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, M.D.[1], Daniel Greene, PhD[2,3], Na Zhu, PhD[4,5], Karyn Megy, PhD[2,3], Marcella Cogliano, M.Sc.[6], Smitha Rajaram, M.D.[7], Divya Pandya, M.Sc.[1], Tobias Tilly, M.Sc.[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, M.St.[3], Carmen M. Treacy, B.Sc.[1], Harm J. Bogaard, M.D., PhD[11], Colin Church, PhD[12], Gerry Coghlan, M.D.[13], Anna W. Coleman, MS[8], Robin Condliffe, M.D.[14], Mélanie Eyries, PhD[15], Henning Gall, M.D., PhD[16], Stefano Ghio, M.D.[17], Simon Holden, PhD[18], Luke Howard, M.D., PhD.[19], David G. Kiely, M.D.[14], Gabor Kovacs, M.D.[20,21], Jim Lordan, PhD[22], Rajiv D. Machado, PhD[10], Robert V. MacKenzie Ross, M.B, B.Chir[23], Shahin Moledina, M.B.Ch.B.[24], David Montani, M.D., PhD[25], Horst Olschewski, M.D.[20,21], Joanna Pepke-Zaba, PhD[26], Christopher J. Rhodes, PhD[19], Werner Seeger, M.D.[16], Florent Soubrier, M.D., PhD[27], Jay Suntharalingam, M.D.[23], Mark R. Toshner, M.D.[1,26], Anton Vonk Noordegraaf, M.D.[11], John Wharton, PhD[19], Stephen John Wort, PhD[28,29], National Cohort Study of Idiopathic and Heritable PAH[30], NIHR Bioresource for Rare Diseases[31], PAH Biobank Enrolling Centers' Investigators[32], Allan Lawrie, PhD[6], Martin R. Wilkins, M.D.[19], Richard C. Trembath, F.R.C.P.[33], Yufeng Shen, PhD[5,34], Chung K. Wendy, M.D.[35], Andrew J. Swift, PhD[6], William C. Nichols, PhD[8,9], Nicholas W. Morrell, M.D.[1,3,18,26], Stefan Gräf, PhD[1,2,3]

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]Addenbrooke's Hospital NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom. [19]Department of Medicine, Imperial College London, London, United Kingdom. [20]Ludwig Boltzmann Institute for Lung Vascular Research, Graz, Austria. [21]Medical University of Graz, Graz, Austria. [22]Freeman Hospital, Newcastle upon Tyne, United Kingdom. [23]Royal United

Hospitals Bath NHS Foundation Trust, Bath, United Kingdom. [24]Great Ormond Street Hospital, London, United Kingdom. [25]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. [26]Royal Papworth Hospital NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom. [27]Département de génétique, hôpital Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, and UMR_S 1166-ICAN, INSERM, UPMC Sorbonne Universités, Paris, France. [28]National Heart & Lung Institute, Imperial College London, London, United Kingdom. [29]Royal Brompton Hospital, London, United Kingdom. [30]www.ipahcohort.com, Cambridge, United Kingdom. [31]University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom. [32]www.pahbiobank.org, Cincinnati, United States. [33]Department of Medical and Molecular Genetics, King's College London, London, United Kingdom. [34]Department of Biomedical Informatics, Columbia University, New York, United States. [35]Columbia University Medical Center, New York, United States.

Corresponding authors:

Dr Stefan Gräf, Ph.D. & Professor Nicholas W. Morrell, M.D.
Department of Medicine, University of Cambridge, Level 5, Cambridge University Hospitals, Box 157, 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: 5013

Abstract

Background

Precision medicine approaches require genotype-phenotype associations that have translational utility and hence can impact disease management and outcomes. To date, approximately one-quarter of patients with pulmonary arterial hypertension harbour rare mutations in disease-causing genes. We hypothesised that integrating deep phenotyping data with whole-genome sequencing data will reveal additional disease variants that are extremely rare and/or have a unique phenotypic signature.

Methods

We analysed 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. In order 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, stratification by age at diagnosis and transfer coefficient of carbon monoxide (KCO).

Results

Protein truncating variants (PTV) in *KDR* were strongly associated with lower KCO tertile (posterior probability (PP)=0.985) and higher age tertile group (PP=0.889). None of the patients harbouring PTV in *KDR* (n=4) had significant parenchymal lung changes that could explain the

reduced KCO. KCO stratification also highlighted an association between Isocitrate Dehydrogenase 3 Gamma (*IDH3G*) and moderately reduced KCO in patients with pulmonary hypertension (PP=0.787). The US PAH Biobank was used to independently assess these findings and identified four additional PAH patients with PTV in *KDR* and two *IDH3G*. We also 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 disease onset, deepening our understanding of the role of VEGF signalling in the pathogenesis of PAH. We also report *IDH3G* as a new PAH risk gene. In addition, we demonstrate that the use of deep clinical phenotyping advances the identification of novel causative rare variants.

Introduction

Pulmonary arterial hypertension is a rare condition characterised by pulmonary vascular narrowing 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 remodelling[1]. More recently, hallmarks of cancer, such as aberrant angiogenesis[2], metabolic reprogramming[3] and resistance to apoptosis[4], have been proposed. 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)[5,6]. It is now established that *BMPR2* mutations are responsible for over 70% of familial cases of PAH (HPAH) and 15-20% of idiopathic cases of PAH (IPAH). Interestingly, the penetrance of *BMPR2* mutations is incomplete, so only a fraction of carriers develop the disease[7]. A smaller proportion (up to 10%) of PAH is caused by mutations in activin-like kinase 1 (*ACVRL1*)[8], endoglin (*ENG*)[9], SMAD family member 9 (*SMAD9*)[10], caveolin-1 (*CAV1*), involved in colocalization of BMP receptors[11], and the potassium channel, *KCNK3*, responsible for membrane potential and vascular tone[12]. Using burden tests, we have recently identified rare pathogenic variants in growth differentiation factor 2 (*GDF2)*, which encodes BMP9, a major ligand for BMPR2, as well as 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[13]. Together, these, and previous findings explain approximately 25% of cases with idiopathic/hereditary pulmonary arterial hypertension (I/HPAH). To further decipher the molecular genetic network of PAH in the remaining 75% of cases, we increased the cohort size and deployed a Bayesian framework incorporating refined phenotype data.

Methods

Study design, ethics, and subject recruitment

The National Institute for Health Research BioResource - Rare Diseases study (NIHRBR-RD), the Rare Disease pilot 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 1148 subjects including individuals diagnosed with either idiopathic or heritable PAH, pulmonary veno-occlusive disease (PVOD) or pulmonary capillary haemangiomatosis (PCH) and a small number of healthy relatives. Adult and paediatric onset cases were eligible, as well as incident and prevalent cases. Recruitment was carried out across the nine PAH specialist centres in the UK and retrospectively by international collaborators at the University of Paris (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 PAH Biobank cohort comprising exome sequencing data from 2572 subjects diagnosed with group 1 PAH[15] and a biobank of 440 PAH patients established at Columbia University Medical Center composed of 29 FPAH, 195 IPAH and 216 APAH individuals[16].

Phenotyping of patients

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

Pseudonymised results of routinely performed clinical tests reported in either clinical case notes or electronic medical records were stored in the *OpenClinica* data capture system. Twenty-one electronic Clinical Case Report Forms (eCRFs) distributed across seven events (diagnostic data, continuous data, follow-up data, epidemiology questionnaire, suspension information, data on relatives and unrelated healthy controls) were constructed to accommodate routinely available clinical information (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 catheterisation. 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 haemodynamic criteria, reported comorbidities (history of pulmonary embolism, chronic obstructive pulmonary disease, interstitial lung disease (ILD), left heart disease, connective tissue disease, structural heart abnormalities, anorexigen use) and results of pulmonary function tests, heart and lung imaging and clinical blood tests (autoantibody screen). Cases in which the diagnosis was questionable were reported back to recruiting centres for verification. Appropriate diagnostic and phenotypic tags were assigned to all recruited patients to be used in the subsequent case-control analysis (Figure S1). The full set of tags, with corresponding numbers of cases, controls and excluded relatives, can be found in Table 1.

Analysis of computerised tomography scans

Diagnostic chest computerised tomography (CT) scans were performed and reported in 613 study participants. The analysis of these scans was done in PAH centres and subsequently transcribed to study eCRFs. Of 613 scans, 294 were available for repeated analysis. The scans were anonymised and transferred to Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, the UK where 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 customised proforma (Table S3).

CT scans were obtained between 2002 and 2018 (n=269), CT pulmonary angiogram (CTPA, n=241), high resolution computed tomography (HRCT no CTPA, n=28). Slice thickness was less than 5mm for all studies, typically $\leq$1mm. Images were analysed on open source software Horos (Annapolis, MD USA). Cardiac and vascular measurements were taken by one observer (MC) and reviewed by the Consultant Radiologist (AS). Thoracic Radiological features were scored semi-quantitatively by two independent Cardiothoracic Radiologist observers each with 9 years experience in pulmonary hypertension imaging (AS, SR) with a very good interobserver agreement (see Supplement, Table S11)

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 a central DNA extraction and QC laboratory in Cambridge (UK).

They were subsequently tested for adequate DNA concentration, DNA degradation and purity. Next-generation paired-end whole-genome sequencing, using three read lengths 100bp (377 samples), 125bp (3,154 samples) and 150bp (9,656 samples), was performed on cases and controls using Illumina HiSeq2500 and HiSeq X (Illumina Inc, San Diego, USA).

