhD[4], Michael W. Pauciulo, BS, MBA[8,9], Laura Southgate, PhD[10], Jennifer M. Martin, MSt[3], Carmen M. Treacy, BSc[1], Christopher J. Penkett, PhD[2,3], Jonathan C. Stephens, BSc[2,3], Harm J. Bogaard, MD, PhD[11], Colin Church, PhD[12], Gerry Coghlan, MD[13], Anna W. Coleman, MS[8], Robin Condliffe, MD[14], Mélanie Eyries, PhD[15], Henning Gall, MD, PhD[16], Stefano Ghio, MD[17], Barbara Girerd, PhD[18], Simon Holden, PhD[19], Luke Howard, MD, PhD[20], Marc Humbert, MD, PhD[18], David G. Kiely, MD[14], Gabor Kovacs, MD[21,22], Jim Lordan, PhD[23], Rajiv D. Machado, PhD[10], Robert V. MacKenzie Ross, MB, BChir[24], Shahin Moledina, MBChB[25], David Montani, MD, PhD[18], Horst Olschewski, MD[21,22], Joanna Pepke-Zaba, PhD[26], Christopher J. Rhodes, PhD[20], Werner Seeger, MD[16], Florent Soubrier, MD, PhD[15], Jay Suntharalingam, MD[24], Mark R. Toshner, MD[1,26], Anton Vonk Noordegraaf, MD[11], John Wharton, PhD[20], Jim Wild, PhD[6], Stephen John Wort, PhD[20,27], NIHR BioResource for

Translational Research - Rare Diseases[28], National Cohort Study of Idiopathic and Heritable PAH[29], PAH Biobank Enrolling Centers' Investigators[30], Allan Lawrie, PhD[6], Martin R. Wilkins, MD[20], Richard C. Trembath, FRCP[31], Yufeng Shen, PhD[5,32], Wendy K. Chung, MD[33], Andrew J. Swift, PhD[6], William C. Nichols, PhD[8,9], Nicholas W. Morrell, MD[1,3,19,26,$], Stefan Gräf, PhD[1,2,3,$]

[$] these authors jointly supervised this work

Affiliations

[1]Department of Medicine, University of Cambridge, Cambridge Biomedical Campus, Cambridge, United Kingdom.

[2]Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, United Kingdom.

[3]NIHR BioResource for Translational Research, Cambridge Biomedical Campus, Cambridge, United Kingdom.

[4]Department of Pediatrics, Columbia University, New York, United States.

[5]Department of Systems Biology, Columbia University, New York, United States.

[6]Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, United Kingdom.

[7]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, United Kingdom.

[8]Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, United States.

[9]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, United States.

[10]Molecular and Clinical Sciences Research Institute, St George's, University of London, London, United Kingdom.

[11]Department of Clinical Genetics, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, Amsterdam, The Netherlands.

[12]Golden Jubilee National Hospital, Glasgow, United Kingdom.

[13]Royal Free Hospital, London, United Kingdom.

[14]Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital, Sheffield, United Kingdom.

[15]Département de génétique, hôpital Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, and UMR_S 1166-ICAN, INSERM, UPMC Sorbonne Universités, Paris, France.

[16]University of Giessen and Marburg Lung Center (UGMLC), member of the German Center for Lung Research (DZL) and of the Excellence Cluster Cardio-Pulmonary Institute (CPI), Giessen, Germany.

[17]Fondazione IRCCS Policlinico San Matteo, Pavia, Italy.

[18]Université Paris-Sud, Faculté de Médecine, Université Paris-Saclay; AP-HP, Service de Pneumologie, Centre de référence de l'hypertension pulmonaire; INSERM UMR_S 999, Hôpital Bicêtre, Le Kremlin-Bicêtre, Paris, France.

[19]Addenbrooke's Hospital NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom.

[20]National Heart & Lung Institute, Imperial College London, London, United Kingdom.

[21]Ludwig Boltzmann Institute for Lung Vascular Research, Graz, Austria.

[22]Medical University of Graz, Graz, Austria.

[23]Freeman Hospital, Newcastle upon Tyne, United Kingdom.

[24]Royal United Hospitals Bath NHS Foundation Trust, Bath, United Kingdom.

[25]Great Ormond Street Hospital, London, United Kingdom.

[26]Royal Papworth Hospital NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom.

[27]Royal Brompton & Harefield NHS Foundation Trust, London, United Kingdom.

[28]University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom.

[29]www.ipahcohort.com, Cambridge, United Kingdom.

[30]www.pahbiobank.org, Cincinnati, United States.

[31]Department of Medical and Molecular Genetics, King's College London, London, United Kingdom.

[32]Department of Biomedical Informatics, Columbia University, New York, United States.

[33]Columbia University Medical Center, New York, United States.

Corresponding authors:

Dr Stefan Gräf, PhD & Professor Nicholas W. Morrell, M.D.

Department of Medicine, University of Cambridge and Cambridge University Hospitals, Box 157, Level 5, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, United Kingdom

Email: sg550@cam.ac.uk / nwm23@cam.ac.uk, tel: (+44) 1223 588036 / (+44) 1223 331666

Total word count: 4537

Abstract

Background

To date, approximately 25% of patients with pulmonary arterial hypertension (PAH) have been found to harbor rare mutations in disease-causing genes. Given the small number of patients affected by mutations in most PAH genes, the identification of the missing heritability in PAH is challenging. We hypothesized that integrating deep phenotyping data with whole-genome sequencing data will reveal additional disease variants that are extremely rare and have a unique phenotypic signature.

Methods

We analyzed whole-genome sequencing data from 13,037 participants enrolled in the NIHR BioResource - Rare Diseases (NIHRBR-RD) study, of which 1148 were recruited to the PAH domain. To test for genetic associations between genes and selected phenotypes of pulmonary hypertension (PH), we used the Bayesian rare-variant association method BeviMed. We defined the groups for comparison by assigning labels ('tags') inferred from the current diagnostic classification of PAH, including stratification by age at diagnosis and the transfer coefficient of carbon monoxide (KCO).

Results

Protein truncating variants (PTV) in *KDR* were strongly associated with the lower KCO tertile (posterior probability (PP)=0.989) and the higher age tertile (PP=0.912) groups. On computed tomographic imaging of the lungs, a range of parenchymal abnormalities were observed in the

patients harboring PTV in *KDR*. KCO stratification also highlighted an association between Isocitrate Dehydrogenase (NAD(+)) 3 Non-Catalytic Subunit Gamma (*IDH3G*) and moderately reduced KCO in patients with pulmonary hypertension (PP=0.649). The US PAH Biobank was used to independently validate these findings and identified four additional PAH cases with PTV in *KDR* and three in *IDH3G*. We confirmed associations between previously established genes and PAH.

Conclusions

PTVs in *KDR*, the gene encoding vascular endothelial growth factor receptor 2 (VEGFR2), are significantly associated with two specific phenotypes of PAH, reduced KCO and later age of onset, highlighting a role for VEGF signaling in the pathogenesis of human PAH. We also report *IDH3G* as a new PAH risk gene. Moreover, we demonstrate that the use of deep clinical phenotyping data advances the identification of novel rare PAH risk variants.

Clinical Perspective

1) What is new?

- The NIHR BioResource PAH cohort is the largest collection of patients with idiopathic and hereditary PAH worldwide with whole-genome sequencing data.
- This is the first study in the field of PAH which combines deep clinical phenotyping with whole-genome sequencing data to discover new genotype-phenotype associations using a Bayesian model comparison method.
- This "smart" experimental design allowed us to discover two new PAH risk genes namely *KDR*, which encodes for vascular endothelial growth factor receptor 2 (VEGFR2) and *IDH3G* (Isocitrate Dehydrogenase (NAD(+)) 3 Non-Catalytic Subunit Gamma). These genes further support central roles for the endothelium and alterations in mitochondrial respiration in the pathobiology of PAH.

2) What are the clinical implications?

- The curation of deep clinical phenotyping data is important in rare diseases such as PAH, where this information could aid the identification of genetic drivers of disease.
- KCO and age stratification enabled patient clustering into subgroups with shared pathobiology and increased power to identify rare causal variants. KCO should be measured in IPAH patients to help decision making about genetic testing.
- Patients harboring PTV in *KDR* show a range of parenchymal changes on chest CT scans, with implications for our understanding of endothelial-epithelial interactions.

Introduction

Pulmonary arterial hypertension (PAH) is a rare condition characterized by pulmonary vascular constriction and obliteration, causing elevation of pulmonary vascular resistance and ultimately, right ventricular failure. Multiple concepts have been proposed to explain the mechanisms leading to pulmonary vessel remodeling[1]. More recently, hallmarks of cancer, such as aberrant angiogenesis[2], metabolic reprogramming[3] and resistance to apoptosis[4], have been identified. A breakthrough in our understanding of PAH pathobiology was the discovery of heterozygous germline mutations in the gene encoding bone morphogenetic protein type 2 receptor (*BMPR2*)[5]. It is now established that *BMPR2* mutations are responsible for over 70% of familial PAH (FPAH) cases and 15-20% of idiopathic PAH (IPAH) cases. The penetrance of *BMPR2* mutations is incomplete, so not all carriers develop the disease[6]. A smaller proportion (up to 10%) of PAH are caused by mutations in activin-like kinase 1 (*ACVRL1*)[7], endoglin (*ENG*)[8], SMAD family member 9 (*SMAD9*)[9], caveolin-1 (*CAV1*), involved in colocalization of BMP receptors[10], and the potassium channel, *KCNK3*, responsible for membrane potential and vascular tone[11]. Using burden tests, we recently identified rare pathogenic variants in growth differentiation factor 2 (*GDF2)*, which encodes BMP9, a major ligand for BMPR2, as well as rare variants in ATPase 13A3 (*ATP13A3*), aquaporin 1 (*AQP1*) and SRY-box 17 (*SOX17*), and reported a list of additional putative genes potentially contributing to the pathobiology of PAH[12]. Together, the established genes explain approximately 25% of cases with IPAH, allowing their reclassification as heritable PAH (HPAH) cases. To identify additional genes harboring potentially causal rare variants in IPAH cases, we increased the cohort size and deployed a recently described Bayesian methodology (BeviMed)[13] that incorporates phenotypic data to increase the power to detect rare risk variants.

Methods

Data availability

The whole-genome sequencing and genotype data of the National Institute for Health Research BioResource - Rare Diseases (NIHRBR-RD) study has been deposited at the European Genome-Phenome Archive[14]. The data and resources of the US PAH Biobank (PAHBB) and the Columbia University Medical Center (CUMC) are made available via an application process[15].

Study design, ethics, and subject recruitment

The NIHRBR-RD, the rare disease pilot study for Genomics England Ltd. 100,000 Genomes Project, was established to identify genetic causes, improve rates of molecular diagnosis and develop new treatments for rare diseases through whole-genome sequencing and deep phenotyping[14]. Of the 18 domains, 15 were defined either as a single rare disease or a group of rare disorders (Table S1). The PAH domain comprised 1,148 subjects, including individuals diagnosed with either IPAH or HPAH, pulmonary veno-occlusive disease (PVOD) or pulmonary capillary haemangiomatosis (PCH) and a small number of healthy relatives. Adult and pediatric onset cases were eligible, as well as incident and prevalent cases. Recruitment of PAH cases was carried out across the nine pulmonary hypertension (PH) specialist centers in the UK and by international collaborators at the Université Paris-Saclay and Sorbonne Université (France), University of Giessen and Marburg (Germany), and hospitals in Graz (Austria), Pavia (Italy) and Amsterdam (The Netherlands). Patients recruited to the NIHRBR-RD study provided written, informed consent for genetic analysis and clinical data capture (REC REF: 13/EE/0325); patients

recruited by European collaborators consented to genetic testing and clinical data collection locally.

Patients with rare diseases recruited to domains other than PAH were used as non-PAH controls in the genetic analysis (Table 1).

For validation, we used the US PAHBB cohort comprising exome sequencing data from 2572 subjects diagnosed with group 1 PAH[15] and a biobank of 440 PAH patients established at CUMC composed of 29 FPAH, 195 IPAH and 216 individuals with forms of PAH associated with other conditions (APAH)[16].

