Bayesian inference associates rare *KDR* variants with specific phenotypes in pulmonary arterial hypertension

A genotype-phenotype association study in PAH

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, Jim Lordan, PhD[23], Rajiv D. Machado, PhD[10], Robert V. MacKenzie Ross, MB, BChir[24], Colm McCabe, PhD[20,25], Shahin Moledina, MBChB[26], David Montani, MD, PhD[18], Horst Olschewski, MD[21,22], Joanna Pepke-Zaba, PhD[27], Laura Price, PhD[20,25], Christopher J. Rhodes, PhD[20], Werner Seeger, MD[16], Florent Soubrier, MD, PhD[15], Jay Suntharalingam, MD[24], Mark R. Toshner, MD[1,27], Anton Vonk Noordegraaf, MD[11], John Wharton, PhD[20], Jim Wild, PhD[6], Stephen John Wort, PhD[20,25], 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[1,3,19,27,$], Stefan Gräf, PhD[1,2,3,$]

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

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

[27]Royal Papworth Hospital NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, 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, 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:

Abstract

Background

Approximately 25% of patients with pulmonary arterial hypertension (PAH) have been found to harbor rare mutations in disease-causing genes. To identify missing heritability in PAH we integrated deep phenotyping with whole-genome sequencing data using Bayesian statistics.

Methods

We analyzed 13,037 participants enrolled in the NIHR BioResource - Rare Diseases (NBR) study, of which 1,148 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.

Results

Heterozygous, high impact, likely loss-of-function variants in the Kinase Insert Domain Receptor (*KDR*) gene were strongly associated with significantly reduced transfer coefficient for carbon monoxide (KCO, posterior probability (PP)=0.989) and older age at diagnosis (PP=0.912). We also provide evidence for familial segregation of a rare nonsense *KDR* variant with these phenotypes. On computed tomographic imaging of the lungs, a range of parenchymal abnormalities were observed in the five patients harboring these predicted deleterious variants in *KDR*. Four additional PAH cases with rare likely loss-of-function variants in *KDR* were independently identified in the US PAH Biobank cohort with similar phenotypic characteristics.

Conclusions

The Bayesian inference approach allowed us to independently validate *KDR*, which encodes for the Vascular Endothelial Growth Factor Receptor 2 (VEGFR2), as a novel PAH candidate gene. Furthermore, this approach specifically associated high impact likely loss-of-function variants in the genetically constrained gene with distinct phenotypes. These findings provide evidence for *KDR* being a clinically actionable PAH gene and further support the central role of the vascular endothelium in the pathobiology of PAH.

Introduction

Pulmonary arterial hypertension (PAH) is characterized by pulmonary vascular constriction and obliteration, causing elevation of pulmonary vascular resistance and ultimately, right ventricular failure. Molecular mechanisms such as aberrant angiogenesis[1], metabolic reprogramming and resistance to apoptosis[2] have been proposed to explain pulmonary vessel remodeling. A breakthrough in our understanding of the pathobiology underlying PAH was the discovery of heterozygous germline mutations in the gene encoding the bone morphogenetic protein receptor type 2 (*BMPR2*)[3], responsible for over 70% of familial PAH (FPAH) cases and 15-20% of idiopathic PAH (IPAH) cases. A smaller proportion (up to 10%) of PAH cases are caused by mutations in activin-like kinase 1 (*ACVRL1*)[4], endoglin (*ENG*)[5], SMAD family member 9 (*SMAD9*)[6], caveolin-1 (*CAV1*), involved in colocalization of BMP receptors[7], and the potassium channel *KCNK3*, responsible for membrane potential and vascular tone[8]. We recently identified rare pathogenic variants in growth differentiation factor 2 (*GDF2*), which encodes BMP9, a major ligand of the BMPR2/ALK1 receptor complex, 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[9]. 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[10] and deployed a recently developed Bayesian methodology (BeviMed)[11] that incorporates phenotypic data to increase the power to detect rare risk variants.