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

In order to identify novel genetic associations with subsets of PAH patients defined by selected diagnostic and phenotype features, we deployed the approach outlined in Figure 1A. In brief, phenotype and diagnostic tags were derived from the collected phenotype data (Figure S1). Filtered variants combined with the defined tags were used as input for the Bayesian-based algorithm BeviMed[18], which calculates a posterior probability of genetic association by model comparison for each tag.

This Bayesian inference procedure is based on the comparison of baseline and association models (dominant and recessive). Considering that distinct groups of patients who share a particular characteristic feature may also share a similar genetic aetiology[19], we used the current

diagnostic classification of pulmonary arterial hypertension and stratification of continuous variables such as age at diagnosis or KCO (% predicted) to define a set of phenotypic tags as described above (Table 1).

The case-control analysis was performed with subjects that had assigned a relevant tag as cases, while both cases without the specific tag or missing data were excluded from the analysis (Figure 1C). The individuals from the non-PAH domains served as controls[14].

Variants were extracted from each gene using the rules described in detail in the NIHRBR-RD manuscript[14] including a $PMAF_x$ (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, a CADD Phred score ≥10 and restricting the analysis to the by Ensembl annotated canonical transcript. For each gene-tag pair, BeviMed was applied to the extracted rare variants from a set of unrelated individuals selected to maximise the number of cases[14]. The baseline model assumed fixed disease risk across all study participants. Under the association model, a latent bipartition of rare variants at a gene locus, into pathogenic and non-pathogenic, the ploidy at each individual variant and the mode of inheritance determined the disease risk. Patients which were labelled as "explained" by genotype (based on identified rare deleterious variants in at least one of the previously established PAH disease genes [*BMPR2*, *ACVRL1*, *ENG*, *CAV1*, *SMAD1*, *SMAD4*, *SMAD9*, *KCNK3*, *EIF2AK4*, *TBX4*, *AQP1*, *ATP13A3*, *GDF2*, *SOX17*]) and being deemed disease-causing by a genetic multidisciplinary team (MDT) according to the ACMG standards and guidelines[20], were excluded from the association testing for other genes to reduce the likelihood of false-positive associations.

Importantly, in order to improve power in scenarios where only a specific variant consequence type was associated with the disease risk, association models were fitted to different subsets of variants according to the severity (impact) rating and consequences provided by Ensembl (https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html): 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*, combining the respective consequence types. The prior probability of association across all association models was set to 0.001. Our choice of prior was informed by the estimation that approximately 30 genes might be involved in the pathogenesis of pulmonary arterial hypertension out of the 32,606 protein-coding and non-coding genes (defined by the selected gene biotypes provided by Ensembl, see supplemental material) tested after applying the filtering described above. The association testing was also performed using the variance component test SKAT-O[21] implemented in the R package SKAT (version 1.3.2.1) using default parameters to compare with results generated using BeviMed.

Descriptive statistics

Statistical analysis and data visualisation 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 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 visualise 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 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 weighted (ordinal data) and unweighted (for non-ordinal data), Cohen's Kappa for two readers was calculated and for continuous variables, intraclass correlation coefficient (ICC) was computed with 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 1148 were recruited to the PAH domain. The PAH domain included 23 unaffected parents and 3 cases with unknown phenotype, which were subsequently removed from the analysis (Table S1 and Figure 1B). Of the remaining 1122 participants, 972 (86.6%) had a clinical diagnosis of IPAH, 73 (6.5%) of HPAH, and 20 (1.8%) were diagnosed with PVOD/PCH. Verification of diagnosis based on the collected clinical information revealed that 57 participants (5%) had a diagnosis other than IPAH, HPAH or PVOD/PCH. These cases were subsequently relabelled and used in the analysis (see Table S4 and Table 1). The population structure of the PAH cohort 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). Among the most common comorbidities were hypertension (24%), diabetes mellitus type 2 (12%) and hypothyroidism (12%). Most patients were treated with combination therapies (44%) followed by monotherapy with sildenafil (24%) (Table S4). Overall survival in the studied population was 97% at 1-year, 91% at 3-years and 84% at 5-years. When the cohort was divided into prevalent

and incident cases 1-, 3-, and 5-year survival was 98%, 93%, 87% and 97%, 84%, 72% respectively.

Transfer coefficient of carbon monoxide (KCO) measured at diagnosis was available for 644 patients (57%) (see Supplemental Material, Table S5 and 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 men, older at diagnosis, had a current or past history of cigarette smoking and an increased number of cardiorespiratory comorbidities (Table S6 and S7). Survival in these groups was significantly worse than in those with preserved or mildly reduced KCO (Figure S3 A&B). Even 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 higher age tertile showed more functional impairment despite milder haemodynamics, 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 1122 cases and interpreted them based on the ACMG standards and guidelines[20]. The majority of these variants have already been described in Gräf *et al.*[13] (see supplemental material). The list of comprehensively annotated SNVs and indels is provided in Table S10. Larger deletions are depicted in Figure S4 A-F.

Rare variant association testing

We used the rare variant association tests BeviMed and SKAT-O to consolidate previously reported and discover novel genotype-phenotype associations. The BeviMed analysis identified 41 significant gene-tag associations with posterior probability (PP) above 0.6 (Table 2 and Figure 2A). *BMPR2, TBX4, EIF2AK4, ACVRL1* show the highest association (PP ≥0.98) and further confirmed significant associations in the majority of other previously established genes[13]. Our analysis showed that individuals with rare variants in *BMPR2*, *TBX4*, *EIF2AK4* (autosomal recessive model) and *SOX17* have a significantly earlier age of disease onset (tag: young age). We also demonstrated the association of rare variants in *AQP1* with HPAH (PP=0.625) supported by familial segregation. The refined phenotype approach corroborated the association between mutations in *BMPR2* and preserved KCO (KCO higher tertile, PP=0.889) as well as an association between biallelic *EIF2AK4* mutations and significantly reduced KCO (KCO <50% predicted, PP=1).

Under an autosomal dominant mode of inheritance, protein-truncating variants (PTVs) in kinase insert domain receptor (*KDR)* were associated with a significantly reduced KCO (KCO lower tertile, PP=0.989), as well as older age at diagnosis (tag: old age, PP=0.889). Interestingly, KCO stratification also highlighted an association between Isocitrate Dehydrogenase 3 Gamma (*IDH3G*) and moderately reduced KCO in patients with pulmonary hypertension (PP = 0.787). We were able to confirm these genotype-phenotype associations independently with the alternative variance component test SKAT-O (data not shown).

Rare variants in the new PAH risk genes: *KDR* and *IHD3G*

We identified a total of five rare protein-truncating variants in *KDR* in the study cohort, four 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) as well as one nonsense variant in exon 27 (p.Glu1206Ter) in a non-PAH control (Table 3). Although this nonsense variant only appears very late in the amino acid chain, is not located in the last exon, thus the resulting modified mRNA sequence is likely to be subject to nonsense-mediated decay. Furthermore, 13 PAH cases (1%) and 102 non-PAH controls (0.9%) harboured 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 seemed to show the opposite trend in KCO (Table 4 and Figure 2 C and D). Importantly, seven of the 13 *KDR* missense variants seen in the PAH cases also were detected in several non-PAH controls. Furthermore, three of the *KDR* missense variants co-occurred with predicted deleterious variants in established PAH risk genes (Table 13).

We also identified three missense variants (c.74C>T, p.Pro25Leu; c.1037C>T, p.Thr346Ile; c.1067T>C, p.Met356Thr) and one large deletion (X:147511939-154854072) in five individuals in the gene encoding isocitrate dehydrogenase subunit gamma (*IDH3G*). The missense variant (c.74C>T, p.Pro25Leu) was present in two IPAH individuals, whereas the large deletion (X:147511939-154854072) was present in one IPAH and one control case. The “Moderate and high” impact category contributed to the detected association. IPAH patients harbouring variants in *IDH3G* were all females with early-onset disease and relatively preserved KCO.

Additionally, two individuals carrying missense variants in *IDH3G* locus were found in US PAH Biobank and Pulmonary Hypertension Center at Columbia University cohorts; one male neonate diagnosed with Scimitar syndrome, hypoplastic right lung and ASD (c.1091C>T, p.Pro364Leu) and a 55-year-old female with large ASD (c.217G>C, p.Val73Leu).

Clinical characterisation of *KDR* mutation carriers

Patients with PTV in *KDR* were older and exhibited significantly reduced KCO when compared with *KDR* missense variant carriers and *BMPR2* mutation carriers (Figure 2C). In order to exclude that the reduction in KCO is the result of coexistent emphysema secondary to smoking or other parenchymal lung diseases, we performed a detailed analysis of imaging studies. Three of the four patients did not have a history of smoking. The CT scans in two out of four patients showed mild parenchymal changes that could account for some of the reduction in KCO but not fully explain it. Two of the four patients carrying a PTV in *KDR* presented with mild centrilobular ground glass opacities (GGO) that are commonly shown in PAH[22], and one had a trace and one a mild non-specific GGO both centrally distributed. Two of the four patients harbouring PTV in *KDR* had mild fibrotic lung changes, whereas the other groups showed less than 10% incidence of fibrotic changes. None of the patients had emphysema, but three showed air trapping (a trace in one patient and mild in two patients). There were no signs of intralobular septal thickening but mediastinal lymphadenopathy was seen in three individuals. Comparisons between patients harbouring deleterious mutations in *BMPR2*, *EIF2AK4*, *KDR,* other PAH risk genes and patients without mutations are presented in Table S11. In summary, there were no major differences between groups, but patients with *KDR* PTVs had significantly less mediastinal lymphadenopathy than patients harbouring deleterious variants in *BMPR2* or other PAH risk genes. Of note, patients with *BMPR2* mutations had the largest bronchial arteries. There were no differences in the frequency of comorbidities between patients harbouring missense and PTV in *KDR* although the frequency of systemic hypertension was high in both groups (44 and 50%, respectively) (Table 4 and Table S12). None of the PTV carriers had a family history of PAH. Survival in this group could not be assessed because of the small number of patients harbouring the mutation, as well as only one event occurring in this group.