Phenotyping of patients

Clinical phenotyping and case-control cohort using phenotypic 'tags'

Pseudonymized results of routinely performed clinical tests were stored across twenty-one electronic Case Report Forms (eCRFs) using the *OpenClinica* data capture system (Table S2). All cases were diagnosed between January 2002 to December 2017, and the diagnostic classification was made according to international guidelines using a multidisciplinary assessment that included echocardiography, comprehensive blood testing, pulmonary function testing, overnight oximetry, isotope perfusion scanning, high-resolution computed tomography, and right heart catheterization. To aid data analysis and improve data quality, a number of quality assurance procedures were introduced (see Supplemental Material). Diagnosis in all patients was verified based on hemodynamic criteria, reported comorbidities and results of pulmonary function tests, heart and lung imaging and autoantibody screen. Cases in which the diagnosis was questionable were reported back to recruiting centers for verification. Appropriate diagnostic and phenotypic tags were assigned to all recruited patients to be used in the

subsequent case-control analysis (Figure S2). The full set of tags, with corresponding numbers of cases, controls, and excluded relatives, can be found in Table 1.

Analysis of computerized tomography (CT) scans

Diagnostic chest CT scans were performed in 613 study participants and the reports were subsequently transcribed to study eCRFs. Of the 613 scans, 294 were available for more detailed analysis. The scans were anonymized and transferred to the University of Sheffield, Sheffield, UK. Here, they were reviewed by two independent cardiothoracic radiologists with expertise in pulmonary hypertension (AS and SR), who were blinded to the underlying diagnosis, mutation and smoking status. For consistency and reproducibility, all measurements were reported on a customized proforma (for further details see Supplemental Material, Table S3 and S10).

Whole-genome sequencing, short read alignment, and variant calling

Samples were received as either DNA extracted from whole blood or as whole blood EDTA samples that were extracted at the central DNA extraction and QC laboratory in Cambridge (UK). Following testing for adequate DNA concentration, DNA degradation, and purity, next-generation paired-end whole-genome sequencing was performed on cases and controls using Illumina HiSeq2500 and HiSeq X (Illumina Inc, San Diego, USA) generating three batches with different read lengths: 100bp (377 samples), 125bp (3,154 samples) and 150bp (9,656 samples). Reads were aligned against the Genome Reference Consortium human genome build 37 (GRCh37, https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/) using the Illumina Isaac Aligner version SAAC00776.15.01.27[17] and variants were called using the Illumina Starling

software version 2.1.4.2 (https://support.illumina.com/help/BS_App_TS_Amplicon_OLH_15055858/Content/Source/Informatics/Apps/IsaacVariantCaller_appENR.htm). The variants were then left-aligned, normalized with *bcftools* and loaded into our Hbase database to produce multi-sample variant calls to undertake the genetic association studies[14].

Genetic association between rare variants and selected diagnostic and phenotypic tags

A schematic of the analysis pipeline is depicted in Figure 1A. We hypothesized that groups of patients who shared a particular feature would also share a similar genetic etiology. In particular, we focused on the current diagnostic classification of pulmonary hypertension and stratification by age at diagnosis and KCO (% predicted), to define a set of phenotypic tags (Table 1). We defined cases as individuals carrying a particular tag whereas the individuals from the non-PAH domains served as controls (Figure 1, Table S1). Variants were extracted from each gene as previously described[14] including a $PMAF_x$ (for a given variant, the probability that the minor allele count is at least the observed minor allele count, given that MAF=1/X) <0.05 with x=1,000 for the recessive and x=10,000 for the dominant association model, and a CADD Phred score ≥10. The analysis was restricted to canonica[14]l transcripts as annotated by Ensembl. The Bayesian model comparison method BeviMed[13] was applied to the extracted rare variants from a set of unrelated individuals to estimate posterior probabilities of gene-tag associations under dominant and recessive modes of inheritance. Importantly, since we selected rare variants that do not correlate strongly with coarse population structure, we did not correct for ethnicity. The association testing was only performed on unrelated individuals (Table 1). Patients with rare deleterious variants in

previously established PAH disease genes (*BMPR2*, *ACVRL1*, *ENG*, *CAV1*, *SMAD1*, *SMAD4*, *SMAD9*, *KCNK3*, *EIF2AK4*, *TBX4*, *AQP1*, *ATP13A3*, *GDF2*, *SOX17*) that were deemed disease-causing by a genetic multidisciplinary team according to the ACMG Standards and Guidelines[18], were excluded from the association testing for other genes to minimize false-positive associations. To increase power in scenarios where only variants of particular consequence types were associated with the disease risk, association models were fitted to different subsets of variants according to the consequences provided by Ensembl (https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html): the High category, comprised only variants of "high" impact, including PTVs and large deletions; the Moderate category contains variants of impact "moderate", including missense variants or consequence "non_coding_transcript_exon_variant"; the combined category Moderate and High, combining the respective consequence types. The prior probability of association across all association models was set to 0.001. Our choice of prior was based on the assumption that approximately 30 genes with rare variants of moderate and high impact might be involved in the pathogenesis of PAH out of the 32,606 protein-coding and non-coding genes (defined by the selected gene biotypes provided by Ensembl, see Supplemental Material).

Descriptive statistics

Statistical analysis and data visualization were performed in R (www.r-project.org). Summary statistics are shown as mean (±SD) or median [IQR] according to data distribution (normality testing was performed with the Shapiro-Wilk test and QQ plots). The number of available data points is reported in tables. Comparisons between the categorical variables were performed using Fisher's exact and Chi-square test, comparisons between continuous non-normally distributed variables were performed with the Mann-Whitney test (for two groups) or the Kruskal-

Wallis test (three and more groups). Adjustment for multiple comparisons was performed when appropriate. The Kaplan-Meier method was used to visualize survival curves; the log-rank test was used to compare survival between two or more groups; Cox proportional hazards regression was used to examine the effect of variables on survival. Testing for the proportional hazards assumption, influential observations and non-linearity were done, and the assumptions were met. To measure the magnitude of agreement between CT scan readers, 22 randomly selected tests were assessed by both radiologists. For categorical variables, the weighted (ordinal data) and unweighted (for non-ordinal data) Cohen's Kappa for two readers were calculated and for continuous variables, the intraclass correlation coefficient (ICC) was computed with the R package ("irr").

Results

Characterization of study cohorts and tag definition

Whole-genome sequencing was performed in 13,037 participants of the NIHRBR-RD study, of which 1,148 were recruited to the PAH domain. The PAH domain included 23 unaffected parents and 3 cases with an unknown phenotype, which were removed from the analysis (Table S1, Figure 1B). Of the remaining 1,122 participants, 972 (86.6%) had a clinical diagnosis of IPAH, 73 (6.5%) of HPAH, and 20 (1.8%) were diagnosed with PVOD/PCH. Diagnosis verification, utilizing the collected clinical information, revealed that 57 participants (5%) had a diagnosis other than IPAH, HPAH or PVOD/PCH. These cases were subsequently relabeled and moved to the respective tag group for analysis (see Table S4, Table 1). The PAH cohort's population structure was comparable to previously studied European PAH populations, with a median age at diagnosis of 49[35;63] years, and female predominance of 68% (760 individuals). Europeans

constituted 84% of the study cohort. (Table S4). Overall survival in the studied population was 97% at one year, 91% at three years and 84% at five years. As expected, there was a significant difference in survival between prevalent and incident cases. In prevalent cases, survival at one, three, and five years was 98%, 93%, 87%, whereas in incident cases it was 97%, 84%, 72%. The transfer coefficient of carbon monoxide (KCO) measured at diagnosis was available for 644 patients (57%) (see Supplemental Material, Table S5, Figure S1). Median KCO in the entire studied population was 71[52;86]% predicted (Figure S2). Cases in the lower tertile or below the KCO threshold of 50% predicted were more commonly male, older at diagnosis, had a current or past history of cigarette smoking and an increased number of cardiorespiratory comorbidities (Table S6, Table S7). Survival in these groups was significantly worse than in those with preserved or mildly reduced KCO (Figure S3 A and B). After adjusting for confounding factors (age, sex, comorbidities, smoking status and whether the case was prevalent or incident), KCO remained an independent predictor of survival (Table S8).

Age at diagnosis was calculated as age at the time of diagnostic right heart catheter (RHC) and was available in all but 10 cases. When patients were divided by age, those in the higher age tertile showed more functional impairment despite milder hemodynamics, lower FEV1/FVC ratio and KCO % predicted, as well as milder emphysematous and fibrotic changes on CT scans (Figure S2 and Table S9).

Rare variants in previously established genes

We identified variants in previously established genes (namely, *BMPR2*, *ACVRL1*, *ENG*, *SMAD1*, *SMAD4*, *SMAD9*, *KCNK3*, *TBX4*, *EIF2AK4*, *AQP1*, *ATP13A3*, *GDF2*, *SOX17*) in 271 (24.2%) of the 1,122 cases and interpreted them based on the ACMG standards and guidelines[18]. The

majority of these variants have already been described in Gräf *et al.*[12] (see supplemental material, Figure S4 A-F).

Rare variant association testing

We used BeviMed to consolidate previously reported and discover novel genotype-phenotype associations. Note, that cases explained by rare deleterious variants in previously established genes were only included for the association testing with the respective disease gene (see Methods). This analysis identified 39 significant gene-tag associations with posterior probability (PP) above 0.6 (Table 2 and Figure 2A). *BMPR2, TBX4, EIF2AK4, ACVRL1* showed the highest association (PP ≥0.99) but we also confirmed significant associations in the majority of other previously established genes[12]. Individuals with rare variants in *BMPR2*, *TBX4*, *EIF2AK4* (autosomal recessive) and *SOX17* had a significantly earlier age of disease onset (tag: young age). We also confirmed the association of rare variants in *AQP1* with PAH (log(BF)=10.02, PP=0.984). The refined phenotype approach corroborated the association between mutations in *BMPR2* and preserved KCO (KCO higher tertile, log(BF)=99.92, PP=1) together with an association between biallelic *EIF2AK4* mutations and significantly reduced KCO (KCO <50% predicted, log(BF)= 29.74, PP=1).

Under an autosomal dominant mode of inheritance, PTVs in the kinase insert domain receptor (*KDR)* were associated with a significantly reduced KCO (KCO lower tertile, log(BF)=11.36, PP=0.989) and older age at diagnosis (tag: old age, log(BF)=9.25, PP=0.912). Furthermore, KCO stratification highlighted an association between isocitrate dehydrogenase (NAD(+)) 3 non-catalytic subunit gamma (*IDH3G*) and moderately reduced KCO in patients with PH (log(BF)=7.523, PP=0.649).

Rare variants in the new PAH risk genes: *KDR* and *IDH3G*

We identified five rare PTVs in *KDR* in the study cohort. Four of these were in PAH cases: 1 frameshift variant in exon 3 of 30 (c.183del, p.Tryp61CysfsTer16), 2 nonsense variants, one in exon 3 (c.183G>A, p.Trp61Ter) and one in exon 22 (c.3064C>T, p.Arg1022Ter) and 1 splice acceptor variant in intron 4 of 29 (c.490-1G>A). In addition, one nonsense variant was identified in exon 27 (p.Glu1206Ter) in a non-PAH control subject (Table 3). This latter nonsense variant appears late in the amino acid sequence, in exon 27 of 30, which might have limited impact on *KDR* function, but this remains to be determined by functional studies. Furthermore, 13 PAH cases (1%) and 108 non-PAH controls (0.9%) harbored ultra-rare, predicted-deleterious *KDR* missense variants (Figure 3). The missense variant carriers, however, did not exhibit a reduced KCO or older age of diagnosis. Instead, these patients show the opposite trend in KCO (Figure 2B and C). Importantly, seven of the 13 *KDR* missense variants seen in PAH cases were also detected in several non-PAH controls and thus, are unlikely causal. Furthermore, three of these variants co-occurred with predicted-deleterious variants in established PAH risk genes (two patients with variants in *BMPR2* and one with a variant in *AQP1*).

We additionally identified three missense variants (c.74C>T, p.Pro25Leu; c.1037C>T, p.Thr346Ile; c.1067T>C, p.Met356Thr) in the gene *IDH3G* in four individuals. The missense variant (c.74C>T, p.Pro25Leu) was present in two IPAH cases. The “Moderate” impact category contributed to the detected association. The missense variant p.Pro25Leu is located in the signal peptide (spanning from position 1 to 39) and may affect protein transit to the mitochondria, although this remains to be experimentally validated. Based on the published structure of the IDH3G protein[19], the functional impact of the other two missense variants stays unclear, reflected in discordant SIFT and PolyPhen predictions (Figure 3B). IPAH patients harboring variants in

IDH3G were all females with the median age of 42[32;55] and relatively preserved KCO. One patient had a patent foramen ovale. (Table 3 and S11).