Methods

Data availability

The data of the NBR study have been deposited in the European Genome-Phenome Archive[10]. The data from the US PAH Biobank (PAHBB) and the Columbia University Medical Center (CUMC) are available via an application[12].

Study design, ethics, and subject recruitment

The NBR was established to identify genetic causes, streamline molecular diagnosis and develop new treatments for rare diseases through whole-genome sequencing (WGS) and deep phenotyping[10]. The study comprised 18 domains (Table I in the Data Supplement), including 1,148 adult and pediatric subjects recruited to the PAH domain. Included were individuals diagnosed with either IPAH or HPAH, pulmonary veno-occlusive disease (PVOD) or pulmonary capillary haemangiomatosis (PCH) and a small number of healthy relatives. Recruitment of PAH cases was carried out across the nine pulmonary hypertension (PH) specialist centers in the UK and by European collaborators (see Data Supplement). Patients recruited to the NBR study provided 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 I in the Data Supplement).

For validation, we used the US PAHBB cohort comprising exome sequencing data from 2,572 subjects diagnosed with group 1 PAH[12] and a biobank of 440 PAH patients established at CUMC[13].

Phenotyping of patients

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

Pseudonymized results of routine clinical tests were stored in the electronic data capture system *OpenClinica* (Table II in the Data Supplement). All cases were diagnosed between 2002 and 2017 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 Data Supplement). Diagnosis was ascertained in all cases. (Figure II in the Data Supplement). The full set of tags, with corresponding numbers of cases, controls, and excluded relatives, is shown in Table 1 and further discussed in the Data Supplement.

Analysis of computerized tomography (CT) scans

Diagnostic chest CT scans were performed in 613 study participants and the reports were transcribed to electronic case report forms. Of the 613 scans, 294 were available for more detailed analysis. The scans were reviewed by two independent cardiothoracic radiologists at the University of Sheffield with expertise in pulmonary hypertension (AS, SR), who were blinded to the underlying diagnosis, mutation and smoking status. For consistency and reproducibility, all measurements were reported on a customized proforma. Twenty-two scans were scored by both radiologists with good interobserver agreement (for further details see Data Supplement, Table III and IV).

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 phenotypically homogenous groups of patients 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), as discussed in Data Supplement, 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 I in the Data Supplement). Rare variants were extracted from each gene as previously described using the WGS data[10] (see Data Supplement). The analysis was restricted to canonical transcripts as annotated by Ensembl. The Bayesian model comparison method BeviMed[11] was applied to the 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, Table I and V in the Data Supplement). 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[14] 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 impact categories defined by Ensembl (high, moderate and the combination of both; see Data Supplement for more details). 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 Data Supplement). Rather than relying on significance thresholds (commonly used in frequentist approaches but discouraged in Bayesian statistical frameworks) we ranked our findings according to the posterior probability truncating the list around the level of the latest previously reported PAH risk genes.

Results

Characterization of study cohorts and tag definition

Whole-genome sequencing was performed in 13,037 participants of the NBR study, of which 1,148 were recruited to the PAH domain[10]. The PAH domain included 23 unaffected parents and three cases with an unknown phenotype, which were removed from the analysis (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 revealed that 57 participants (5%) had a diagnosis other than IPAH, HPAH or PVOD/PCH. These cases were subsequently relabelled and moved to the respective tag group for analysis. The comprehensive clinical characterisation of the study cohort is shown in Table VI in the Data Supplement. In summary, the median age at diagnosis was 49[35;63] years with a female predominance of 68%. Europeans constituted 84% of the study cohort. 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%, and 87%, whereas in incident cases it was 97%, 84%, and 72%, respectively. Median transfer

coefficient for carbon monoxide (KCO) in the entire studied population was 71[52;86]% predicted. 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 (see Figure III, Table VII, VIII, and IX in the Data Supplement). Survival in these groups was significantly worse than in those with preserved or mildly reduced KCO (Figure III A-D in the Data Supplement). 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 X in the Data Supplement).