Additional cases in the US PAH Biobank and Pulmonary Hypertension Center at Columbia University cohorts

To replicate our findings, we used patients recruited to the US PAH Biobank[15] and the Pulmonary Hypertension Center at Columbia University[16] to identify patients carrying predicted pathogenic rare variants in the new PAH risk genes. Four individuals harbouring *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 appears in both cohorts (Figure 3). Patient-level data for these individuals are summarised in Table S13. 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. Diffusion capacity of carbon monoxide was available for one patient and was significantly decreased at 35% predicted, and only minor pleural scarring in the left upper lobe was found in this individual. Two out of four patients harbouring PTV in *KDR* had also been diagnosed with systemic hypertension.

Discussion

One of the critical translational steps in identifying novel, causative genes in rare disorders is the discovery of genotype-phenotype associations to inform patient care and impact outcomes. A pragmatic focus on deeply-phenotyped individuals and “smart” experimental design cannot be overestimated[23]. With this in mind, we continued to study the molecular genetic architecture of PAH using the Bayesian approach BeviMed[18]. 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 rare gene variants.

Our findings strongly suggest a link between rare protein-truncating KDR variants and significantly reduced KCO and older age at diagnosis. The human KDR, located on chromosome 4q11–q12, encodes vascular endothelial growth factor receptor 2 (VEGFR-2)[24]. VEGFR-2 is composed of an extracellular domain, which comprises seven Ig-like domains (I–VII), of which domains II and III bind VEGF-A, a critical growth factor for physiological and pathological angiogenesis in vascular endothelial cells. In mice, even though VegfA haploinsufficiency is embryonically lethal[25], heterozygosity of its receptor, Vegfr2, is compatible with life and unimpaired vascular development[26].

The role of VEGF signalling in the pathogenesis of PAH has been a matter of research since the reports of increased expression of VEGF, VEGFR1 and VEGFR2 in rat lung tissue in response to acute and chronic hypoxia[27]. An increase in lung VEGF has also been reported in rats with PH following monocrotaline exposure[28]. In humans, VEGFA is highly expressed in plexiform lesions in patients with IPAH[29], tracheal aspirates from neonates with a persistent PH of the newborn[30] and small pulmonary arteries from infants with PH associated with a congenital diaphragmatic hernia[31]. In view of these findings, it is surprising that the overexpression of VEGFA ameliorates hypoxia-induced PAH[32]. In contrast, inhibition of VEGF signalling by SU5416 (sugen) combined with chronic hypoxia triggers severe angioproliferative PH[33]. SU5416, a small-molecule inhibitor of the tyrosine kinase segment of VEGF receptors inhibits VEGFR1[34] and VEGFR2[35] causing endothelial cell apoptosis, loss of lung capillaries and emphysema[36]. In combination with chronic hypoxia, SU5416 causes cell-death dependent compensatory pulmonary endothelial cell proliferation and severe PH[33]. Interestingly, sugen in combination with other stimuli such as immune insufficiency[37] or overexpression of HIF-1α[38] also leads to severe PH in rats. Further evidence supporting the role of VEGF inhibition in the pathobiology of PAH comes from reports

of PH in patients treated with bevacizumab[39] and the multi-tyrosine kinase inhibitors, dasatinib[40] and bosutinib, have also been associated with PAH[41]. Both preclinical and patient data show that inhibition of VEGF is associated with considerable cardiovascular side effects[42]. Among common side effects of VEGF inhibitors are systemic hypertension (HTN), proteinuria, renal impairment and thyroid dysfunction. The overall incidence of HTN induced by bevacizumab and RTKIs scale from 9 to 67% and is dose-dependent[43]. Mechanisms implicated in HTN include impairment of nitric oxide (NO) signalling, increased arterial stiffness[44], reduced capillary density[45] or functional rarefaction[46] and activation of the endothelin system[47], all of which are relevant to the pathobiology of PAH. Notably, two out of four of our cases with PTVs at the KDR locus had systemic hypertension, also the frequency of thyroid dysfunction seemed to be higher (although not statistically significant) in patients with KDR PTVs (25% UK cohort, 50% US cohort) than in patients without mutations in PAH risk genes (13.2%). The proportion of patients with renal impairment was not different between KDR PTV and missense variant carriers or the rest of the study population. Mutations in KDR were also reported in other cardiovascular diseases; Bleyl et al. reported that KDR might be a candidate for familial total anomalous pulmonary venous return[48]. Besides, haploinsufficiency in KDR locus has also been associated with tetralogy of Fallot[49]. We report one patient (US cohort) with PAH associated with congenital heart disease and KDR protein-truncating splice donor variant (c.161+1G>T). The impact of these variants on congenital heart malformations remains to be elucidated but previous research indicates that Flk1+ cells contribute to normal development, capillarity and metabolism of both cardiac and skeletal muscle[50,51].

Isocitrate dehydrogenase (NAD(+)) 3 non-catalytic subunit gamma (IDH3G) is a protein-coding gene encoding enzyme catalyzing the decarboxylation of isocitrate (ICT) into alpha-ketoglutarate, a tricarboxylic acid (TCA) cycle intermediate. Metabolomic[3] and imaging studies[52] have previously shown disrupted bioenergetics in IPAH characterised by the accumulation of

TCA cycle intermediates. This indicates suppression of mitochondrial glucose oxidation, central to which is inhibition of pyruvate dehydrogenase (PDH)[53]. Alpha-ketoglutarate is a required cofactor for PHD, the enzyme that under normal conditions causes proteasomal degradation of hypoxia-inducible factor (HIF)[54]. Citrate and alpha-ketoglutarate have also been implicated in acetylation[55] and methylation[56] of nuclear histones. Interestingly IDH activity has been reported to be increased both in PAEC and serum in patients harbouring BMPR2 pathogenic variants[57]. IDH has the capacity to catalyze against TCA flow so to convert alpha-ketoglutarate to isocitrate leading to depletion of PHD co-factor alpha-ketoglutarate and causing decreased hydroxylation of HIF necessary for its proteasomal degradation[57]. Those findings have potential therapeutic implications, as pyruvate dehydrogenase kinase inhibitor (dichloroacetate) has shown some efficacy in genetically susceptible PAH patients[58].

With this study, we highlight that deep clinical phenotyping in combination with genotype data can accelerate the identification of novel disease risk genes and disease subtypes, which may have prognostic and therapeutic implications. 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 registry data[59] to be an independent predictor of survival. Decreased KCO was also found in patients with PVOD/PCH with or without biallelic EIF2AK4 mutations[60]. Although some reduction in KCO is one of the typical features of PH, PVOD patients show the lowest KCO values when compared to IPAH or CTEPH. In contrast, KCO is relatively preserved in BMPR2 mutation carriers[61]. Strong association with survival and a link with other causative mutations makes the KCO phenotype particularly attractive for genetic studies, and KCO should be consistently collected in future PAH registries.

As lung disease should always be taken under consideration as a cause of low KCO, we applied the World Symposium on PH criteria[62] 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 PTV KDR 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[63]. 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[64] and smoking-induced emphysema[65]; 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 HRCT. After excluding known causes of significantly reduced KCO, one can hypothesize that PTVs in the KDR locus leads to severe angioproliferative obstruction of small capillaries and subsequent decreased capillary blood volume available for gas exchange. An alternative explanation could be that PTVs in KDR are associated with the development of ILD. The latter hypothesis can be indirectly supported by the high percentage of air trapping seen in these patients (75%); small airway obstruction has been previously reported in ILD[66]. Further studies are needed to determine the contribution of lung capillary volume and alveolar-capillary membrane diffusing capacity to the overall diffusing capacity in patients with PTVs in KDR.

Recent registries have shown a considerable shift in PAH demographics[59,67]. Particularly in ageing western populations, PAH is now diagnosed in older patients, with a significant burden of comorbidities, a weaker response to treatment and poorer survival[68]. Although genetic disorders tend to present earlier in life, better phenotypic and genetic characterisation of older patients is required as this group now constitutes the majority of the adult PAH population. The occurrence of PAH at a relatively old age in KDR PTV carriers may be indicative of the necessity of a second hit (similarly to what is seen in animal sugen-hypoxia PH model) for the development of the disease. Although not identified in our study, such a hit might be an environmental

exposure or age-related accumulation of somatic mutations. The latter concept has recently gained traction. It is estimated that new somatic mutations occur at rate of 40 per year per cell across various tissues[69] and they vary in number from <500 in newborns to >3000 per cell in centenarians contributing to various age-related diseases including cancer[69,70].

In our study deep phenotyping enabled patient stratification into subgroups with shared pathobiology and therefore increased 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. Based on in silico analysis we showed that the associated variants were predicted to be deleterious while occurring at highly conserved positions. Finally, we performed an in-depth literature review supporting the functional importance of these genes in the pathogenesis of PAH.