Clinical characterization of *KDR* mutation carriers

Patients with PTV in *KDR* were older and exhibited significantly reduced KCO compared to *KDR* missense variant carriers and *BMPR2* mutation carriers (Figure 2C). In order to determine whether the reduction in KCO was the result of coexistent emphysema secondary to smoking or other parenchymal lung diseases, we performed a detailed analysis of thoracic CT imaging. Three of the four cases did not have a history of smoking. CT scans for all four patients showed a range of lung parenchymal changes (Figure 4). W000229 had evidence of mild mainly subpleural ILD, mild emphysema, and air trapping. W000274 had signs of ILD with traction bronchiectasis in the lower zones, mild air trapping, and mild diffuse ground-glass opacities (GGO) and neovascularity. Also, E001392 showed mild centrilobular ground-glass nodularity in addition to moderate pleural effusion and a trace of air trapping, but no ILD. In these cases, it seemed likely that the observed parenchymal changes contributed to the low KCO. In contrast, E003448 had a low KCO despite only a trace of central nonspecific ground-glass change on the CT images. Comparisons of CT findings between patients harboring deleterious mutations in *BMPR2*, *EIF2AK4*, *KDR,* other PAH risk genes and patients without mutations are presented in Table S10. There were no differences in the frequency of comorbidities between patients harboring missense and PTV in *KDR* although the frequency of systemic hypertension was high (44%) (Table S11). Survival analysis could not be conducted due to the low frequency of mutation carriers, as well as only two events occurring in this group. Following the death of W000229, his daughter, aged 53, was diagnosed with PAH and a reduced KCO at 40% predicted. On the CT scan, mild interstitial fibrosis was visible (Figure S5). Sanger sequencing confirmed that father

and daughter carried the same deleterious *KDR* nonsense variant p.Trp61Ter suggesting familial segregation of the phenotype with the genotype (Figure 3B).

Additional cases in US PAH cohorts

To seek further evidence for *KDR* and *ID3HG* as new risk genes for PAH, we analyzed subjects recruited to the PAHBB[15] and the CUMC[16] to identify additional patients carrying predicted pathogenic rare variants. Four additional individuals harboring *KDR* PTVs were identified. These comprised, 2 nonsense variants, one in exon 3 (c.303C>A, p.Tyr101Ter) and one in exon 22 (c.3064C>T, p.Arg1022Ter) and two splice donor variants, one in intron 2 of 29 (c.161+1G>T) and one in intron 5 (c.658+1G>A). Interestingly, the nonsense variant p.Arg1022Ter appeared in both cohorts (Figure 3). Patient-level data for these individuals are summarized in Table S3. Three of the four patients were diagnosed with idiopathic PAH at 72, 65 and 42 years respectively, whereas one patient was diagnosed at age 4 with PAH associated with double outlet right ventricle. The diffusion capacity of carbon monoxide was available for one patient and was decreased at 35% predicted, with only minor pleural scarring in the left upper lobe found on CT imaging. Two out of four patients (50%) harboring a PTV in *KDR* had also been diagnosed with systemic hypertension.

Additionally, three further individuals carrying rare missense variants at the *IDH3G* locus were found in the US PAHBB and CUMC cohorts; a male neonate diagnosed with Scimitar syndrome, hypoplastic right lung and atrial septal defect (ASD) (c.1091C>T, p.Pro364Leu), a 55-year-old female with large ASD (c.217G>C, p.Val73Leu) and a two-year-old female with CHD-PAH (c.5C>T, p.Ala2Val) (Table 3 and Figure 3B).

Discussion

One of the critical steps in identifying novel, causative genes in rare disorders is the discovery of genotype-phenotype associations to inform patient care and outcomes. A pragmatic focus on deeply-phenotyped individuals and "smart" experimental design provides additional leverage to identify novel risk variants[20]. To deploy this approach in PAH we brought together phenotypic and genetic data using BeviMed[13]. This Bayesian framework allows the inclusion of prior information about the hypothesis in a flexible manner and compares a range of possible genetic models in a single analysis. To generate case-control labels, we tagged PAH cases with diagnostic labels and stratified them by age at diagnosis and KCO. Analyses were then performed to identify associations between tags and ultra-rare gene variants under dominant and recessive modes of inheritance and different variant consequence types.

Our analysis provides strong statistical evidence of an association between ultra-rare PTVs in *KDR* and PAH with reduced KCO and older age at diagnosis under a dominant mode of inheritance. The high impact variants in *KDR* were predicted to truncate the protein. PTVs are expected to have strong effects on disease risk as they dramatically alter protein sequence. *KDR,* which encodes the vascular endothelial growth factor receptor 2 (VEGFR-2)[21] binds VEGF-A, a critical growth factor for physiological and pathological angiogenesis in vascular endothelial cells. In mice, even though VegfA haploinsufficiency is embryonically lethal[22], heterozygosity of its receptor, Vegfr2, is compatible with life and unperturbed vascular development[23].

The role of VEGF signaling in the pathogenesis of PAH has been an area of intense interest since reports of increased expression of VEGF, VEGFR1 and VEGFR2 in rat lung tissue in response to acute and chronic hypoxia[24]. An increase in lung VEGF has also been reported in rats with PH following monocrotaline exposure[25]. In humans, VEGF-A is highly expressed in plexiform lesions in patients with IPAH[26]. In addition, inhibition of VEGF signaling by SU5416 (sugen) combined

with chronic hypoxia triggers severe angioproliferative PH[27]. SU5416, a small-molecule inhibitor of the tyrosine kinase segment of VEGF receptors inhibits VEGFR1[28] and VEGFR2[29] causing endothelial cell apoptosis, loss of lung capillaries and emphysema[30]. Further evidence supporting the role of VEGF inhibition in the pathobiology of PAH comes from reports of PH in patients treated with bevacizumab[31] and the multi-tyrosine kinase inhibitors, dasatinib[32], bosutinib, ponatinib, ruxolitinib, and nilotinib[33]. Besides statistical evidence accompanied by biological plausibility, we also identified one case with a family history, which together with a recently published case report of two families in which PTVs in *KDR* segregated with the phenotype of PAH and significantly reduced KCO[34] amounts to three reported cases with familial segregation. Mutations in *KDR* have also been reported in other cardiovascular diseases. Bleyl et al. reported that *KDR* might be a candidate for familial total anomalous pulmonary venous return[35]. In addition, haploinsufficiency at the *KDR* locus has also been associated with tetralogy of Fallot[36]. We identified one patient in the UCMC cohort with PAH associated with congenital heart disease harboring a *KDR* protein-truncating splice donor variant (c.161+1G>T).

Metabolic disturbances in pulmonary arteries and the right ventricle[37] also reflected in plasma metabolite levels[3] are well-documented drivers of the PAH process. In particular, PAH is associated with a shift from oxidative phosphorylation to glycolysis facilitated by the inhibition of pyruvate dehydrogenase (PDH)[38]. *IDH3G* encodes the gamma subunit of isocitrate dehydrogenase 3 (IDH3). IDH3 is one of the three IDHs, enzymes that are highly conserved from yeast to mammals and are composed of multiple subunits[39]. IDH3 catalyzes the irreversible conversion of isocitrate to alpha-ketoglutarate as a part of the tricarboxylic acid (TCA) cycle[40]. Under normal circumstances, IDH1 and 2 participate in synthetic pathways when the cellular energy level is high enough to inhibit IDH3 and to divert citrate to the cytosol instead of the TCA cycle. Functional studies, using human endothelial cells, indicated that IDH3 activity remained comparable between wild type lines and those harboring disease-causing *BMPR2* variants,

however, mutants displayed a significant increase in IDH1 and 2 activity. Moreover, IDH1 and 2 activity was significantly higher in the serum of patients with H/IPAH irrespective of mutation status, when compared to healthy controls, and correlated with disease severity[41]. We have identified 3 deleterious, heterozygous missense variants in *IDH3G* in 4 PAH patients, which may imbalance the TCA cycle and affect levels of TCA cycle intermediates, that possess signaling properties as well as serve for biosynthesis purposes. This needs confirmation in future functional studies. In the present study, we highlight that deep clinical phenotyping, in combination with genotype data, can accelerate the identification of novel disease risk genes and disease subtypes. Of particular interest is the association of *KDR* PTVs with significantly reduced KCO. Reduced KCO, which reflects impairment of alveolar-capillary membrane function, has been noted in the analysis of early PAH registry data[42] to be an independent predictor of survival. Decreased KCO was also found in patients with PVOD/PCH with or without biallelic *EIF2AK4* mutations[43]. Although some reduction in KCO is one of the typical features of pulmonary vascular disease, PVOD patients show the lowest KCO values when compared to IPAH or CTEPH. In contrast, KCO is relatively preserved in *BMPR2* mutation carriers[44]. Strong association with survival and a link with other causative mutations makes the KCO phenotype particularly attractive for stratification in genetic studies.

As lung disease should always be taken under consideration as a cause of low KCO, we applied the World Symposium on PH criteria[45] to exclude lung disease as a cause of PH: TLC ≥70% pred., FVC ≥70% pred., FEV1 ≥60% pred., and no severe fibrosis and/or emphysema on chest HRCT. None of the *KDR* PTV cases met these criteria, although two of the four patients did show evidence of early ILD. Another potential reason for low KCO in the PAH population is the diagnosis of PVOD/PCH[46]. Again, careful analysis of CT scans and clinical data did not reveal convincing evidence for this diagnosis in *KDR* PTV carriers. Cigarette smoking is a well-known factor leading to the decrease of KCO, which can be explained by increased carboxyhemoglobin

levels[47] and smoking-induced emphysema[48]. Only one of the 4 *KDR* PTV carriers was a previous smoker with 15 pack-years of exposure but non-smoker for over 20 years prior to diagnosis and with no signs of emphysema on CT. These findings suggest that PTVs in *KDR* are associated with a form of PAH characterized by a range of lung parenchymal abnormalities, including small airways disease, emphysema and ILD, as two of the four patients harboring PTVs in *KDR* had mild fibrotic lung changes. Of note, the patients with mutations in other PAH risk genes, or those without the identified genetic mutation, showed less than 10% incidence of fibrotic changes on CT imaging. Further larger studies are needed to determine the full range of lung parenchymal abnormalities in PAH cases with PTVs in *KDR*.

In this study, we have assumed that PAH is a monogenic condition, which is caused by either deleterious heterozygous or biallelic variants in a single gene. This assumption, although widely supported by the literature, may not be accurate in all cases. An alternative hypothesis is that some cases of PAH represent an oligogenic disorder involving two or more genes. Although not statistically explored in the current analysis we found a total of 22 PAH cases carrying deleterious variants in more than one PAH gene. These variants could contribute as genetic modifiers, impacting penetrance and/or expressivity. Finally, we have explored only a limited number of clinical phenotypes. Further studies with larger numbers of phenotypic tags derived from clinical and molecular data will further increase the power to detect new associations.

In summary, this study shows how deep phenotyping enabled patient stratification into subgroups with shared pathobiology and increased the power to detect genotype-phenotype associations. We provided statistical evidence of a strong association between PTVs in the gene *KDR* and significantly decreased KCO, as well as later age of disease onset, and moderate impact variants in *IDH3G* and preserved KCO.

Acknowledgments

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.

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. Exome sequencing and genotyping data were generated by the Regeneron Genetics Center.

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; StevenNathan, 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 Barne

Sources of Funding

The UK National Cohort of Idiopathic and Heritable PAH is supported by the NIHRBR-RD, the British Heart Foundation (BHF) (SP/12/12/29836 and SP/18/10/33975), the BHF Cambridge Centre of Cardiovascular Research Excellence, and the UK Medical Research Council (MR/K020919/1), the Dinosaur Trust, BHF Programme grants to RCT (RG/08/006/25302) and NWM (RG/13/4/30107), and the UK NIHR National Institute for Health Research Cambridge Biomedical Research Centre. NWM is a BHF Professor and NIHR Senior Investigator. AL is supported by a BHF Senior Basic Science Research Fellowship (FS/13/48/30453).

Samples and/or data from the National Biological Sample and Data Repository for PAH, funded by an NIH investigator-initiated resources grant (R24 HL105333 to WCN), were used in this study.

Disclosures

NWM is a Director and Co-founder of Morphogen-IX. JW received personal fees from Actelion Pharmaceuticals. GK reports personal fees and non-financial support from Actelion Pharmaceuticals, Bayer, GSK, MSD, Boehringer Ingelheim, Novartis, Chiesi and Vitalaire outside the submitted work. CJR declares fees from Actelion Pharmaceuticals and United Therapeutics. AL received support and fees from GSK and Actelion Pharmaceuticals.

References

1. Voelkel NF, Gomez-Arroyo J, Abbate A, Bogaard HJ, Nicolls MR. Pathobiology of pulmonary arterial hypertension and right ventricular failure [Internet]. European Respiratory Journal. 2012;40:1555–1565. Available from: http://dx.doi.org/10.1183/09031936.00046612

2. Tuder RM, Groves B, Badesch DB, Voelkel NF. Exuberant endothelial cell growth and elements of inflammation are present in plexiform lesions of pulmonary hypertension. *Am J Pathol*. 1994;144:275–285.