Age at diagnosis was calculated as age at the time of diagnostic right heart catheterisation and was available in all but 10 cases. Patients in the higher age tertile showed more functional impairment despite milder hemodynamics, lower FEV1/FVC ratio and KCO [% predicted], as well as mild emphysematous and fibrotic changes on CT scans (Figure III E and F and Table XI in the Data Supplement).

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[14]. The majority of these variants have already been described in Gräf *et al.*[9] (see Data Supplement).

Rare variant association testing

We used BeviMed to consolidate previously reported PAH genes and to discover novel genotype-phenotype associations. Of note, 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 40 significant gene-tag associations with posterior probability (PP) above 0.75 (Table 2 and Figure 2A). *BMPR2, TBX4, EIF2AK4, ACVRL1* and *AQP1* showed the highest association (PP≥0.99) but we also confirmed significant associations in the majority of other previously identified genes. Individuals with rare variants in *BMPR2*, *TBX4* (high impact), *EIF2AK4* (biallelic) and *SOX17* had a significantly younger age of disease onset (tag: young age). We also confirmed the association of rare variants in *AQP1* with FPAH (log(BF)=10.023, PP=0.958). The refined phenotype approach corroborated the association between high impact variants in *BMPR2* and preserved KCO (KCO higher tertile, log(BF)=99.923, PP=1) together with an association of biallelic *EIF2AK4* mutations with significantly reduced KCO (KCO <50% predicted, log(BF)= 29.741, PP=1).

Under an autosomal dominant mode of inheritance, high impact variants in the Kinase Insert Domain Receptor (*KDR)* were associated with a significantly reduced KCO (KCO lower tertile, log(BF)=11.362, PP=0.989) and older age at diagnosis (tag: old age, log(BF)=9.249, PP=0.912).

Rare high impact variants in the new PAH candidate gene *KDR*

We identified five ultra-rare high impact variants in *KDR* in the study cohort. Ultra-rare variants exist in the general population only at a frequency of less than 1 in 10,000 (0.01%). Four of these were in PAH cases: one frameshift variant in exon 3 of 30 (c.183del, p.Tryp61CysfsTer16), two nonsense variants, one in exon 3 (c.183G>A, p.Trp61Ter) and one in exon 22 (c.3064C>T, p.Arg1022Ter) and one 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, and hence is likely to escape nonsense-mediated decay, but this remains to be studied functionally. Furthermore, 13 PAH cases (1%) and 108 non-PAH

controls (0.9%) harbored rare, predicted-deleterious *KDR* missense variants of moderate impact (Figure 3). The missense variant carriers, however, did not exhibit a reduced KCO or older age at 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 of unknown significance. Furthermore, three of these missense variants co-occurred with a predicted-deleterious variant in an established PAH risk gene (two patients carried also a variant in *BMPR2* and one a variant in *AQP1*).

Clinical characterization of *KDR* mutation carriers

Patients with high impact variants in *KDR* were older and exhibited significantly reduced KCO similar to biallelic *EIF2AK4* mutation carriers and in contrast to *KDR* missense variant and *BMPR2* mutation carriers (Figure 2B and C). Three of the four cases did not have a history of smoking. CT scans for all four patients showed a range of mild lung parenchymal changes (Figure 4). W000229 had evidence of mild mainly subpleural interstitial lung disease (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. E001392 showed mild centrilobular GGO 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 GGO 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 IV in the Data Supplement. There were no differences in the frequency of comorbidities between patients harboring missense and loss-of-function variants in *KDR* although the frequency of systemic hypertension was high (44%) (Table XII in the Data Supplement). Survival analysis could not

be conducted due to the small number 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 had a reduced KCO at 40% predicted. On the CT scan, mild interstitial fibrosis was observed (Figure V in the Data Supplement). Sanger sequencing confirmed that father and daughter carried the same deleterious *KDR* nonsense variant p.Trp61Ter (Figure 3B).