Acknowledgements

We thank NIHR BioResource volunteers for their participation, and gratefully acknowledge NIHR BioResource centres, 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 centres involved in this study. The UK National Cohort of Idiopathic and Heritable PAH is supported by the NIHRBR-RD, the British Heart Foundation (BHF) (SP/12/12/29836), 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). 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 Organisation 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. 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. 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

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. Thomson J, Machado R, Pauciulo M, Morgan N, Yacoub M, Corris P, McNeil K, Loyd J, Nichols W, Trembath R. Familial and sporadic primary pulmonary hypertension is caused by BMPR2 gene mutations resulting in haploinsufficiency of the bone morphogenetic protein tùype II receptor. *J Heart Lung Transplant*. 2001;20:149.

6. 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.

7. 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.

8. 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.

9. 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

10. 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

11. 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.

12. 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

13. 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.

14. The NIHR BioResource, on behalf of the 100,000 Genomes Project. Whole-genome sequencing of rare disease patients in a national healthcare system. *Nature* [Internet]. Available from: https://www.biorxiv.org/content/10.1101/507244v1

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. 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.

19. Lentaigne C, Greene D, Sivapalaratnam S, Favier R, Seyres D, Thys C, Grassi L, Mangles S, Sibson K, Stubbs MJ, Burden F, Bordet J-C, Armari-Alla C, Erber W, Farrow S, Gleadall N, Gomez K, Megy K, Papadia S, Penkett CJ, Sims MC, Stefanucci L, Stephens JC, Read RJ, Stirrups KE, Ouwehand WH, Laffan MA, Frontini M, Freson K, Turro E. Germline mutations in the transcription factor IKZF5 cause thrombocytopenia. *Blood* [Internet]. 2019;Available from: http://dx.doi.org/10.1182/blood.2019000782

20. 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.

21. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011;89:82–93.

22. Rajaram S, Swift AJ, Wild JM, Kiely DG. Response to: “CT assessment for pulmonary hypertension requires systematic assessment of cardiac, vascular and parenchymal signs” by Marloes et al. Thorax. 2015;70:1087–1088.

23. 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

24. 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.

25. 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.

26. 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

27. 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.

28. 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.

29. 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

30. Lassus P, Turanlahti M, Heikkilä P, Andersson LC, Nupponen I, Sarnesto A, Andersson S. Pulmonary Vascular Endothelial Growth Factor and Flt-1 in Fetuses, in Acute and Chronic Lung Disease, and in Persistent Pulmonary Hypertension of the Newborn [Internet]. American Journal of Respiratory and Critical Care Medicine. 2001;164:1981–1987. Available from: http://dx.doi.org/10.1164/ajrccm.164.10.2012036

31. Shehata SM, Mooi WJ, Okazaki T, El-Banna I, Sharma HS, Tibboel D. Enhanced expression of vascular endothelial growth factor in lungs of newborn infants with congenital diaphragmatic hernia and pulmonary hypertension. *Thorax*. 1999;54:427–431.

32. Partovian C, Adnot S, Raffestin B, Louzier V, Levame M, Mavier IM, Lemarchand P, Eddahibi S. Adenovirus-mediated lung vascular endothelial growth factor overexpression protects against hypoxic pulmonary hypertension in rats. *Am J Respir Cell Mol Biol*. 2000;23:762–771.

33. 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.

34. 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.

35. 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.

36. 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.

37. Tamosiuniene R, Nicolls MR. Regulatory T cells and pulmonary hypertension. *Trends Cardiovasc Med*. 2011;21:166–171.

38. Mizuno S, Farkas L, Al Husseini A, Farkas D, Gomez-Arroyo J, Kraskauskas D, Nicolls MR, Cool CD, Bogaard HJ, Voelkel NF. Severe Pulmonary Arterial Hypertension Induced by SU5416 and Ovalbumin Immunization [Internet]. American Journal of Respiratory Cell and Molecular Biology. 2012;47:679–687. Available from: http://dx.doi.org/10.1165/rcmb.2012-0077oc

39. 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

40. 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

41. Ashraf O, Naddour M, Babar L, Beg M, Malik K. SECOND GENERATION TYROSINE KINASE INHIBITOR BOSUTINIB AS NOVEL CAUSE OF SEVERE PULMONARY HYPERTENSION: UNMASKING THE CULPRIT [Internet]. Chest. 2018;154:1047A. Available from: http://dx.doi.org/10.1016/j.chest.2018.08.945

42. Kamba T, McDonald DM. Mechanisms of adverse effects of anti-VEGF therapy for cancer. *Br J Cancer*. 2007;96:1788–1795.

43. Zheng Y, Chen J, Lu Y. Incidence and risk of hypertension with bevacizumab in non-small-cell lung cancer patients: a meta-analysis of randomized controlled trials [Internet]. Drug Design, Development and Therapy. 2015;4751. Available from: http://dx.doi.org/10.2147/dddt.s87258

44. Steeghs N, Gelderblom H, Roodt JO 't, Christensen O, Rajagopalan P, Hovens M, Putter H, Rabelink TJ, de Koning E. Hypertension and rarefaction during treatment with telatinib, a small molecule angiogenesis inhibitor. *Clin Cancer Res*. 2008;14:3470–3476.

45. van den Meiracker AH, Danser AHJ, Sleijfer S, Kappers MHW. Re: Hypertension as a Biomarker of Efficacy in Patients With Metastatic Renal Cell Carcinoma Treated With Sunitinib [Internet]. JNCI Journal of the National Cancer Institute. 2011;103:1557–1557. Available from: http://dx.doi.org/10.1093/jnci/djr328

46. Levy BI, Ambrosio G, Pries AR, Struijker-Boudier HAJ. Microcirculation in Hypertension [Internet]. Circulation. 2001;104:735–740. Available from: http://dx.doi.org/10.1161/hc3101.091158

47. Kappers MHW, van Esch JHM, Sluiter W, Sleijfer S, Jan Danser AH, van den Meiracker AH. Hypertension Induced by the Tyrosine Kinase Inhibitor Sunitinib Is Associated With Increased Circulating Endothelin-1 Levels [Internet]. Hypertension. 2010;56:675–681. Available from: http://dx.doi.org/10.1161/hypertensionaha.109.149690

48. 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.

49. 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.

50. Motoike T, Markham DW, Rossant J, Sato TN. Evidence for novel fate of Flk1+ progenitor:

contribution to muscle lineage. *Genesis*. 2003;35:153–159.

51. Baum O, Jentsch L, Odriozola A, Tschanz SA, Olfert IM. Ultrastructure of Skeletal Muscles in Mice Lacking Muscle-Specific VEGF Expression. *Anat Rec* . 2017;300:2239–2249.

52. 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.

53. 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

54. Salceda S, Caro J. Hypoxia-inducible Factor 1α (HIF-1α) Protein Is Rapidly Degraded by the Ubiquitin-Proteasome System under Normoxic Conditions [Internet]. Journal of Biological Chemistry. 1997;272:22642–22647. Available from: http://dx.doi.org/10.1074/jbc.272.36.22642

55. Wellen KE, Hatzivassiliou G, Sachdeva UM, Bui TV, Cross JR, Thompson CB. ATP-citrate lyase links cellular metabolism to histone acetylation. *Science*. 2009;324:1076–1080.

56. Johnson D. Cell Death Signaling in Cancer Biology and Treatment. Springer Science & Business Media; 2012.

57. 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.

58. Michelakis ED, Gurtu V, Webster L, Barnes G, Watson G, Howard L, Cupitt J, Paterson I, Thompson RB, Chow K, O'Regan DP, Zhao L, Wharton J, Kiely DG, Kinnaird A, Boukouris AE, White C, Nagendran J, Freed DH, Wort SJ, Gibbs JSR, Wilkins MR. Inhibition of pyruvate dehydrogenase kinase improves pulmonary arterial hypertension in genetically susceptible patients. *Sci Transl Med* [Internet]. 2017;9. Available from: http://dx.doi.org/10.1126/scitranslmed.aao4583

59. 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

60. 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.

61. 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.

62. 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

63. 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.

64. 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.

65. 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

66. Mello GCF de, de Mello GCF, Carvalho CRR, Kairalla RA, Saldiva PHN, Fernezlian S, Silva LFF, Dolhnikoff M, Mauad T. Small Airway Remodeling in Idiopathic Interstitial Pneumonias: A Pathological Study [Internet]. Respiration. 2010;79:322–332. Available from: http://dx.doi.org/10.1159/000235722

67. 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.

68. 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.

69. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, Nijman IJ, Martincorena I, Mokry M, Wiegerinck CL, Middendorp S, Sato T, Schwank G, Nieuwenhuis EES, Verstegen MMA, van der Laan LJW, de Jonge J, IJzermans JNM, Vries RG, van de Wetering M, Stratton MR, Clevers H, Cuppen E, van Boxtel R. Tissue-specific mutation accumulation in human adult stem cells during life [Internet]. Nature. 2016;538:260–264. Available from:

http://dx.doi.org/10.1038/nature19768

70. Zhang L, Dong X, Lee M, Maslov AY, Wang T, Vijg J. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan [Internet]. Available from: http://dx.doi.org/10.1101/535906

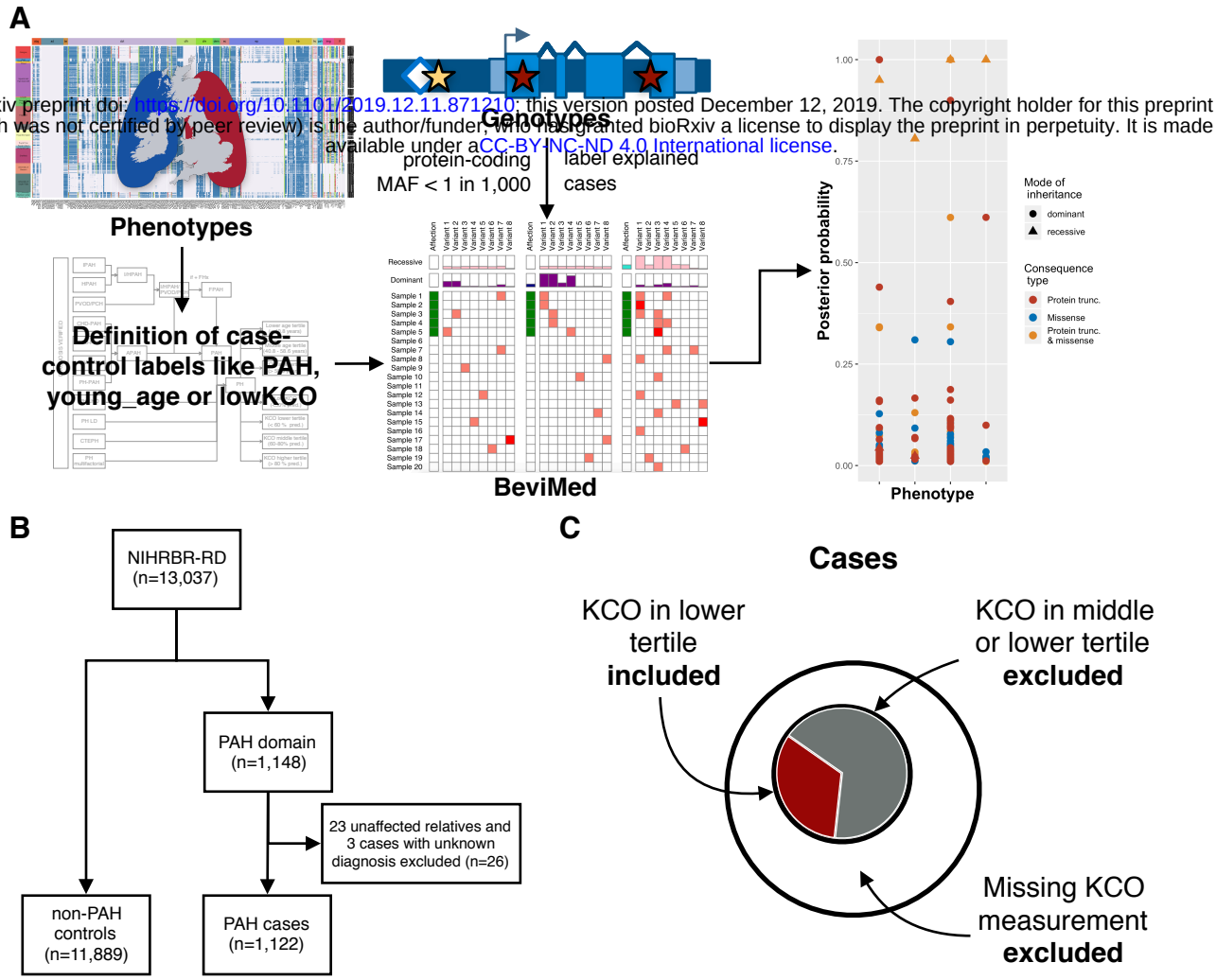


Figure 1. Design of the genetic association study. A. Overview of the analysis approach. Using deep phenotyping data tags were assigned to sub sets of patients with shared diagnostic and phenotypic features (see Figure S1 for more details). Rare sequence variants called from whole genome sequencing data were filtered and explained cases were labelled. These data served as input to BeviMed in order to estimate genome-wide the posterior probability of gene loci being associated with any of the given tags. B. Consort diagram summarising the size of the study cohort. C. Schematic representation of the definition of cases to account for missing data and minimise the likelihood of false positive associations exemplified by the KCO lower tertile tag.

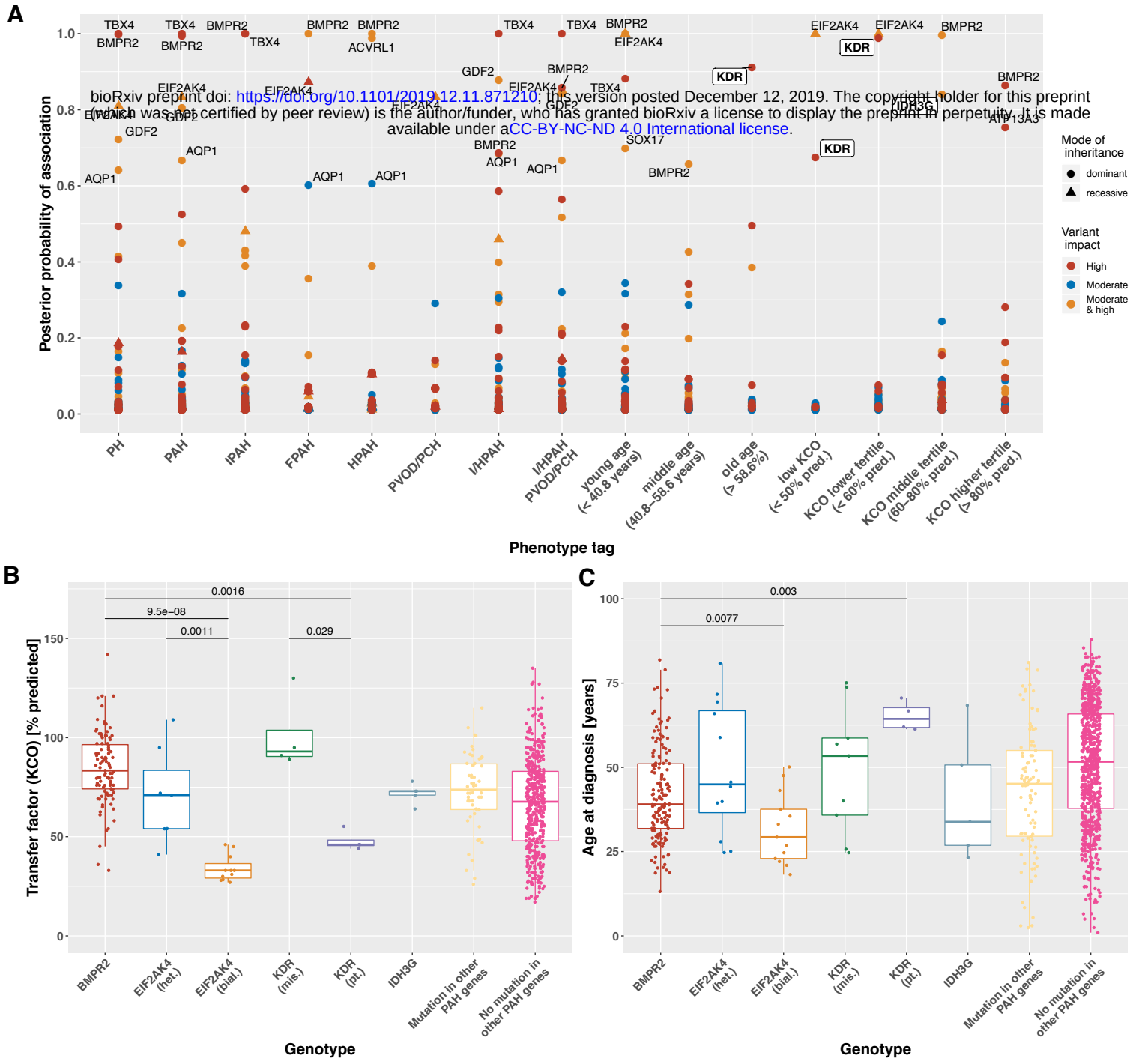
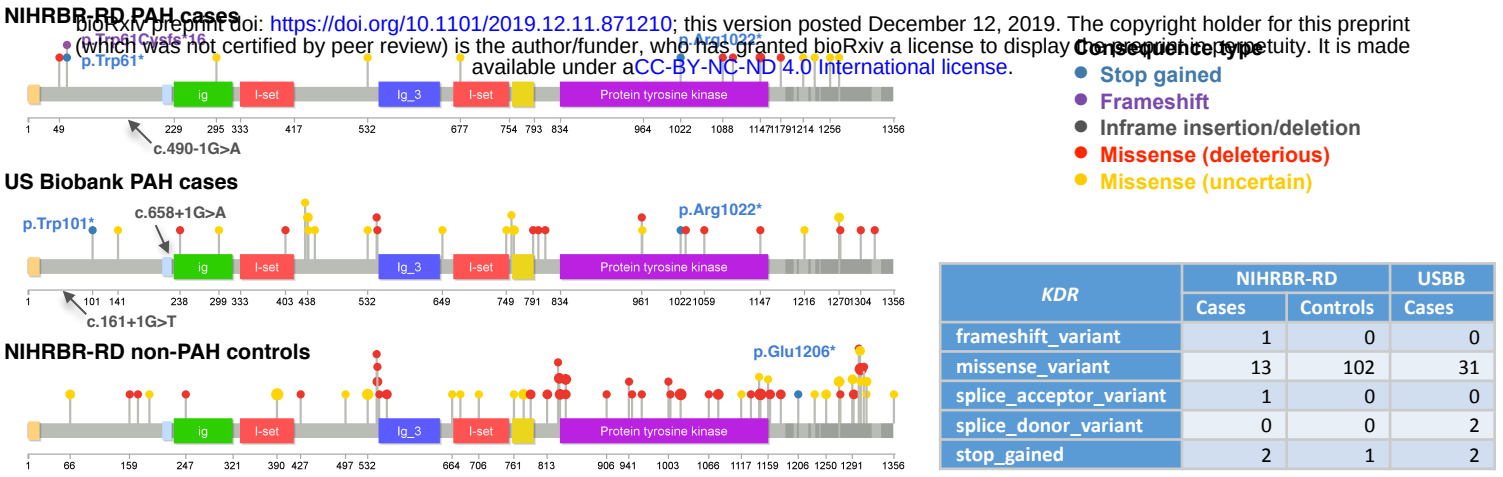


Figure 2. Genetic association study results revealing established and novel genotype-phenotype links. (A) With BeviMed estimated posterior probability of rare predicted deleterious variants in a given gene being associated with a diagnostic or phenotypic tag for both. Shape and colour of points indicate mode of inheritance and consequence type of variants driving the association. Box-and-whisker plots showing the distribution of (B) transfer factor and (C) age at diagnosis stratified by genotype across the PAH domain. Significant differences in the means of the distributions are indicated by the bars at the top of the figures providing the respective p-values.