3. Rhodes CJ, Ghataorhe P, Wharton J, Rue-Albrecht KC, Hadinnapola C, Watson G, Bleda M, Haimel M, Coghlan G, Corris PA, Howard LS, Kiely DG, Peacock AJ, Pepke-Zaba J, Toshner MR, Wort SJ, Gibbs JSR, Lawrie A, Gräf S, Morrell NW, Wilkins MR. Plasma Metabolomics Implicates Modified Transfer RNAs and Altered Bioenergetics in the Outcomes of Pulmonary Arterial Hypertension. *Circulation*. 2017;135:460–475.

4. Pullamsetti SS, Savai R, Seeger W, Goncharova EA. Translational Advances in the Field of Pulmonary Hypertension. From Cancer Biology to New Pulmonary Arterial Hypertension Therapeutics. Targeting Cell Growth and Proliferation Signaling Hubs. *Am J Respir Crit Care Med*. 2017;195:425–437.

5. Machado RD, Pauciulo MW, Thomson JR, Lane KB, Morgan NV, Wheeler L, Phillips JA 3rd, Newman J, Williams D, Galiè N, Manes A, McNeil K, Yacoub M, Mikhail G, Rogers P, Corris P, Humbert M, Donnai D, Martensson G, Tranebjaerg L, Loyd JE, Trembath RC, Nichols WC. BMPR2 haploinsufficiency as the inherited molecular mechanism for primary pulmonary hypertension. *Am J Hum Genet*. 2001;68:92–102.

6. Larkin EK, Newman JH, Austin ED, Hemnes AR, Wheeler L, Robbins IM, West JD, Phillips JA 3rd, Hamid R, Loyd JE. Longitudinal analysis casts doubt on the presence of genetic anticipation in heritable pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2012;186:892–896.

7. Trembath RC. Mutations in the TGF-beta type 1 receptor, ALK1, in combined primary pulmonary hypertension and hereditary haemorrhagic telangiectasia, implies pathway specificity. *J Heart Lung Transplant*. 2001;20:175.

8. Chaouat A. Endoglin germline mutation in a patient with hereditary haemorrhagic telangiectasia and dexfenfluramine associated pulmonary arterial hypertension [Internet]. Thorax. 2004;59:446–448. Available from: http://dx.doi.org/10.1136/thx.2003.11890

9. Shintani M, Yagi H, Nakayama T, Saji T, Matsuoka R. A new nonsense mutation of SMAD8 associated with pulmonary arterial hypertension [Internet]. Journal of Medical Genetics. 2009;46:331–337. Available from: http://dx.doi.org/10.1136/jmg.2008.062703

10. Austin ED, Ma L, LeDuc C, Berman Rosenzweig E, Borczuk A, Phillips JA 3rd, Palomero T, Sumazin P, Kim HR, Talati MH, West J, Loyd JE, Chung WK. Whole exome sequencing to identify a novel gene (caveolin-1) associated with human pulmonary arterial hypertension. *Circ Cardiovasc Genet*. 2012;5:336–343.

11. Ma L, Roman-Campos D, Austin ED, Eyries M, Sampson KS, Soubrier F, Germain M, Trégouët D-A, Borczuk A, Rosenzweig EB, Girerd B, Montani D, Humbert M, Loyd JE, Kass RS, Chung WK. A Novel Channelopathy in Pulmonary Arterial Hypertension [Internet]. New England Journal of Medicine. 2013;369:351–361. Available from:

http://dx.doi.org/10.1056/nejmoa1211097

12. Gräf S, Haimel M, Bleda M, Hadinnapola C, Southgate L, Li W, Hodgson J, Liu B, Salmon RM, Southwood M, Machado RD, Martin JM, Treacy CM, Yates K, Daugherty LC, Shamardina O, Whitehorn D, Holden S, Aldred M, Bogaard HJ, Church C, Coghlan G, Condliffe R, Corris PA, Danesino C, Eyries M, Gall H, Ghio S, Ghofrani H-A, Gibbs JSR, Girerd B, Houweling AC, Howard L, Humbert M, Kiely DG, Kovacs G, MacKenzie Ross RV, Moledina S, Montani D, Newnham M, Olschewski A, Olschewski H, Peacock AJ, Pepke-Zaba J, Prokopenko I, Rhodes CJ, Scelsi L, Seeger W, Soubrier F, Stein DF, Suntharalingam J, Swietlik EM, Toshner MR, van Heel DA, Vonk Noordegraaf A, Waisfisz Q, Wharton J, Wort SJ, Ouwehand WH, Soranzo N, Lawrie A, Upton PD, Wilkins MR, Trembath RC, Morrell NW. Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nat Commun*. 2018;9:1416.

13. Greene D, NIHR BioResource, Richardson S, Turro E. A Fast Association Test for Identifying Pathogenic Variants Involved in Rare Diseases. *Am J Hum Genet*. 2017;101:104–114.

14. Ouwehand WH. Whole-genome sequencing of rare disease patients in a national healthcare system [Internet]. bioRxiv. 2020 [cited 2020 Feb 19];507244. Available from: https://www.biorxiv.org/content/10.1101/507244v3.abstract

15. Zhu N, Pauciulo MW, Welch CL, Lutz KA, Coleman AW, Gonzaga-Jauregui C, Wang J, Grimes JM, Martin LJ, He H, PAH Biobank Enrolling Centers' Investigators, Shen Y, Chung WK, Nichols WC. Novel risk genes and mechanisms implicated by exome sequencing of 2572 individuals with pulmonary arterial hypertension. *Genome Med*. 2019;11:69.

16. Zhu N, Welch CL, Wang J, Allen PM, Gonzaga-Jauregui C, Ma L, King AK, Krishnan U, Rosenzweig EB, Ivy DD, Austin ED, Hamid R, Pauciulo MW, Lutz KA, Nichols WC, Reid JG, Overton JD, Baras A, Dewey FE, Shen Y, Chung WK. Rare variants in SOX17 are associated with pulmonary arterial hypertension with congenital heart disease. *Genome Med*. 2018;10:56.

17. Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, Chuang H-Y, Källberg M, Kumar SA, Liao A, Little KM, Strömberg MP, Tanner SW. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics*. 2013;29:2041–2043.

18. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–424.

19. Ma T, Peng Y, Huang W, Ding J. Molecular mechanism of the allosteric regulation of the αγ heterodimer of human NAD-dependent isocitrate dehydrogenase. *Sci Rep*. 2017;7:40921.

20. FitzGerald G, Botstein D, Califf R, Collins R, Peters K, Van Bruggen N, Rader D. The future of humans as model organisms [Internet]. Science. 2018;361:552–553. Available from: http://dx.doi.org/10.1126/science.aau7779

21. Terman BI, Carrion ME, Kovacs E, Rasmussen BA, Eddy RL, Shows TB. Identification of a new endothelial cell growth factor receptor tyrosine kinase. *Oncogene*.

1991;6:1677–1683.

22. Ferrara N, Carver-Moore K, Chen H, Dowd M, Lu L, O'Shea KS, Powell-Braxton L, Hillan KJ, Moore MW. Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature*. 1996;380:439–442.

23. Oladipupo SS, Smith C, Santeford A, Park C, Sene A, Wiley LA, Osei-Owusu P, Hsu J, Zapata N, Liu F, Nakamura R, Lavine KJ, Blumer KJ, Choi K, Apte RS, Ornitz DM. Endothelial cell FGF signaling is required for injury response but not for vascular homeostasis [Internet]. Proceedings of the National Academy of Sciences. 2014;111:13379–13384. Available from: http://dx.doi.org/10.1073/pnas.1324235111

24. Tuder RM, Flook BE, Voelkel NF. Increased gene expression for VEGF and the VEGF receptors KDR/Flk and Flt in lungs exposed to acute or to chronic hypoxia. Modulation of gene expression by nitric oxide. *J Clin Invest*. 1995;95:1798–1807.

25. Cho YJ, Han JY, Lee SG, Jeon BT, Choi WS, Hwang YS, Roh GS, Lee JD. Temporal changes of angiopoietins and Tie2 expression in rat lungs after monocrotaline-induced pulmonary hypertension. *Comp Med*. 2009;59:350–356.

26. Tuder RM, Chacon M, Alger L, Wang J, Taraseviciene-Stewart L, Kasahara Y, Cool CD, Bishop AE, Geraci M, Semenza GL, Yacoub M, Polak JM, Voelkel NF. Expression of angiogenesis-related molecules in plexiform lesions in severe pulmonary hypertension: evidence for a process of disordered angiogenesis [Internet]. The Journal of Pathology. 2001;195:367–374. Available from: http://dx.doi.org/10.1002/path.953.abs

27. Taraseviciene-Stewart L, Kasahara Y, Alger L, Hirth P, Mc Mahon G, Waltenberger J, Voelkel NF, Tuder RM. Inhibition of the VEGF receptor 2 combined with

chronic hypoxia causes cell death-dependent pulmonary endothelial cell proliferation and severe pulmonary hypertension. *FASEB J*. 2001;15:427–438.

28. Itokawa T, Nokihara H, Nishioka Y, Sone S, Iwamoto Y, Yamada Y, Cherrington J, McMahon G, Shibuya M, Kuwano M, Ono M. Antiangiogenic effect by SU5416 is partly attributable to inhibition of Flt-1 receptor signaling. *Mol Cancer Ther*. 2002;1:295–302.

29. Fong TA, Shawver LK, Sun L, Tang C, App H, Powell TJ, Kim YH, Schreck R, Wang X, Risau W, Ullrich A, Hirth KP, McMahon G. SU5416 is a potent and selective inhibitor of the vascular endothelial growth factor receptor (Flk-1/KDR) that inhibits tyrosine kinase catalysis, tumor vascularization, and growth of multiple tumor types. *Cancer Res*. 1999;59:99–106.

30. Kasahara Y, Tuder RM, Taraseviciene-Stewart L, Le Cras TD, Abman S, Hirth PK, Waltenberger J, Voelkel NF. Inhibition of VEGF receptors causes lung cell apoptosis and emphysema. *J Clin Invest*. 2000;106:1311–1319.

31. Garcia AA, Hirte H, Fleming G, Yang D, Tsao-Wei DD, Roman L, Groshen S, Swenson S, Markland F, Gandara D, Scudder S, Morgan R, Chen H, Lenz H-J, Oza AM. Phase II Clinical Trial of Bevacizumab and Low-Dose Metronomic Oral Cyclophosphamide in Recurrent Ovarian Cancer: A Trial of the California, Chicago, and Princess Margaret Hospital Phase II Consortia [Internet]. Journal of Clinical Oncology. 2008;26:76–82. Available from: http://dx.doi.org/10.1200/jco.2007.12.1939

32. Montani D, Bergot E, Günther S, Savale L, Bergeron A, Bourdin A, Bouvaist H, Canuet M, Pison C, Macro M, Poubeau P, Girerd B, Natali D, Guignabert C, Perros F, O'Callaghan DS, Jaïs X, Tubert-Bitter P, Zalcman G, Sitbon O, Simonneau G, Humbert M.

Pulmonary Arterial Hypertension in Patients Treated by Dasatinib [Internet]. Circulation. 2012;125:2128–2137. Available from: http://dx.doi.org/10.1161/circulationaha.111.079921

33. El-Dabh A, Acharya D. EXPRESS: Pulmonary hypertension with dasatinib and other tyrosine kinase inhibitors. *Pulm Circ*. 2019;2045894019865704.

34. Eyries M, Montani D, Girerd B, Favrolt N, Riou M, Faivre L, Manaud G, Perros F, Gräf S, Morrell NW, Humbert M, Soubrier F. Familial pulmonary arterial hypertension by KDR heterozygous loss of function [Internet]. European Respiratory Journal. 2020;1902165. Available from: http://dx.doi.org/10.1183/13993003.02165-2019

35. Bleyl S, Nelson L, Odelberg SJ, Ruttenberg HD, Otterud B, Leppert M, Ward K. A gene for familial total anomalous pulmonary venous return maps to chromosome 4p13-q12. *Am J Hum Genet*. 1995;56:408–415.

36. Reuter MS, Jobling R, Chaturvedi RR, Manshaei R, Costain G, Heung T, Curtis M, Hosseini SM, Liston E, Lowther C, Oechslin E, Sticht H, Thiruvahindrapuram B, van Mil S, Wald RM, Walker S, Marshall CR, Silversides CK, Scherer SW, Kim RH, Bassett AS. Haploinsufficiency of vascular endothelial growth factor related signaling genes is associated with tetralogy of Fallot. *Genet Med*. 2019;21:1001–1007.