Additional *KDR* cases in US PAH cohorts

To seek further evidence for *KDR* as a new candidate gene for PAH, we analyzed subjects recruited to the PAHBB[12] and the CUMC[13] to identify additional patients carrying predicted pathogenic rare variants. Four additional individuals harboring rare high impact *KDR* variants were identified. These comprised, two 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 3. Three of the four patients were diagnosed with idiopathic PAH at 72, 65 and 42 years respectively, whereas one patient was diagnosed at age four with PAH associated with double outlet right ventricle. The diffusing capacity of carbon monoxide was available for one patient and was decreased at 35% predicted, with minor pleural scarring in the left upper lobe found on CT imaging. Two out of four patients (50%) harboring a high impact variant in *KDR* had been diagnosed with systemic hypertension.

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[15]. To deploy this approach in PAH we brought together phenotypic and genetic data using BeviMed[11]. This Bayesian framework allows the inclusion of prior information regarding the hypothesis being tested 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 impact categories.

Our BeviMed analysis provided strong statistical evidence of an association between ultra-rare, high impact variants in *KDR* and PAH with significantly reduced KCO and older age at diagnosis under a dominant mode of inheritance. Strikingly, likely loss-of-function variants in *KDR* exist in the general population with a frequency of only 4-7 per 100,000 (see Table 4 in the Data Supplement). In contrast, we identified four PAH cases in the NBR cohort which equates to almost 2 in 1,000. Additionally, the statistical constraint metrics provided by gnomAD[16] strongly suggest that loss-of-function variants in *KDR* are not tolerated (pLI = 1; o/e = 0.15 (0.09 - 0.25)). Besides the statistical evidence, we also identified one additional case with a family history, which together with a recently published case report of two families in which loss-of-function variants in *KDR* segregated with PAH and significantly reduced KCO[17], amounts to three reported familial cases with a distinct phenotype.

VEGFR2*,* which is encoded by *KDR*[18], binds VEGFA, a critical growth factor for physiological and pathological angiogenesis in vascular endothelial cells. In mice, even though VegfA haploinsufficiency is embryonically lethal[19], heterozygosity of its receptor, Vegfr2, is

compatible with life and unperturbed vascular development[20]. The role of VEGF signaling in the pathogenesis of PAH has been an area of intense interest since increased expression of VEGF, VEGFR1 and VEGFR2 were reported in rat lung tissue in response to acute and chronic hypoxia[21]. An increase in lung VEGF has also been reported in rats with PH following monocrotaline exposure[22]. In humans, VEGF-A is highly expressed in plexiform lesions in patients with IPAH[23]. In addition, inhibition of VEGF signaling by SU5416 (sugen) combined with chronic hypoxia triggers severe angioproliferative PH[24]. SU5416, a small-molecule inhibitor of the tyrosine kinase segment of VEGF receptors, inhibits VEGFR1[25] and VEGFR2[26] causing endothelial cell apoptosis, loss of lung capillaries and emphysema[27]. Further evidence supporting the role of VEGF inhibition in the pathobiology of PAH comes from reports of PH in patients treated with bevacizumab[28] and the multi-tyrosine kinase inhibitors[29,30]. Mutations in *KDR* have also been linked to congenital heart diseases. Bleyl *et al.* reported that *KDR* might be a candidate for familial total anomalous pulmonary venous return[31]. In addition, haploinsufficiency at the *KDR* locus has also been associated with tetralogy of Fallot[32]. We identified one patient in the CUMC cohort with PAH associated with congenital heart disease harboring a *KDR* likely protein-truncating splice donor variant (c.161+1G>T).