A

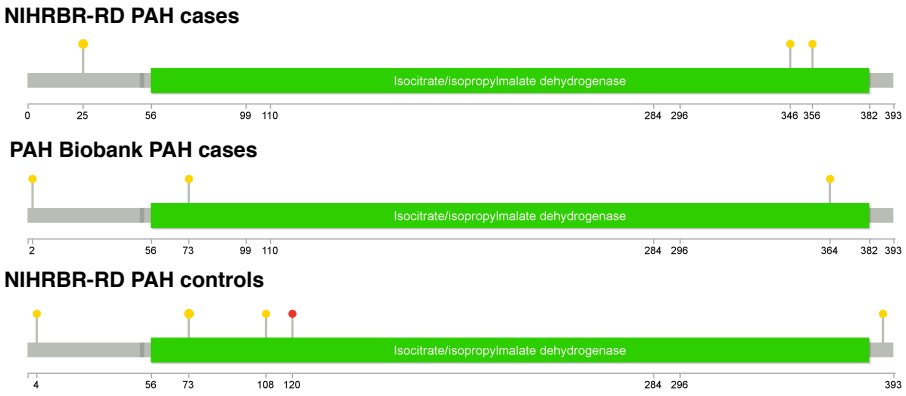
KDR SNVs and indels



KDR	NIHRBR-RD		USBB
	Cases	Controls	Cases
frameshift_variant	1	0	0
missense_variant	13	102	31
splice_acceptor_variant	1	0	0
splice_donor_variant	0	0	2
stop_gained	2	1	2

B

IDH3G SNVs and indels



IDH3G	NIHRBR-RD		USBB
	Cases	Controls	Cases
large_deletion	1	2	0
missense_variant	4	6	3

C

IDH3G large deletion

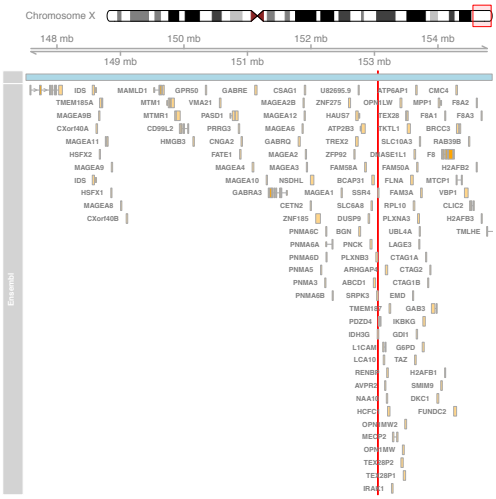


Figure 3. Summary of single nucleotide variants (SNVs), small insertions and deletions (indels) and large deletions identified in the two novel candidate PAH disease risk genes *KDR* (A) and *IDH3G* (B, C) after filtering (MAF < 1/10,000 and CADD > 15). SNVs and indels are represented by according to their consequence type coloured lollipops on top of the protein sequence with domain annotations retrieved from Uniprot (accession numbers P35968 (*KDR* (A)) and P51553 (*IDH3G* (D))). The size of the circles represents how often this variant occurs. PTVs are labeled with the respective HGVS notation. Splice variants are marked by dark grey arrows. The large deletion identified in *IDH3G* (C) is depicted in light blue, the respective gene locus is highlighted in red.

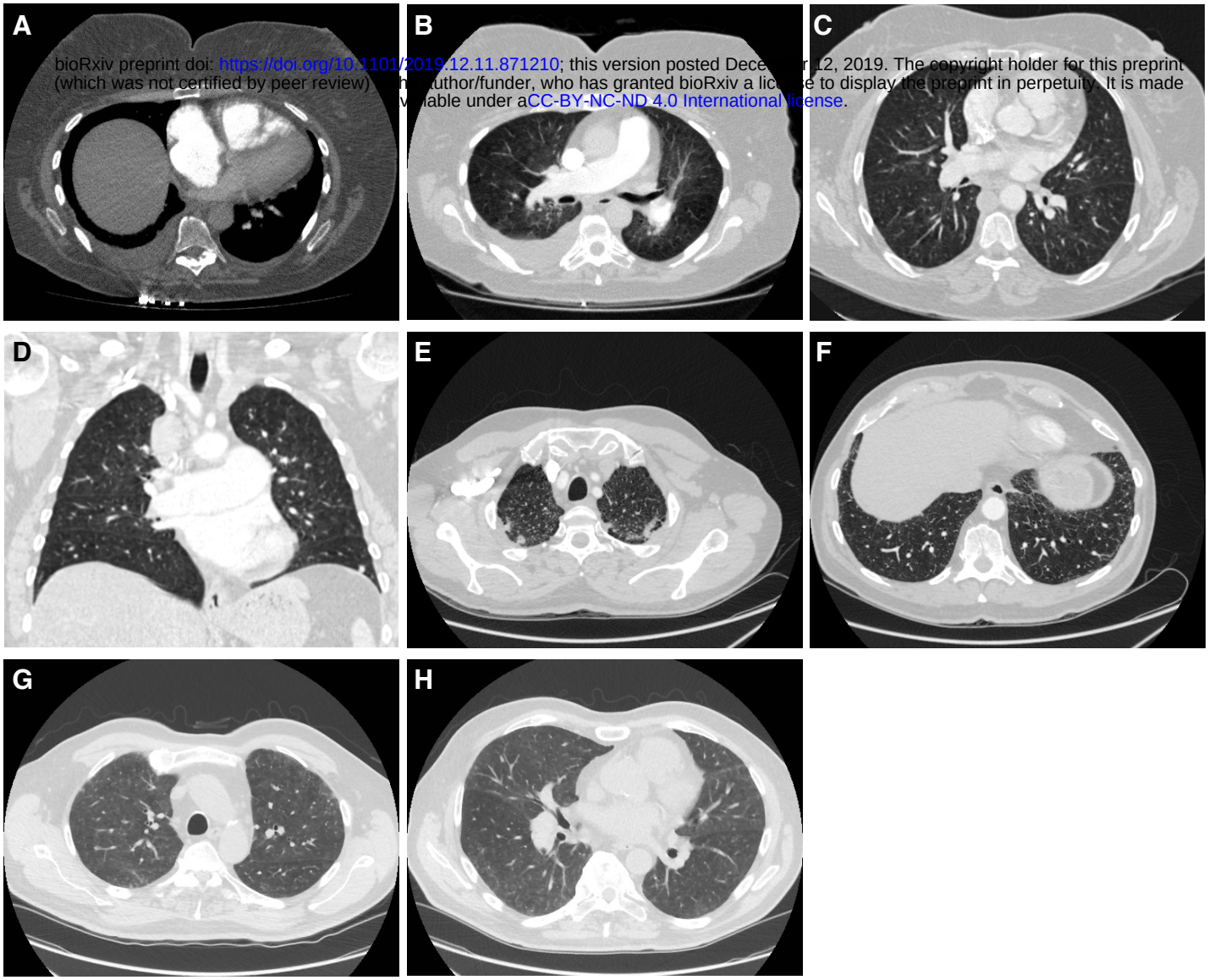


Figure 4. Pulmonary computerised tomography (CT) scans of patients carrying protein-truncating *KDR* mutations. (A) Axial image of pulmonary CT angiogram at the level of the right ventricle (RV) moderator band showing flattening of intraventricular septum, leftwards bowing of the interatrial septum and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

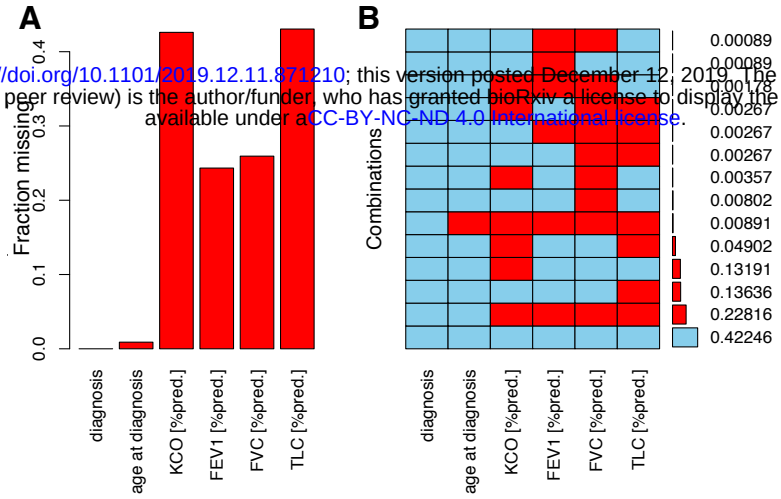


Figure S1. Summary of missing data. The missing rate (A) and missing pattern (B) in KCO in relation to missingness in diagnosis, age at diagnosis and other lung function tests (FEV1: forced expiratory volume in 1st, FVC: forced vital capacity, TLC: total lung capacity).