37. Sakao S, Miyauchi H, Voelkel NF, Sugiura T, Tanabe N, Kobayashi Y, Tatsumi K. Increased Right Ventricular Fatty Acid Accumulation in Chronic Thromboembolic Pulmonary Hypertension. *Ann Am Thorac Soc*. 2015;12:1465–1472.

38. Sutendra G, Dromparis P, Bonnet S, Haromy A, McMurtry MS, Chris Bleackley R, Michelakis ED. Pyruvate dehydrogenase inhibition by the inflammatory cytokine TNFα contributes to the pathogenesis of pulmonary arterial hypertension [Internet]. Journal of

Molecular Medicine. 2011;89:771–783. Available from: http://dx.doi.org/10.1007/s00109-011-0762-2

39. McAlister-Henn L. Ligand binding and structural changes associated with allostery in yeast NAD(+)-specific isocitrate dehydrogenase. *Arch Biochem Biophys*. 2012;519:112–117.

40. Huh T-L, Kim Y-O, Oh I-U, Song BJ, Inazawa J. Assignment of the Human Mitochondrial NAD -Specific Isocitrate Dehydrogenase α Subunit (IDH3A) Gene to 15q25.1 → q25.2 byin SituHybridization [Internet]. Genomics. 1996;35:274. Available from: http://dx.doi.org/10.1006/geno.1996.0357

41. Fessel JP, Hamid R, Wittmann BM, Robinson LJ, Blackwell T, Tada Y, Tanabe N, Tatsumi K, Hemnes AR, West JD. Metabolomic analysis of bone morphogenetic protein receptor type 2 mutations in human pulmonary endothelium reveals widespread metabolic reprogramming. *Pulm Circ*. 2012;2:201–213.

42. Test VJ, Farber HW, McGoon MD, Parsons L, Channick RN. Pulmonary Arterial Hypertension in the Elderly: Baseline Characteristics and Evaluation of Therapeutics. An Examination of the Reveal Registry [Internet]. B27. FROM ALPHA TO OMEGA: ASSESSMENT AND OUTCOMES IN PULMONARY HYPERTENSION. 2009;Available from: http://dx.doi.org/10.1164/ajrccm-conference.2009.179.1_meetingabstracts.a2649

43. Hadinnapola C, Bleda M, Haimel M, Screaton N, Swift A, Dorfmüller P, Preston SD, Southwood M, Hernandez-Sanchez J, Martin J, Treacy C, Yates K, Bogaard H, Church C, Coghlan G, Condliffe R, Corris PA, Gibbs S, Girerd B, Holden S, Humbert M, Kiely DG, Lawrie A, Machado R, MacKenzie Ross R, Moledina S, Montani D, Newnham M,

Peacock A, Pepke-Zaba J, Rayner-Matthews P, Shamardina O, Soubrier F, Southgate L, Suntharalingam J, Toshner M, Trembath R, Vonk Noordegraaf A, Wilkins MR, Wort SJ, Wharton J, NIHR BioResource–Rare Diseases Consortium; UK National Cohort Study of Idiopathic and Heritable PAH, Gräf S, Morrell NW. Phenotypic Characterization of Mutation Carriers in a Large Cohort of Patients Diagnosed Clinically With Pulmonary Arterial Hypertension. *Circulation*. 2017;136:2022–2033.

44. Trip P, Girerd B, Bogaard H-J, de Man FS, Boonstra A, Garcia G, Humbert M, Montani D, Vonk-Noordegraaf A. Diffusion capacity and BMPR2 mutations in pulmonary arterial hypertension. *Eur Respir J*. 2014;43:1195–1198.

45. Nathan SD, Barbera JA, Gaine SP, Harari S, Martinez FJ, Olschewski H, Olsson KM, Peacock AJ, Pepke-Zaba J, Provencher S, Weissmann N, Seeger W. Pulmonary hypertension in chronic lung disease and hypoxia. *Eur Respir J* [Internet]. 2019;53. Available from: http://dx.doi.org/10.1183/13993003.01914-2018

46. Montani D, Dorfmuller P, Maitre S, Jaïs X, Sitbon O, Simonneau G, Humbert M. [Pulmonary veno-occlusive disease and pulmonary capillary hemangiomatosis]. *Presse Med*. 2010;39:134–143.

47. Sansores RH, Pare PD, Abboud RT. Acute effect of cigarette smoking on the carbon monoxide diffusing capacity of the lung. *Am Rev Respir Dis*. 1992;146:951–958.

48. López-Campos JL, Soler-Cataluña JJ, Miravitlles M. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2019 Report: Future Challenges [Internet]. Archivos de Bronconeumología. 2019;Available from: http://dx.doi.org/10.1016/j.arbres.2019.06.001

49. Pang C, Sollie A, Sijtsma A, Hendriksen D, Charbon B, de Haan M, de Boer T, Kelpin F, Jetten J, van der Velde JK, Smidt N, Sijmons R, Hillege H, Swertz MA. SORTA: a system for ontology-based re-coding and technical annotation of biomedical phenotype data [Internet]. Database. 2015;2015:bav089. Available from: http://dx.doi.org/10.1093/database/bav089

50. Ling Y, Johnson MK, Kiely DG, Condliffe R, Elliot CA, Gibbs JSR, Howard LS, Pepke-Zaba J, Sheares KKK, Corris PA, Fisher AJ, Lordan JL, Gaine S, Coghlan JG, Wort SJ, Gatzoulis MA, Peacock AJ. Changing demographics, epidemiology, and survival of incident pulmonary arterial hypertension: results from the pulmonary hypertension registry of the United Kingdom and Ireland. *Am J Respir Crit Care Med*. 2012;186:790–796.

51. Hoeper MM, Huscher D, Ghofrani HA, Delcroix M, Distler O, Schweiger C, Grunig E, Staehler G, Rosenkranz S, Halank M, Held M, Grohé C, Lange TJ, Behr J, Klose H, Wilkens H, Filusch A, Germann M, Ewert R, Seyfarth H-J, Olsson KM, Opitz CF, Gaine SP, Vizza CD, Vonk-Noordegraaf A, Kaemmerer H, Gibbs JSR, Pittrow D. Elderly patients diagnosed with idiopathic pulmonary arterial hypertension: results from the COMPERA registry. *Int J Cardiol*. 2013;168:871–880.

52. van der Bruggen CE, Spruijt OA, Nossent EJ, Trip P, Marcus JT, de Man FS, Jan Bogaard H, Noordegraaf AV. Treatment response in patients with idiopathic pulmonary arterial hypertension and a severely reduced diffusion capacity. *Pulm Circ*. 2017;7:137–144.

53. Laveneziana P, Montani D, Dorfmüller P, Girerd B, Sitbon O, Jaïs X, Savale L, Eyries M, Soubrier F, Similowski T, Simonneau G, Humbert M, Garcia G. Mechanisms of exertional dyspnoea in pulmonary veno-occlusive disease with EIF2AK4 mutations

[Internet]. European Respiratory Journal. 2014;44:1069–1072. Available from: http://dx.doi.org/10.1183/09031936.00088914

54. Montani D, Dorfmüller P, Girerd B, Le Pavec J, Fadel E, Simonneau G, Sitbon O, Humbert M. Natural History over 8 Years of Pulmonary Vascular Disease in a Patient Carrying Biallelic EIF2AK4 Mutations [Internet]. American Journal of Respiratory and Critical Care Medicine. 2018;198:537–541. Available from: http://dx.doi.org/10.1164/rccm.201802-0317le

55. Godinas L, Amar D, Montani D, Lau EM, Jaïs X, Savale L, Jevnikar M, Sitbon O, Simonneau G, Humbert M, Laveneziana P, Garcia G. Lung capillary blood volume and membrane diffusion in precapillary pulmonary hypertension. *J Heart Lung Transplant*. 2016;35:647–656.

56. Eichstaedt CA, Song J, Benjamin N, Harutyunova S, Fischer C, Grünig E, Hinderhofer K. EIF2AK4 mutation as “second hit” in hereditary pulmonary arterial hypertension. *Respir Res*. 2016;17:141.

57. Barozzi C, Galletti M, Tomasi L, De Fanti S, Palazzini M, Manes A, Sazzini M, Galiè N. A Combined Targeted and Whole Exome Sequencing Approach Identified Novel Candidate Genes Involved in Heritable Pulmonary Arterial Hypertension. *Sci Rep*. 2019;9:753.

Figure legends

Figure 1. Design of the genetic association study. A, Overview of the analytical approach. Using deep phenotyping, data tags were assigned to patients who shared phenotypic features

(see Figure S2 for more details). Rare sequence variants, called from whole-genome sequencing data, were filtered and explained cases were labeled. BeviMed was applied to a set of unrelated individuals, to test the posterior probability of gene-tag associations. **B**, Consort diagram summarising the size of the study cohort. **C**, Schematic representation of the definition of cases, exemplified by the KCO lower tertile tag. Cases were defined as individuals carrying a particular tag, whereas patients with missing information or those without a tag were removed from the gene-tag association testing. Individuals from non-PAH domains served as controls. KCO - transfer coefficient of carbon monoxide, MAF - minor allele frequency.

Figure 2. Genetic association study results revealing established and novel genotype-phenotype links. A, Figure showing phenotype tags on the x-axis and corresponding posterior probability of genotype-phenotype association on the y-axis, as calculated by BeviMed. The definitions of the tags are listed in Table 1. Shape and colour of points indicate the mode of inheritance and consequence type/impact of variants driving the association. Box-and-whisker plots showing the distribution of transfer coefficient of carbon monoxide (**B**) and age at diagnosis (**C**) stratified by genotype across the PAH domain. The two-tailed Wilcoxon signed-rank test was used to determine differences in the medians of the distributions, which are indicated by the bars at the top of the figures providing the respective p-values. bial. - biallelic, mis. - missense, pt. - protein-truncating.

Figure 3. Summary of single nucleotide variants (SNVs) and small insertions and deletions (indels) identified in the two novel candidate PAH disease risk genes. Only predicted deleterious variants in *KDR* (**A, B**) and *IDH3G* (**C**) are shown (MAF<1/10,000 and CADD≥10).

SNVs and indels are represented by colored lollipops on top of the protein sequence. The domain annotations were retrieved from Uniprot (accession numbers P35968 [*KDR* (**A**)] and P51553 [*IDH3G* (**B**)]). Lollipop color indicates the consequence type and size represents the variant frequency within a cohort. Missense variants that are supported by both SIFT (deleterious) and PolyPhen-2 (damaging) are colored in red, otherwise in yellow (i.e. SIFT and PolyPhen-2 disagree). Relevant variants are labelled with the respective HGVS notation. The number of variants by predicted consequence type and cohort is provided in the tables. **C**, Familial segregation of *KDR* nonsense variant c.183G>A (p.Trp61Ter) with PAH from father (W000229) to daughter (W000229.d). Sanger sequencing results are shown in the chromatograms.

Figure 4. Chest computerized tomography (CT) scans of patients carrying protein-truncating *KDR* mutations. A, Axial image of pulmonary CT angiogram at the level of the right ventricle (RV) moderator band, showing flattening of intraventricular septum, leftwards bowing of the interatrial septum and the enlargement of the right atrium (RA) and RV, indicative of RV strain; bilateral pleural effusion, larger on the right side. **B**, Axial image of a pulmonary CT angiogram demonstrating enlarged pulmonary artery and mild central lung ground-glass opacity (GGO). **C**, Axial high-resolution CT slice of the chest in the lung window showing a trace of non-specific GGO with a central distribution. **D**, Coronal image showing the trace of central GGO and enlarged central pulmonary arteries. Axial high-resolution CT slice of the chest in the lung window showing apical subpleural fibrosis (**E**), and very minor subpleural fibrosis at the lung bases (**F**). Axial high-resolution CT slice of the chest in the lung window showing subpleural GGO at apical level (**G**), and mild GGO at mid-thoracic level (**H**). Patients: E001392 (**A**, **B**), E003448 (**C**, **D**), W000229 (**E**, **F**), W000274 (**G** and **H**).

Table legends

Table 1. Definitions of labels and the number of unrelated cases and controls for genetic association analysis with BeviMed. mPAP - mean pulmonary artery pressure, PH - pulmonary hypertension, PAH - pulmonary arterial hypertension, I/HPAH - Idiopathic/Hereditary Pulmonary Arterial Hypertension, PVOD - Pulmonary veno-occlusive disease, PCH - Pulmonary capillary haemangiomatosis, APAH - Associated Pulmonary Arterial Hypertension, CHD - Congenital Heart Disease, CTD - Connective Heart Disease, PPH, LHD - Left Heart Disease, LD - Lung Disease, CTEPH - Chronic Thromboembolic Pulmonary Hypertension, KCO - transfer coefficient of carbon monoxide.