In the present study, we highlight that deep clinical phenotyping, in combination with genotype data, can improve the identification of novel disease risk genes and disease subtypes. *KDR* was already identified as a possible candidate gene, which did not achieve genome-wide significance, in our previous rare variant association study[9]. In combination with deep phenotyping data, *KDR* reached in the present study a significance level comparable to the most commonly affected genes in PAH. Reduced KCO, which reflects impairment of alveolar-capillary membrane function, has been noted in the analysis of early PAH registry data[33] to be an independent predictor of survival. Decreased KCO was also found in patients with PVOD/PCH with or without biallelic *EIF2AK4* mutations[34]. 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[35]. 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[36] 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 CT. None of the cases carrying a high impact variant in *KDR* 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[37]. Careful analysis of CT scans and clinical data did not reveal convincing evidence for this diagnosis in *KDR* mutation carriers. Cigarette smoking is a well-known factor leading to the decrease of KCO. Only one of the four *KDR* high impact variant carriers had a significant 15 pack-years smoking history, but with no signs of emphysema on CT. These findings suggest that loss-of-function variants 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 a high impact variant in *KDR* had mild fibrotic lung changes. Notably, 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 deleterious variants 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 entirely accurate. Alternatively, some cases of PAH might represent an oligogenic inheritance 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. In this analysis, 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 increase the power to detect new associations. Finally, KCO measurements were missing for a proportion of patients which could introduce a selection bias, although all the deleterious variants in *KDR* had phenotypic data available in the UK cohort.

In summary, this study shows that deep phenotyping enables patient stratification into subgroups with shared pathobiology and with increased power to detect new genotype-phenotype associations. We provide statistical evidence for an association between high impact, likely loss-of-function variants in *KDR* and significantly decreased KCO and later disease onset, further supported by familial segregation.

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 also thank Kathryn Auckland for proofreading the manuscript.

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; Steven Nathan, MD; Ronald Oudiz, MD; Ivan Robbins, MD; Robert Schilz, DO, PhD; Terry Fortin, MD; Jeffrey Wilt, MD; Delphine Yung, MD; Eric Austin, MD; Ferhaan Ahmad, MD, PhD; Nitin Bhatt, MD; Tim Lahm, MD; Adaani Frost, MD; Zeenat Safdar, MD; Zia Rehman, MD; Robert Walter, MD; Fernando Torres, MD; Sahil Bakshi, DO; Stephen Archer, MD; Rahul Argula, MD; Christopher Barne

Sources of Funding

The UK National Cohort of Idiopathic and Heritable PAH is supported by the National Institute for Health Research (NIHR), 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).

All research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre.

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. 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.

2. 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.

3. International PPH Consortium, Lane KB, Machado RD, Pauciulo MW, Thomson JR, Phillips JA 3rd, Loyd JE, Nichols WC, Trembath RC. Heterozygous germline mutations in

BMPR2, encoding a TGF-beta receptor, cause familial primary pulmonary hypertension. *Nat Genet*. 2000;26:81–84.

4. 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.

5. 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

6. 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

7. 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.

8. 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

9. 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.

10. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, Allen HL, Sanchis-Juan A, Frontini M, Thys C, Stephens J, Mapeta R, Burren OS, Downes K, Haimel M, Tuna S, Deevi SVV, Aitman TJ, Bennett DL, Calleja P, Carss K, Caulfield MJ, Chinnery PF, Dixon PH, Gale DP, James R, Koziell A, Laffan MA, Levine AP, Maher ER, Markus HS, Morales J, Morrell NW, Mumford AD, Ormondroyd E, Rankin S, Rendon A, Richardson S, Roberts I, Roy NBA, Saleem MA, Smith KGC, Stark H, Tan RYY, Themistocleous AC, Thrasher AJ, Watkins H, Webster AR, Wilkins MR, Williamson C, Whitworth J, Humphray S, Bentley DR, NIHR BioResource for the 100,000 Genomes Project, Kingston N, Walker N, Bradley JR, Ashford S, Penkett CJ, Freson K, Stirrups KE, Raymond FL, Ouwehand WH. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*. 2020;583:96–102.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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

16. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database Consortium, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–443.