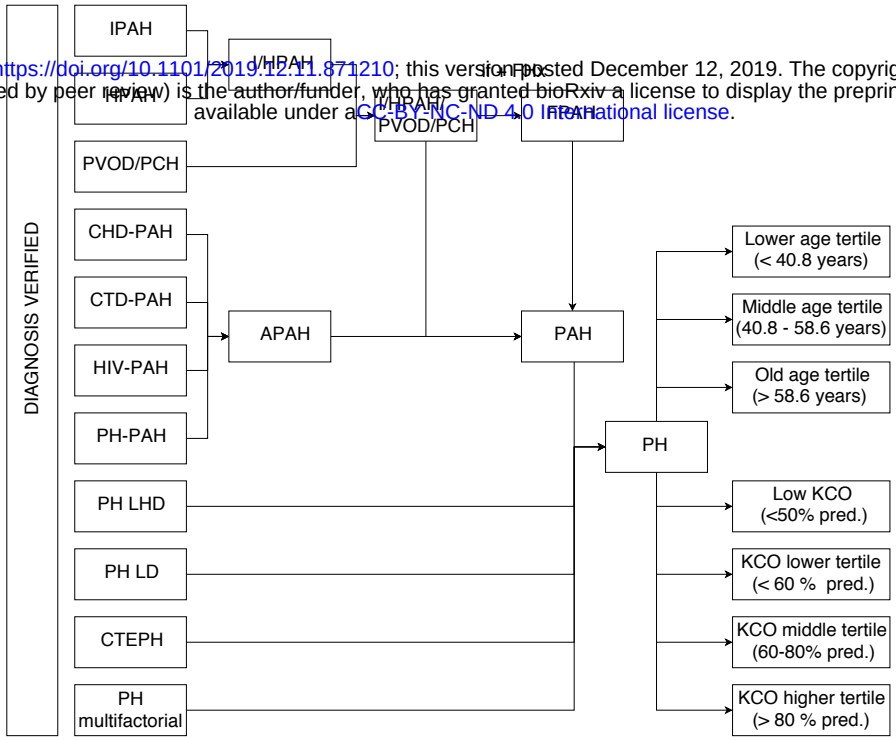


Figure S2. Flowchart describing the definition of diagnostic and phenotypic tags. For detailed description see supplementary information.

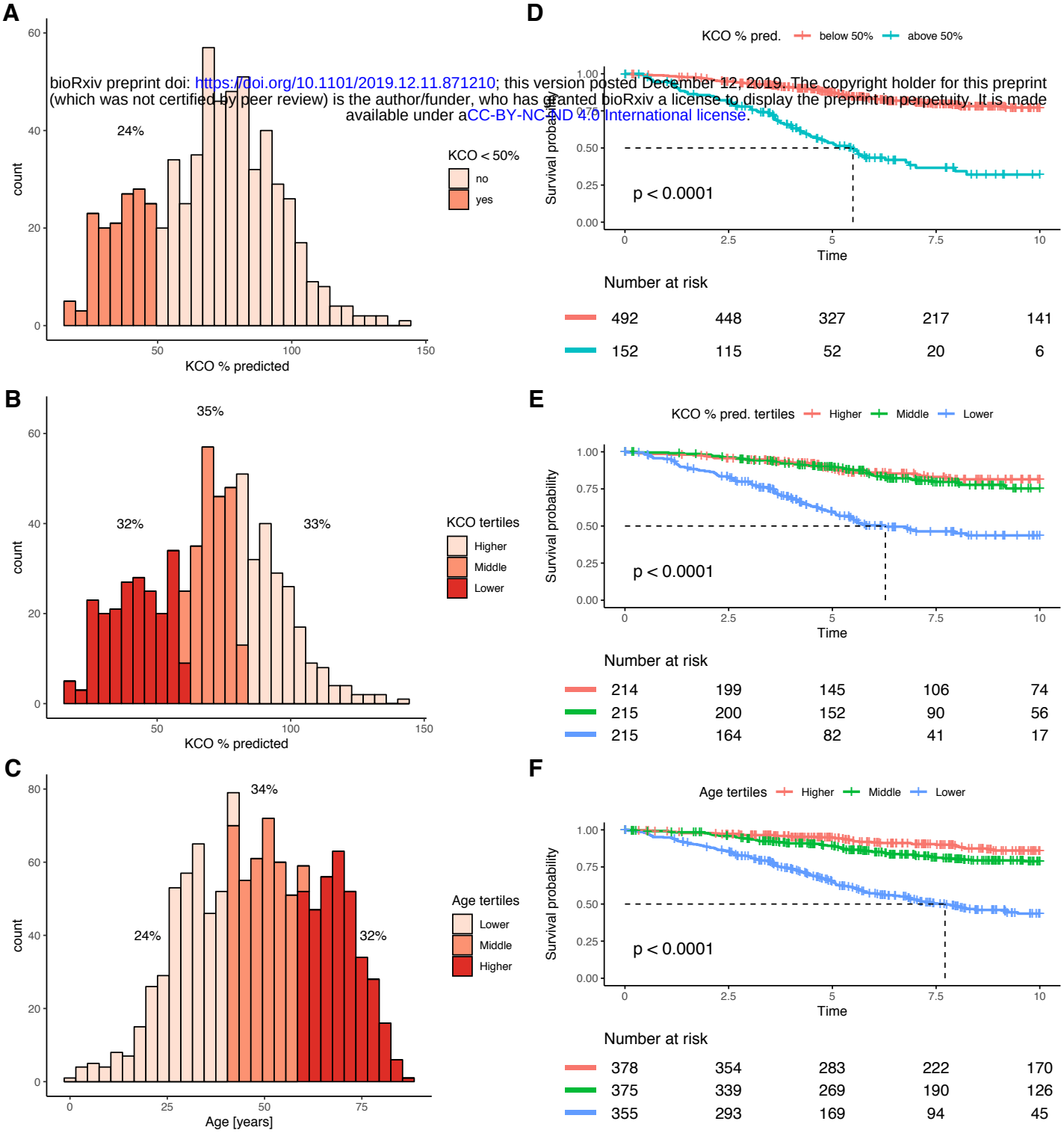


Figure S3. A-C. Distribution of transfer factor coefficient A. Coloured by KCO tertiles B. Coloured by KCO below and above threshold at 50% predicated. C. Distribution of age tertiles. D-F. Kaplan-Meier survival curves for KCO tertiles. There was a significant difference in survival between higher and lower and middle and lower tertile. Only lower tertile achieved median survival at 6.3 years. B. Kaplan-Meier survival curves for KCO below and above threshold at 50% predicted. Patients with KCO below 50% threshold median survival of 5.5 years. C. Kaplan-Meier survival curves for age tertiles. Survival in the higher age group was significantly lower than in low and middle tertile groups.

Figure S4. Summary of large deletions identified in previously established disease genes. A. Affected region containing *BMPR2*. B. Zoom into *BMPR2* locus. C. Affected region containing *GDF2*. D. Zoom into *GDF2* locus. E. Affected region containing *TBX4*. B. Zoom into *TBX4* locus.

Table 1: Definitions of tags and number of unrelated cases and controls for genetic association analysis with SeqOne.

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 aclinical diagnosis of IPAH	972	9134	2785
HPAH	Patients with aclinical 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,	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: PPH-PAH	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; age (0.96 - 40.7 years)	378	9136	2785
middle age	Lower age tertile; age (0.96 - 40.7 years)	376	9134	2779
old age	Old age tertile; age (58.6 - 88.1 years)	355	9136	2778
low KCO	KCO < 50% pred.	152	9136	2778
KCO lower tertile	KCO range 17-59% pred.	211	9136	2778
KCO middle tertile	KCO range 60-80% pred.	215	9136	2778
KCO higher tertile	KCO range 80-142% pred.	215	9134	2779

Table 2: BeviMed analysis results. Posterior probabilities of gene-tag associations. The High category comprise only variants of higher 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.