Table 2. BeviMed analysis results. Posterior probabilities and Bayes factors (the ratio between the probabilities of the data under H1 and under H0. The observed data are BF times more likely under H1 than under H0, and so the larger the BF, the stronger the support in the data for H1 compared with H0) of gene-tag associations. The "High" category, comprises only variants of “high” impact, including PTVs and large deletions; the Moderate category contains variants of impact “moderate”, including missense variants or consequence “non_coding_transcript_exon_variant”; the combined category Moderate and High, include both respective consequence types.

Table 3. Gene changes for IPAH patients harboring protein-truncating variants (PTV) in the KDR gene and PTV and missense variants in the *IDH3G* gene. None of the *KDR* variants has been previously reported in gnomAD, ExAC or internal controls. For *KDR* HGVSc notations are based on transcript sequence ENST00000263923.4. HGVSp notations are based on amino acid sequence ENSP00000263923.4. None of the patients harboring PTV in *KDR* had capillary

hemangioma. For *IDH3G* HGVSc notations are based on transcript sequence ENST00000217901.5, HGVSp notations are based on amino acid sequence ENSP00000217901.5. Protein truncating variants were defined as stop gained, splice acceptor variants or frameshift variants.

Abbreviations: *KDR* - Kinase insert domain receptor, *IDH3G* - Isocitrate dehydrogenase (NAD(+)) 3 non-catalytic subunit gamma, WHO FC - World Health Organization functional class, 6MWD - 6-minute walk distance, SpO_2 - arterial oxygen saturation, mRAP - mean right atrial pressure, mPAP - mean pulmonary artery pressure, mPAWP - mean pulmonary artery wedge pressure, CO - cardiac output, PVR - pulmonary vascular resistance, FEV_1 - forced expiratory volume in 1 sec, FVC - forced vital capacity, KCO - transfer factor coefficient for carbon monoxide, HTN - systemic hypertension, CAD - coronary artery disease, OA - osteoarthritis, DM - diabetes mellitus, OSA - obstructive sleep apnea, ASD - atrial septal defect, VSD - ventricular septal defect., * DLCO% predicted

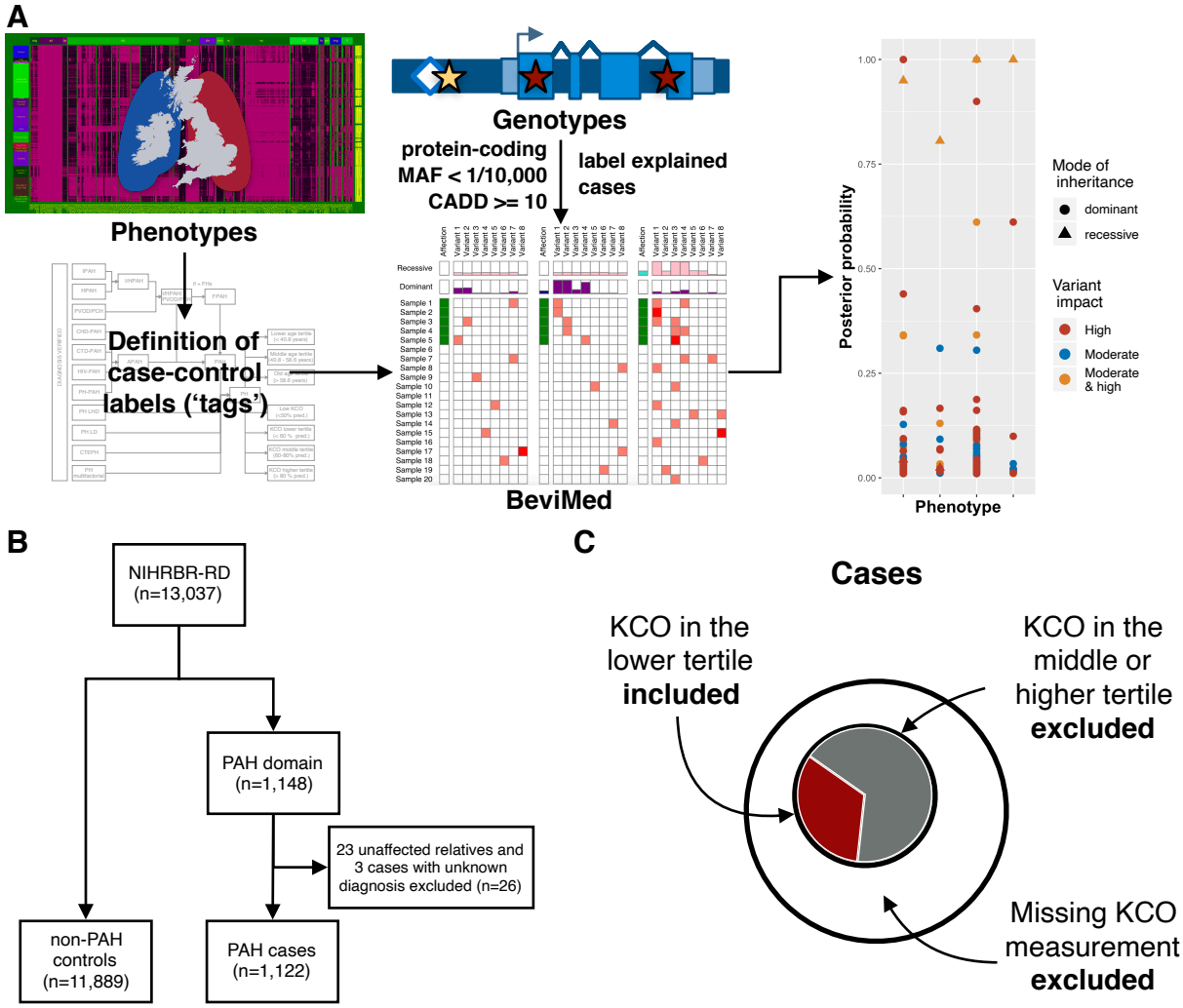


Figure 1. Design of the genetic association study. A, Overview of the analytical approach. Using deep phenotyping, data tags were assigned to patients who shared phenotypic features (see Figure S1 for more details). Rare sequence variants, called from whole-genome sequencing data, were filtered and explained cases were labelled. BeviMed was applied to a set of unrelated individuals, to test the posterior probability of gene-tag associations. **B**, Consort diagram summarising the size of the study cohort. **C**, Schematic representation of the definition of cases, exemplified by the KCO lower tertile tag. Cases were defined as individuals carrying a particular tag, whereas patients with missing information or those without a tag were removed from the gene-tag association testing. Individuals from non-PAH domains served as controls. KCO - transfer coefficient of carbon monoxide, MAF - minor allele frequency.

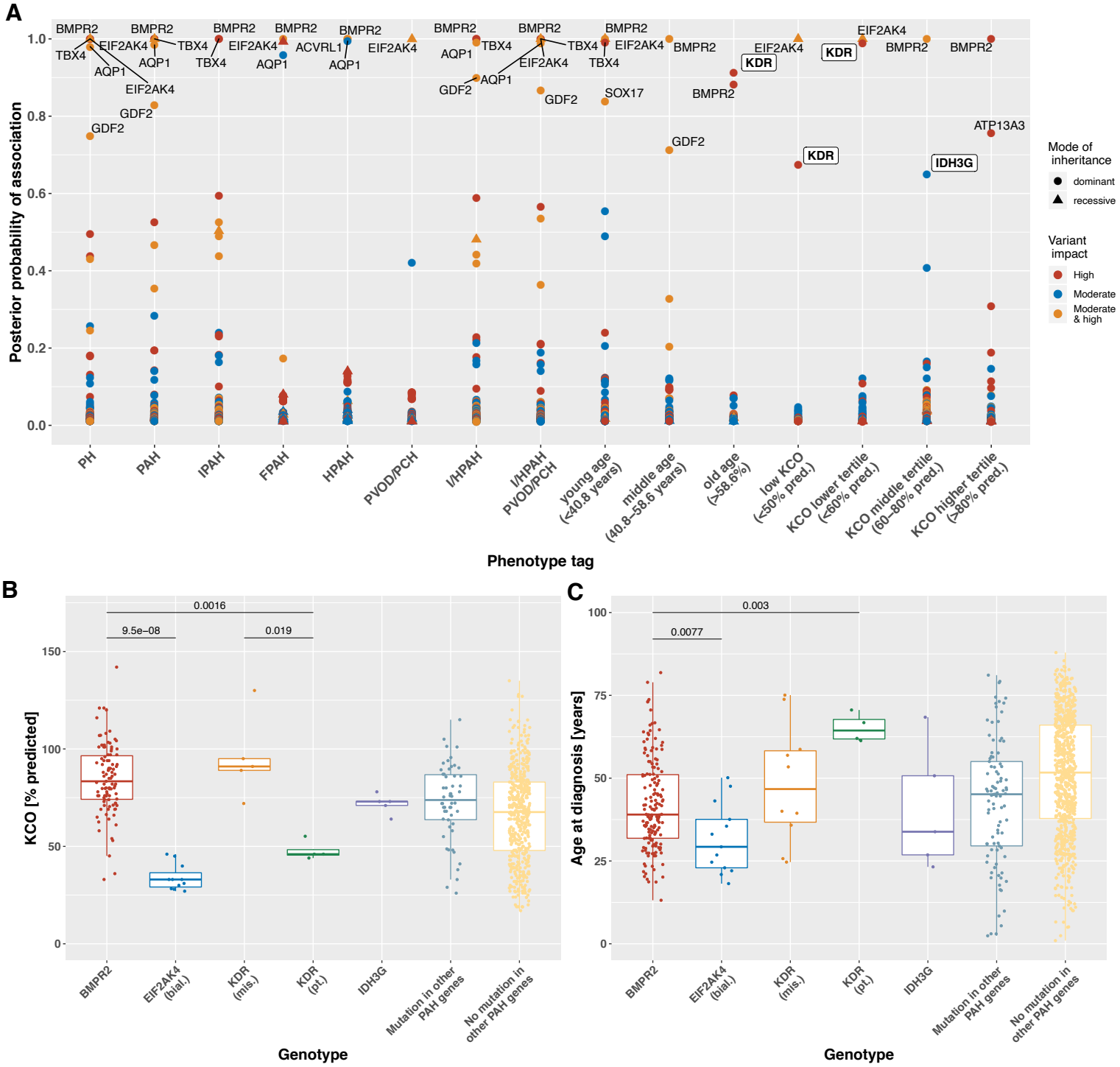
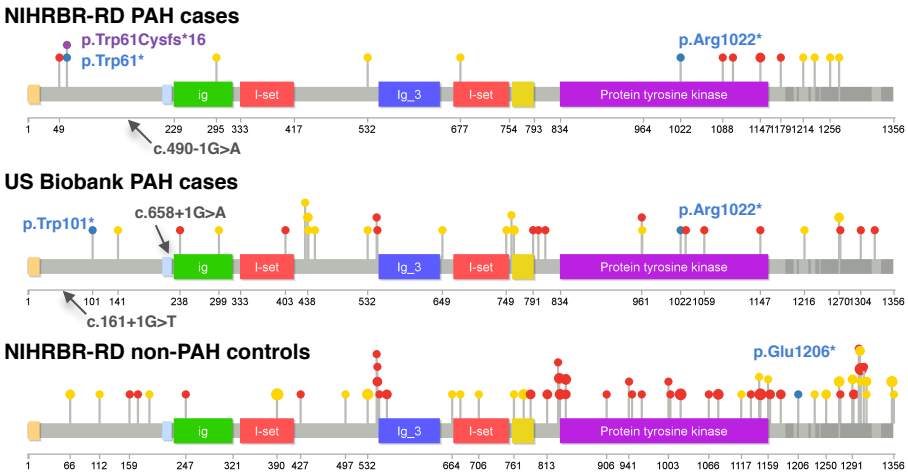


Figure 2. Genetic association study results revealing established and novel genotype-phenotype links. A, Figure showing phenotype tags on the x-axis and corresponding posterior probability of genotype-phenotype association on the y-axis, as calculated by BeviMed. The definitions of the tags are listed in Table 1. Shape and colour of points indicate the mode of inheritance and consequence type/impact of variants driving the association. Box-and-whisker plots showing the distribution of transfer coefficient of carbon monoxide (**B**) and age at diagnosis (**C**) stratified by genotype across the PAH domain. The two-tailed Wilcoxon signed-rank test was used to determine differences in the medians of the distributions, which are indicated by the bars at the top of the figures providing the respective p-values. bial. - biallelic, mis. - missense, pt. - protein-truncating.