17. 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

18. 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.

19. 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.

20. 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

21. 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.

22. 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.

23. 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

24. 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.

25. 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.

26. 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.

27. 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.

28. 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

29. 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

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

31. 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.

32. 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.

33. 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

34. 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.

35. 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.

36. 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

37. 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.

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 (for more details see Figure II in the Data Supplement). 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 estimate the posterior probability of gene-tag associations. **B**, Consort diagram summarizing 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 for carbon monoxide, MAF - minor allele frequency.

Figure 2. Rare variant 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 impact/consequence type of variants driving the association. Box-and-whisker plots showing the distribution of (**B**) the transfer coefficient for carbon monoxide (KCO) and (**C**) the age at diagnosis 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. Abbreviations: bial. - biallelic, lof - loss-of-function, mis. - missense.

Figure 3. Summary of rare single nucleotide variants (SNVs) and small insertions and deletions (indels) identified in the novel PAH candidate gene *KDR*. A, Only rare predicted deleterious variants in *KDR* 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 number P35968). Lollipop colors indicate the consequence type and sizes represent the variant frequency within a cohort. Missense variants that are predicted to be deleterious (SIFT) and damaging (PolyPhen-2) are colored in red, otherwise in yellow (i.e. SIFT and PolyPhen-2 disagree). High impact variants are labelled with the respective HGVS notation. The number of variants by predicted consequence type and cohort is provided in the table. **B**, Familial segregation of *KDR* nonsense variant c.183G>A (p.Trp61*) with PAH (i.e. reduced KCO and late onset) 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 high impact *KDR* mutations. A, Axial image of pulmonary CT angiogram at the level of the right ventricle (RV) moderator band, showing flattening of interventricular 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**,**H**).

Table legends

Table 1. Definitions of labels and the number of unrelated cases and controls in the rare variant association analysis with BeviMed. See paragraph on "Number of PAH domain samples in the analysis" in the Data Supplement for more details. Abbreviations: 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, LHD - Left Heart Disease, LD - Lung Disease, CTEPH - Chronic Thromboembolic Pulmonary Hypertension, KCO - transfer coefficient for carbon monoxide.

Table 2. BeviMed analysis results. Posterior probabilities and Bayes Factors (BF) of gene-tag associations. The BF is 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. The "High" category comprises only variants of high impact, including loss-of-function variants and large deletions; the “Moderate” category contains variants of moderate impact, including missense variants or variants of consequence type “non_coding_transcript_exon_variant”; the combined category “Moderate and High” includes both respective consequence types.

Table 3. Gene changes for IPAH patients harboring likely loss-of-function variants in the *KDR* gene. None of the *KDR* variants have previously been reported in gnomAD, ExAC or internal controls. HGVSc notations are based on transcript sequence ENST00000263923.4. HGVSp notations are based on the amino acid sequence ENSP00000263923.4. Abbreviations: *KDR* - Kinase insert domain receptor, 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.

Table 4. Comparison of high impact likely loss-of-function variants in *KDR* in the Human large-scale sequencing reference populations gnomAD and TOPMed with the NBR non-PAH controls and PAH cases. Abbreviations: gnomAD - The Genome Aggregation Database, TOPMed - The Trans-Omics for Precision Medicine program, KDR - kinase insert domain receptor, NBR - NIHR BioResource - Rare Diseases, LoF - loss of function.