Gene	Tag	Bayes Factor	Posterior prbability	Consequence type	Mode of inheritance
BMPR2	HPAH	151	1.000	Moderate and high	dominant
BMPR2	FPAH	149	1.000	Moderate and high	dominant
BMPR2	young age	151	1.000	Moderate and high	dominant
BMPR2	IPAH	146	1.000	High	dominant
TBX4	I/HPAH/PVOD/PCH	25	1.000	High	dominant
TBX4	I/HPAH	26	1.000	High	dominant
TBX4	PH	25	1.000	High	dominant
TBX4	PAH	25	1.000	High	dominant
TBX4	IPAH	22	1.000	High	dominant
EIF2AK4	young_age	23	1.000	Moderate and high	recessive
EIF2AK4	low KCO	32	1.000	Moderate and high	recessive
EIF2AK4	KCO lower tertile	28	1.000	Moderate and high	recessive
BMPR2	PH	265	0.998	High	dominant
BMPR2	KCO middle tertile	55	0.997	Moderate and high	dominant
BMPR2	PAH	267	0.995	High	dominant
ACVRL1	HPAH	17	0.989	Moderate and high	dominant
KDR	KCO lower tertile	13	0.989	High	dominant
BMPR2	I/HPAH/PVOD/PCH	265	0.924	High	dominant
KDR	old age	11	0.915	High	dominant
BMPR2	KCO higher tertile	102	0.889	High	dominant
TBX4	young age	13	0.876	High	dominant
GDF2	I/HPAH	11	0.872	Moderate and high	dominant
EIF2AK4	FPAH	14	0.865	High	recessive
EIF2AK4	I/HPAH/PVOD/PCH	17	0.855	Moderate and high	recessive
GDF2	I/HPAH/PVOD/PCH	11	0.846	Moderate and high	dominant
EIF2AK4	PAH	17	0.845	Moderate and high	recessive
IDH3G	KCO middle tertile	11	0.830	Moderate and high	dominant
EIF2AK4	PH	17	0.820	Moderate and high	recessive
EIF2AK4	PVOD/PCH	16	0.819	Moderate and high	recessive
GDF2	PAH	10	0.801	Moderate and high	dominant
ATP13A3	KCO higher tertile	10	0.741	High	dominant
BMPR2	middle age	65	0.740	Moderate and high	dominant
GDF2	PH	10	0.731	Moderate and high	dominant
SOX17	young age	10	0.725	Moderate and high	dominant
KDR	low KCO	9	0.665	High	dominant
AQP1	PAH	12	0.663	Moderate and high	dominant
AQP1	I/HPAH/PVOD/PCH	13	0.660	Moderate and high	dominant
AQP1	PH	12	0.649	Moderate and high	dominant
AQP1	I/HPAH	13	0.649	Moderate and high	dominant
BMPR2	I/HPAH	267	0.640	High	dominant
AQP1	HPAH	13	0.625	Moderate	dominant

Table 3. Gene changes for IPAH patients harbouring protein truncating variants (PTV) in the KDR gene and PTV and missence variants in the IDH3G gene. WHO FC - World Health Organisation functional class, 6MWD - six 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 volumen in 1 sec, FVC - forced vital capacity, KCO - transfer factor coefficient for carbon monoxide. 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 harborung PTV in KDR had capillary hemanngioma, * DLCO% predicted; 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.

Gene	***KDR***								***IDH3G***						
Cohort	UK				US				UK					US	
WGS ID	W000229	E003448	W000274	E001392	CUMC-JM161	CCHMC12-190	CCHMC-19-023	CCHMC-27-015	E004190	E004149	E004194	E001063	W000031	CCHMC_22-105	CCHMC_10-074
Exon	3		22	3	2	3	5	22	1-13	1	1	12	12	13	4
HGVSc	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.217G>C
HGVSp	p.Trp61Ter	-	p.Arg1022Ter	p.Trp61CysfsTer16		p.Tyr101Ter		p.Arg1022Ter		p.Met356Thr	p.Thr346Ile	p.Pro25Leu	p.Pro25Leu	p.Pro364Leu	p.Val73Leu
Consequence type	stop gained	splice acceptor variant	stop gained	frameshift variant	splice donor variant	stop gained	stop gained	stop gained	large deletion	missense variant	missense variant	missense variant	missense variant	missense variant	missense variant
Shared	PAH(1)	PAH(1)	PAH(1)	PAH(1)	No	No	No	No	GEL(1); PAH(1)	PAH(1)	PAH(1)	PAH(2)	PAH(2)	gnomAD_exome_ALL 5.47E-06	gnomAD_exome_ALL-1.09E-05
CADD_PHRED_v1.4	40	34	36	33	26	38	24	37		23.9	17.15	23.7	23.7	23.3	21.7
GerpN	5.93	5.75	5.95	5.93	5.83	5.83	5.8	5.95		5.46	5.46	5.22	5.22		
Ansestry	European	European	European	European	East-Asian	European	European	European	East-Asian	European	European	European	European	European	European
Sex	male	female	male	female	female	male	female	female	female	female	female	female	female	female	male
Diagnosis	IPAH	IPAH	IPAH	IPAH	APAH-CHD secondary to double outlet RV	IPAH	IPAH	IPAH	IPAH	IPAH	IPAH	IPAH	IPAH	CHD-PAH	CHD-PAH
Age at diagnosis [years]	71	62	67	61	4	72	65	42	23	27	34	51	68	0	55
WHO FC	2	3	3	3		NA	NA	NA	4	3	4	4	2	3	3
6MWD [m]	472	422	660	180		380	NA	245	350		414		414	NA	316
SpO2 pre [%]	95	97	98	97	NA	NA	NA	NA	99	96	95	98	96		
SpO2 post [%]	86	86		91	NA	NA	NA	NA	97		99	96	95		
FEV1 [% pred.]	116	90	83	67.3	85%	NA	77%	NA	74	87	104	95	99.1		
FVC [% pred.]	115	94	91	72.8	92%	NA	83%	NA	76	90	109	95.8	96.3		
TLC [% pred.]	NA	NA	NA	NA	NA	NA	65%	NA	NA	NA	105	76	98		
KCO [% pred.]	44	46	46	55.2	NA	NA	35%*	NA	73	71	64	78	73		
Smoking history	Never	Never	Ex-smoker	Never	Never	Never	Ex-smoker	Never	Never	Ex-smoker	Never	Never	Never	Never	
mRAP [mmHg]	5	8	8	3	NA	5	29	14	15	14	8	12	6	3	7
mPAP [mmHg]	62	57	41	44	NA	49	66	60	58	64	49	50	62	46	69
PAWP [mmHg]	4	15	12	9	NA	5	16	15	15	8	10	12	7		10
CO [L/min]	3.6	4.58	5.966667	5.23	NA	4.33	1.8	4.6	2.37	3.23		3.29	4.1	4.4	
PVR	16.11	9.17	4.86	6.69	NA	NA	27.9	9.8	18.1	17.3		11.6	13.4		
Comrobidities	hyperlipidemia, HTN, DM type 2	HTN, hypothyrodism	DM type 2	CAD, DM type 2	No	HTN, hyperlipidemia,	HTN, hypothyrpoidism, OA	Obesity, CAD, DM type 2, hypothyroidism	No	No	No	PFO	No	Scimitar syndrome, hypoplastic rght lung, ASD with spontaneous closure	Large ASD
Family history	No	No	No	No	?	No	No	No	No	No	No	No	No	?	?
Status	alive	alive	alive	dead	?	alive	alive	alive	alive	alive	alive	death	alive	alive	alive

Table 4. Clinical characteristics of IPAH patients harbouring protein truncating variants in the KDR gene. IPAH - idiopathic pulmonary arterial hypertension, WHO FC - World Health Organisation functional class, 6MWD - six 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, NO - nitric oxide challenge, FEV1 - forced expiratory volume in 1 second, FVC - forced vital capacity, KCO - transfer factor coefficient for carbon monoxide, COPD - chronic obstructive pulmonary disease, OSA - obstructive sleep apnea, CAD - coronary artery disease, HTN - sytemic hypertension, CKD - chronic kidney disease, Hb - haemoglobin, WBC - white blood cells, TSH - thyroid-stimulating hormone. Comorbidities are reported as the number and percentage of cases possessing a disease entity. None of of the patients had a history of pulmonary embolism or asthma. Three of the KDR missense variants co-occurred with predicted deleterious variants in established PAH risk genes (BMPR2 and AQP1)

	***KDR* missense N=13**	***KDR* PTV N=4**	**p.overall**	**N**
Diagnosis verified: IPAH	13 (100%)	4 (100%)	.	17
Age[years]	46 [36;59]	64 [62;68]	0.113	17
Sex: female	9 (69%)	2 (50%)	0.584	17
BMI[kg/m^2]	29 [24;32]	26 [26;30]	1	13
WHO FC: II/III/IV [%]	23.1/9.2/7.7	25/75/0	1	17
6MWD[m]	312 [150;355]	301 [240;362]	0.814	11
SpO2 pre	95 [93;97]	97 [96;97]	0.335	11
SpO2 post	90 [80;96]	86 [86;88]	0.926	12
mRAP[mmHg]	8 [6;13]	6 [4;8]	0.431	14
mPAP[mmHg]	53 [42;62]	50 [43;58]	0.896	15
mPAWP[mmHg]	10 [8;13]	10 [8;13]	0.642	13
CO[L/min]	4.0 [3.0;5.5]	4.9 [4.3;5.4]	0.514	15
PVR[WU]	10.2 [4.56;14.3]	7.93 [6.23;10.9]	1	13
Acute NO challenge: vasoresponder	1 (33.3%)	1 (25.0%)	1	7
FEV1[% pred.]	84 [65;94]	86 [79;96]	0.48	14
FVC[% pred.]	86 [72;97]	92 [86;99]	0.723	14
FEV1/FVC ratio	0.78 [0.75;0.87]	0.78 [0.76;0.79]	0.671	14
KCO [% pred.]	89 [74;93]	46 [46;48]	0.008	11
Smoking history	6 (54.5%)	1 (25.0%)	0.677	15
COPD	1 (7.69%)	0 (0.00%)	1	17
Pulmonary fibrosis	0 (0.00%)	2 (50.0%)	0.044	17
CAD	1 (7.69%)	1 (25.0%)	0.426	17
HTN	5 (38.5%)	2 (50.0%)	1	17
CKD	1 (7.69%)	0 (0.00%)	1	17
Hb[g/l]	154 [140;166]	148 [135;152]	0.214	15
WBC[x10e9/l]	9.20 [6.30;11.0]	8.80 [8.23;9.55]	0.844	15
Platelets[x10e9/l]	262 [209;294]	216 [188;251]	0.361	15
Creatinine[umol/l]	78.0 [61.5;98.0]	67.0 [66.5;96.5]	0.866	13
TSH[mU/l]	3.65 [1.80;6.90]	1.76 [1.72;1.84]	0.234	12