A

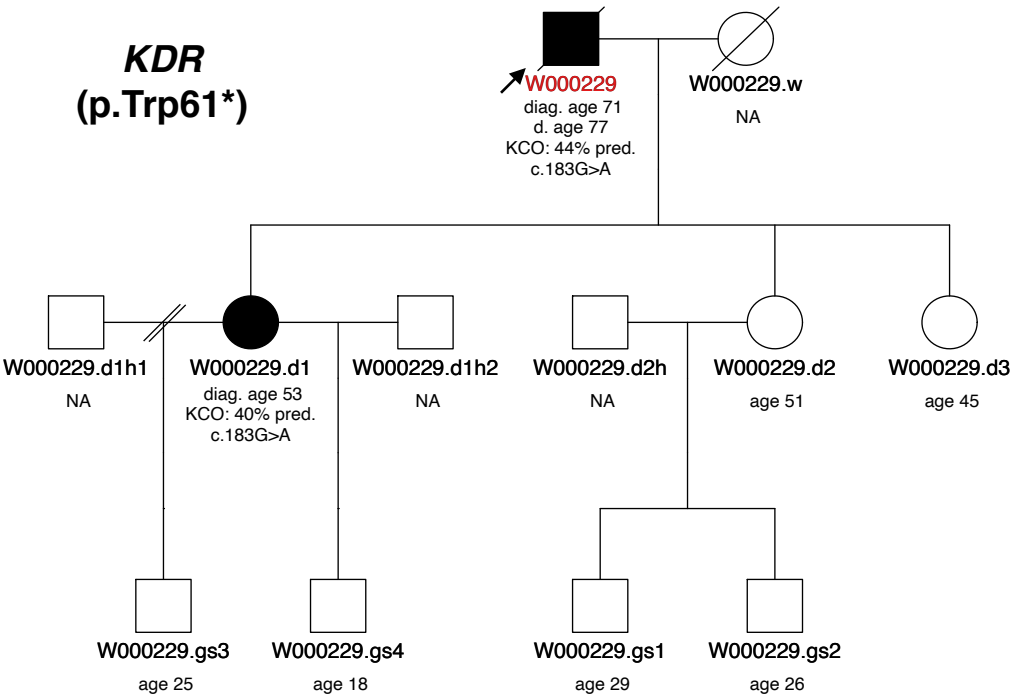
KDR SNVs and indels

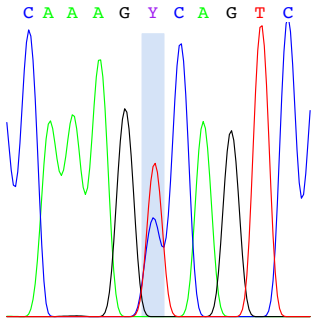


Consequence type
- Stop gained
- Frameshift
- Splice site
- Missense (SIFT / PolyPhen-2 agree)
- Missense (SIFT / PolyPhen-2 disagree)

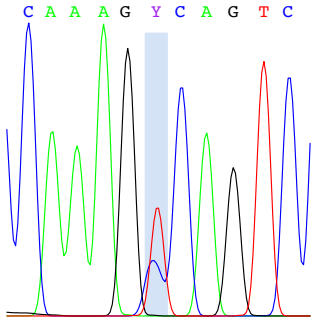
KDR	NIHRBR-RD		USBB
	Cases	Controls	Cases
frameshift_variant	1	0	0
missense_variant	13	107	31
splice_acceptor_variant	1	0	0
splice_donor_variant	0	0	2
stop_gained	2	1	2

B





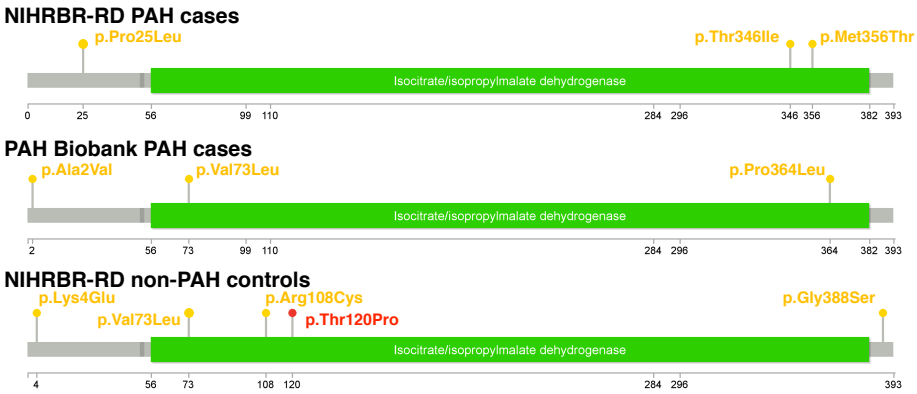
W000229



W000229.d1

C

IDH3G SNVs and indels



IDH3G	NIHRBR-RD		USBB
	Cases	Controls	Cases
missense_variant	4	6	3

Figure 3. Summary of single nucleotide variants (SNVs) and small insertions and deletions (indels) identified in the two novel candidate PAH disease risk genes. Only predicted deleterious variants in *KDR* (**A, B**) and *IDH3G* (**C**) are shown (MAF<1/10,000 and CADD≥10). SNVs and indels are represented by coloured lollipops on top of the protein sequence. The domain annotations were retrieved from Uniprot (accession numbers P35968 [*KDR* (**A**)] and P51553 [*IDH3G* (**C**)]). Lollipop colour indicates the consequence type and size represents the variant frequency within a cohort. Missense variants that are supported by both SIFT (deleterious) and PolyPhen-2 (damaging) are coloured in red, otherwise in yellow (i.e. SIFT and PolyPhen-2 disagree). Relevant variants are labelled with the respective HGVS notation. The number of variants by predicted consequence type and cohort is provided in the tables. **C**, Familial segregation of *KDR* nonsense variant c.183G>A (p.Trp61*) with PAH from father (W000229) to daughter (W000229.d). Sanger sequencing results are shown in the chromatograms.

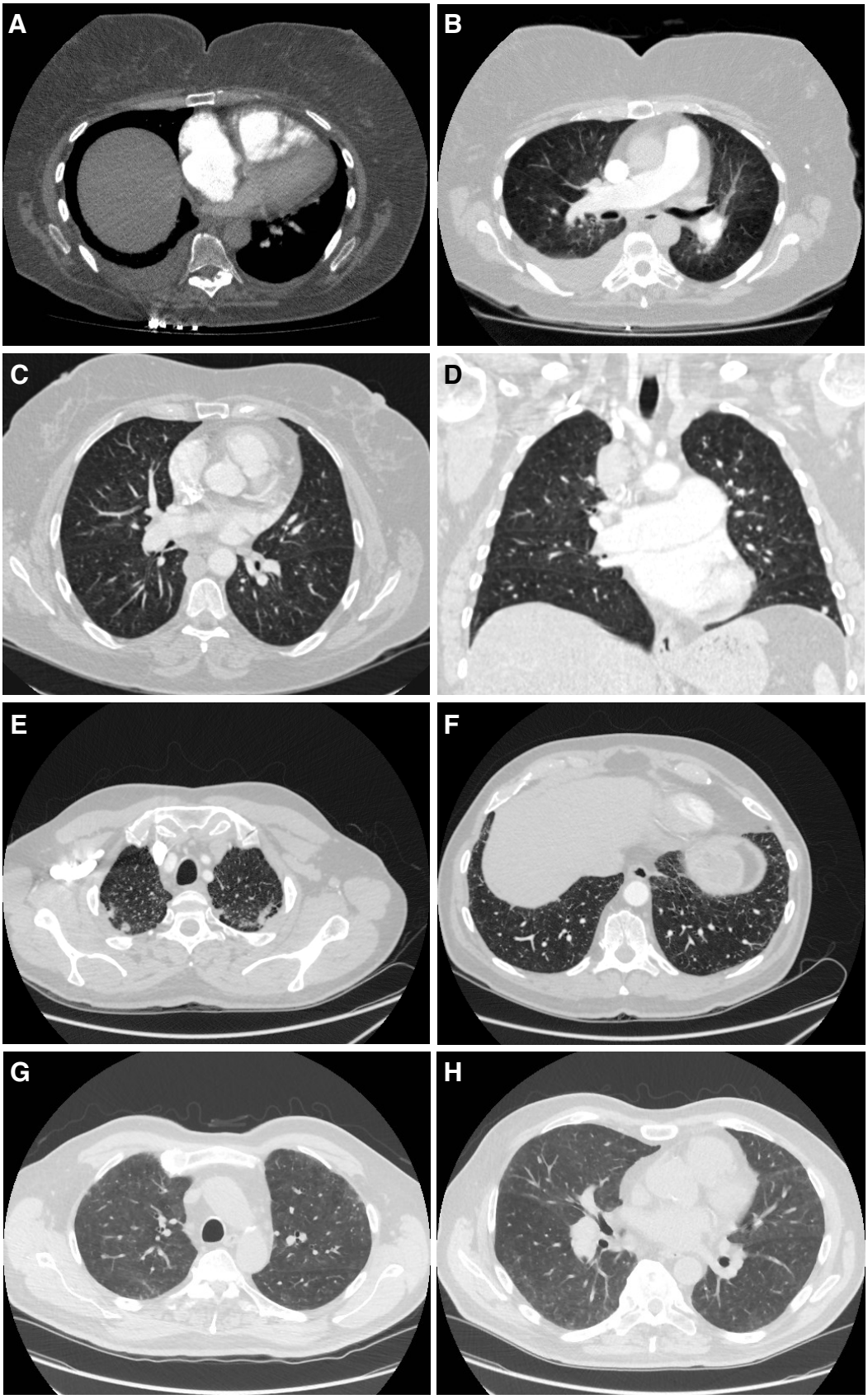


Figure 4. Chest computerized tomography (CT) scans of patients carrying protein-truncating *KDR* mutations. A, Axial image of pulmonary CT angiogram at the level of the right ventricle (RV) moderator band, showing flattening of intraventricular septum, leftwards bowing of the interatrial septum and the enlargement of the right atrium (RA) and RV, indicative of RV strain; bilateral pleural effusion, larger on the right side. **B**, Axial image of a pulmonary CT angiogram demonstrating enlarged pulmonary artery and mild central lung ground glass opacity (GGO). **C**, Axial high-resolution CT slice of the chest in the lung window showing a trace of non-specific GGO with a central distribution. **D**, Coronal image showing the trace of central GGO and enlarged central pulmonary arteries. Axial high-resolution CT slice of the chest in the lung window showing apical subpleural fibrosis (**E**), and very minor subpleural fibrosis at the lung bases (**F**). Axial high-resolution CT slice of the chest in the lung window showing subpleural GGO at apical level (**G**), and mild GGO at mid-thoracic level (**H**). Patients: E001392 (**A**, **B**), E003448 (**C**, **D**), W000229 (**E**, **F**), W000274 (**G** and **H**).

Table 1. Definitions of labels and the number of unrelated cases and controls for genetic association analysis with BeviMed. mPAP - mean pulmonary artery pressure, PH - pulmonary hypertension, PAH - pulmonary arterial hypertension, I/HPAH - Idiopathic/Hereditary Pulmonary Arterial Hypertension, PVOD - Pulmonary veno-occlusive disease, PCH - Pulmonary capillary haemangiomatosis, APAH - Associated Pulmonary Arterial Hypertension, CHD - Congenital Heart Disease, CTD - Connective Heart Disease, PPH, LHD - Left Heart Disease, LD - Lung Disease, CTEPH - Chronic Thromboembolic Pulmonary Hypertension, KCO - transfer coefficient of carbon monoxide.

Tag	Tag description	Cases	Controls	Excluded relatives
PH	Individuals with mPAP > 25 mmHg	1112	9134	2786
PAH	Patients with one of the following diagnoses: IPAH, HPAH, PVOD, PCH, APAH: CHD-PAH, APAH:CTD-PAH, APAH:HIV-PAH, APAH:PH-PAH	1085	9134	2786
I/HPAH	Patients with a clinical diagnosis of IPAH or HPAH	1036	9134	2786
IPAH	Patients with a clinical diagnosis of IPAH	972	9134	2785
HPAH	Patients with a clinical diagnosis of HPAH	67	9136	2779
PVOD/PCH	Patients with a clinical diagnosis of PVOD/PCH	20	9136	2778
I/HPAH/PVOD/PCH	Patients with one of the following diagnoses: IPAH, HPAH, PVOD, PCH	1056	9134	2786
FPAH	Patients with one of the following diagnoses: IPAH, HPAH, PVOD, PCH and a positive family history	80	9136	2781
APAH	Patients with one of the following diagnoses: APAH:CHD-PAH, APAH:CTD-PAH, APAH:HIV-PAH, APAH:PH-PAH	29	9136	2778
APAH: CHD-PAH	Patients with PAH associated with congenital heart disease	17	9136	2778
APAH: CTD-PAH	Patients with PAH associated with connective tissue disease	10	9136	2778
APAH: PoPH	Patients with PAH associated with portopulmonary hypertension	1	9136	2778
APAH: HIV-PAH	Patients with PAH associated with HIV	1	9136	2778
PH-LHD	Patients with pulmonary hypertension associated with left heart disease (Group 2)	7	9136	2778
PH-LD	Patients with pulmonary hypertension associated with lung disease (Group 3)	8	9136	2778
CTEPH	Chronic thromboembolic pulmonary hypertension (Group 4)	6	9136	2778
PH-multifactorial	Multifactorial pulmonary hypertension (Group 5)	6	9136	2778
young age	Lower age tertile (<40.8 years)	378	9136	2785
middle age	Middle age tertile (40.8 - 58.6 years)	376	9134	2779
old age	Higher age tertile (>58.6 years)	355	9136	2778
low KCO	KCO <50% pred.	152	9136	2778
KCO lower tertile	KCO <60% pred.	211	9136	2778
KCO middle tertile	KCO 60-80% pred.	215	9136	2778
KCO higher tertile	KCO >80% pred.	215	9134	2779

Table 2. BeviMed analysis results. Posterior probabilities and Bayes factors (the ratio between the probabilities of the data under H1 and under H0. The observed data are BF times more likely under H1 than under H0, and so the larger the BF, the stronger the support in the data for H1 compared with H0) of gene-tag associations. The "High" category, comprise only variants of “high” impact, including PTVs and large deletions; the Moderate category contains variants of impact “moderate”, including missense variants or consequence “non_coding_transcript_exon_variant”; the combined category Moderate and High, include both respective consequence types.

Gene	Transcript	Tag	Log Bayes Factor	Posterior probability	Consequence type	Mode of inheritance
BMPR2	ENST00000374580	I/HPAH	265.762	1.000	High	dominant
BMPR2	ENST00000374580	PAH	265.639	1.000	High	dominant
BMPR2	ENST00000374580	I HPAH PVOD PCH	263.481	1.000	High	dominant
BMPR2	ENST00000374580	young age	149.576	1.000	Moderate and high	dominant
BMPR2	ENST00000374580	HPAH	149.091	1.000	Moderate and high	dominant
BMPR2	ENST00000374580	FPAH	147.822	1.000	Moderate and high	dominant
BMPR2	ENST00000374580	IPAH	144.582	1.000	High	dominant
BMPR2	ENST00000374580	KCO higher tertile	99.923	1.000	High	dominant
BMPR2	ENST00000374580	middle age	63.119	1.000	Moderate and high	dominant
BMPR2	ENST00000374580	KCO middle tertile	52.706	1.000	Moderate and high	dominant
EIF2AK4	ENST00000263791	lowDLCO	29.741	1.000	Moderate and high	recessive
EIF2AK4	ENST00000263791	KCO lower tertile	26.247	1.000	Moderate and high	recessive
TBX4	ENST00000240335	I/HPAH	23.783	1.000	High	dominant
TBX4	ENST00000240335	I/HPAH/PVOD/PCH	23.549	1.000	High	dominant
TBX4	ENST00000240335	PAH	23.141	1.000	High	dominant
EIF2AK4	ENST00000263791	young age	20.547	1.000	Moderate and high	recessive
TBX4	ENST00000240335	IPAH	19.990	1.000	High	dominant
EIF2AK4	ENST00000263791	I/HPAH/PVOD/PCH	15.718	1.000	Moderate and high	recessive
ACVRL1	ENST00000388922	HPAH	15.501	1.000	Moderate and high	dominant
EIF2AK4	ENST00000263791	PAH	15.407	1.000	Moderate and high	recessive
EIF2AK4	ENST00000263791	PVOD/PCH	14.441	0.999	Moderate and high	recessive
AQP1	ENST00000311813	HPAH	12.075	0.994	Moderate	dominant
EIF2AK4	ENST00000263791	FPAH	11.858	0.993	High	recessive
TBX4	ENST00000240335	young age	11.500	0.990	High	dominant
AQP1	ENST00000311813	I/HPAH	11.466	0.990	Moderate and high	dominant
KDR	ENST00000263923	KCO lower tertile	11.362	0.989	High	dominant
AQP1	ENST00000311813	I/HPAH/PVOD/PCH	11.291	0.988	Moderate and high	dominant
AQP1	ENST00000311813	PAH	11.047	0.984	Moderate and high	dominant
AQP1	ENST00000311813	FPAH	10.023	0.958	Moderate	dominant
KDR	ENST00000263923	old age	9.249	0.912	High	dominant
GDF2	ENST00000249598	I/HPAH	9.091	0.899	Moderate and high	dominant
BMPR2	ENST00000374580	old age	8.913	0.881	High	dominant
GDF2	ENST00000249598	I/HPAH/PVOD/PCH	8.775	0.866	Moderate and high	dominant
SOX17	ENST00000297316	young age	8.554	0.839	Moderate and high	dominant
GDF2	ENST00000249598	PAH	8.478	0.828	Moderate and high	dominant
ATP13A3	ENST00000439040	KCO higher tertile	8.035	0.755	High	dominant
GDF2	ENST00000249598	middle age	7.818	0.713	Moderate and high	dominant
KDR	ENST00000263923	lowDLCO	7.636	0.675	High	dominant
IDH3G	ENST00000217901	KCO middle tertile	7.523	0.649	Moderate	dominant

Table 3. Gene changes for IPAH patients harboring protein-truncating variants (PTV) in the KDR gene and PTV and missense variants in the IDH3G gene. None of the KDR variants has been previously reported in gnomAD, ExAC or internal controls. For KDR HGVSc notations are based on transcript sequence ENST00000263923.4. HGVSp notations are based on amino acid sequence ENSP00000263923.4. None of the patients harbouring PTV in KDR had capillary hemangioma. For IDH3G HGVSc notations are based on transcript sequence ENST00000217901.5, HGVSp notations are based on amino acid sequence ENSP00000217901.5. Protein truncating variants were defined as stop gained, splice acceptor variants or frameshift variants. OC133E is an affected daughter of W000229.
Abbreviations: KDR - Kinase insert domain receptor, IDH3G - Isocitrate dehydrogenase (NAD(+)) 3 non-catalytic subunit gamma, WHO FC - World Health Organisation functional class, 6MWD - 6-minute walk distance, SpO2 - arterial oxygen saturation, mRAP - mean right atrial pressure, mPAP - mean pulmonary artery pressure, mPAWP - mean pulmonary artery wedge pressure, CO - cardiac output, PVR - pulmonary vascular resistance, FEV1 - forced expiratory volume in 1 sec, FVC - forced vital capacity, KCO - transfer factor coefficient for carbon monoxide, HTN - systemic hypertension, CAD - coronary artery disease, OA - osteoarthritis, DM - diabetes mellitus, OSA - obstructive sleep apnea, ASD - atrial septal defect, VSD - ventricular septal defect, *DLCO% predicted

Gene	***KDR***									***IDH3G***						
Cohort	UK					US				UK				US		
ID	W000229	W000229.d	E003448	W000274	E001392	CUMC-JM161	CCHMC-12-190	CCHMC-19-023	CCHMC-27-015	E004149	E004194	E001063	W000031	CCHMC_22-105	JM1415	CCHMC_10-074
Exon	3/30	3/30		22/30	3/30	2/30	3/30	5/30	22/30	1	1/13	12/13	12/13	13/13		4/13
HGVSc	c.183G>A	c.183G>A	c.490-1G>A	c.3064C>T	c.183del	c.161+1G>T	c.303C>A	c.658+1G>A	c.3064C>T	c.1067T>C	c.1037C>T	c.74C>T	c.74C>T	c.1091C>T	c.161+1G>T	c.217G>C
HGVSp	p.Trp61Ter	p.Trp61Ter	-	p.Arg1022Ter	.Trp61CysfsTer16		p.Tyr101Ter		p.Arg1022Ter	p.Met356Thr	p.Thr346Ile	p.Pro25Leu	p.Pro25Leu	p.Pro364Leu		p.Val73Leu
Consequence type	stop gained	stop gained	splice acceptor variant	stop gained	frameshift variant	splice donor variant	stop gained	stop gained	stop gained	missense variant	missense variant	missense variant	missense variant	missense variant	splice donor variant	missense variant
Shared	PAH(1)	PAH(1)	PAH(1)	PAH(1)	PAH(1)	No	No	No	No	PAH(1)	PAH(1)	PAH(2)	PAH(2)	NA	NA	NA
gnomAD	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.68E-05	NA	NA	NA	5.47E-06	2.62E-05	1.09E-05
CADD PHRED v1.:	38	38	26	37	35	26.4	38	24.3	37	23.3	23.3	25.6	12.02	23.3	26.4	22.9
SIFT	-		-	-	-	-	-	-	-	deleterious(0)	deleterious(0.04)	deleterious(0.04)	deleterious(0.04)	deleterious(0.04)		deleterious(0.02)
PolyPhen	-		-	-	-	-	-	-	-	benign(0.297)	benign(0.297)	benign(0.006)	benign(0.006)	benign(0.104)		bening(0.022)
GerpN	5.93	5.93	5.75	5.95	5.93	5.83	5.48	5.8	5.95	5.22	5.22	5.46	5.46	5.18	5.22	4.72
Ancestry	European	European	European	European	European	East-Asian	European	European	European	European	European	European	European	European	European	European
Sex	male	female	female	male	female	female	male	female	female	female	female	female	female	female	female	male
Diagnosis	IPAH	IPAH	IPAH	IPAH	IPAH	APAH-CHD secondary to double outlet RV	IPAH	IPAH	IPAH	IPAH	IPAH	IPAH	IPAH	CHD-PAH	CHD-PAH	CHD-PAH
e at diagnosis [yea	71	53	62	67	61	4	72	65	42	27	34	51	68	0	19	55
WHO FC	2	3	3	3	3	2	NA	NA	NA	3	4	4	2	3	2	3
6MWD [m]	472	202	422	660	180	NA	380	NA	245		414		414	316	364.2	316
SpO_2 pre [%]	95	96	97	98	97	NA	NA	NA	NA	96	95	98	96	NA	98	NA
SpO_2 post [%]	86	87	86	NA	91	NA	NA	NA	NA		99	96	95	NA	99	NA
FEV_1 [% pred.]	116	70	90	83	67.3	85%	NA	77%	NA	87	104	95	99.1	NA	59	NA
FVC [% pred.]	115	76	94	91	72.8	92%	NA	83%	NA	90	109	95.8	96.3	NA	67	NA
TLC [% pred.]	NA	69	NA	NA	NA	NA	NA	65%	NA	NA	105	76	98	NA	85	NA
KCO [% pred.]	44	40	46	46	55.2	NA	NA	35%*	NA	71	64	78	73	NA	NA	NA
Smoking history	Never	Never	Never	Ex-smoker	Never	Never	Never	Ex-smoker	Never	Ex-smoker	Never	Never	Never	Never	Never	NA
mRAP [mmHg]	5	13	8	8	3	NA	5	29	14	14	8	12	6	3	2	7
mPAP [mmHg]	62	45	57	41	44	NA	49	66	60	64	49	50	62	46	37	69
PAWP [mmHg]	4	5	15	12	9	NA	5	16	15	8	10	12	7	NA	5	10
CO [L/min]	3.6	3.3	4.58	5.97	5.23	NA	4.33	1.8	4.6	3.23	NA	3.29	4.1	4.4	6.32	
PVR	16.11	12.12	9.17	4.86	6.69	NA	NA	27.9	9.8	17.3	NA	11.6	13.4	NA	5.2	
Comrobidities	hyperlipidemia, HTN, DM type 2	DM type 2, OSA, pulmonary fibrosis	HTN, hypothyrodism	DM type 2	CAD, DM type 2	No	HTN, hyperlipidemia,	HTN, hypothyrpoidism, OA	Obesity, CAD, DM type 2, hypothyroidism	No	No	PFO	No	Scimitar syndrome, hypoplastic rght lung, ASD with spontaneous closure	VSD	Large ASD
Family history	Yes, daugther	Yes, father	No	No	No	No	No	No	No	No	No	No	No	No	No	NA
Status	dead	alive	alive	alive	dead	alive	alive	alive	alive	alive	alive	dead	alive	alive	alive	alive