Tables

Table 1. Definitions of labels and the number of unrelated cases and controls included in the rare variant association analysis with BeviMed. See paragraph on "Number of PAH domain samples in the analysis" in the Data Supplement for more details. Abbreviations: 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, LHD - Left Heart Disease, LD - Lung Disease, CTEPH - Chronic Thromboembolic Pulmonary Hypertension, KCO - transfer coefficient for 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 (BF) of gene-tag associations. The BF is 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. The "High" category comprises only variants of high impact, including loss-of-function variants and large deletions; the “Moderate” category contains variants of moderate impact, including missense variants or variants of consequence type “non_coding_transcript_exon_variant”; the combined category “Moderate and High” includes 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	PH	262.625	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	low KCO	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
TBX4	ENST00000240335	PH	22.877	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	PH	15.071	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	PH	10.791	0.980	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

Table 3. Gene changes for IPAH patients harboring likely loss-of-function variants in the *KDR* gene. None of the *KDR* variants have previously been reported in gnomAD, ExAC or internal controls. HGVSc notations are based on transcript sequence ENST00000263923.4. HGVSp notations are based on the amino acid sequence ENSP00000263923.4. Abbreviations: *KDR* - Kinase insert domain receptor, WHO FC - World Health Organization 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.

Cohort	**UK**					**US**			
ID	**W000229**	**W000229.d**	**E003448**	**W000274**	**E001392**	**CUMC-JM161**	**CCHMC-12-190**	**CCHMC-19-023**	**CCHMC-27-015**
Exon	3/30	3/30		22/30	3/30	2/30	3/30	5/30	22/30
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
HGVSp	p.Trp61Ter	p.Trp61Ter	-	p.Arg1022Ter	p.Trp61CysfsTer16		p.Tyr101Ter		p.Arg1022Ter
Consequence type	stop gained	stop gained	splice acceptor variant	stop gained	frameshift variant	splice donor variant	stop gained	stop gained	stop gained
Shared	PAH(1)	PAH(1)	PAH(1)	PAH(1)	PAH(1)	No	No	No	No
gnomAD	NA	NA	NA	NA	NA	NA	NA	NA	NA
CADD PHRED v1.3	38	38	26	37	35	26.4	38	24.3	37
SIFT	-		-	-	-	-	-	-	-
PolyPhen	-		-	-	-	-	-	-	-
GerpN	5.93	5.93	5.75	5.95	5.93	5.83	5.48	5.8	5.95
Ancestry	European	European	European	European	European	East-Asian	European	European	European
Sex	male	female	female	male	female	female	male	female	female
Diagnosis	IPAH	IPAH	IPAH	IPAH	IPAH	APAH-CHD secondary to double outlet RV	IPAH	IPAH	IPAH
Age at diagnosis [years]	71	53	62	67	61	4	72	65	42
WHO FC	2	3	3	3	3	2	NA	NA	NA

6MWD [m]	472	202	422	660	180	NA	380	NA	245
SpO_2 pre [%]	95	96	97	98	97	NA	NA	NA	NA
SpO_2 post [%]	86	87	86	NA	91	NA	NA	NA	NA
FEV_1 [% pred.]	116	70	90	83	67.3	85%	NA	77%	NA
FVC [% pred.]	115	76	94	91	72.8	92%	NA	83%	NA
TLC [% pred.]	NA	69	NA	NA	NA	NA	NA	65%	NA
KCO [% pred.]	44	40	46	46	55.2	NA	NA	35%*	NA
Smoking history	Never	Never	Never	Ex-smoker	Never	Never	Never	Ex-smoker	Never
mRAP [mmHg]	5	13	8	8	3	NA	5	29	14
mPAP [mmHg]	62	45	57	41	44	NA	49	66	60
PAWP [mmHg]	4	5	15	12	9	NA	5	16	15
CO [L/min]	3.6	3.3	4.58	5.97	5.23	NA	4.33	1.8	4.6
PVR	16.11	12.12	9.17	4.86	6.69	NA	NA	27.9	9.8
Comorbidities	hyperlipidemia, HTN, DM type 2	DM type 2, OSA, pulmonary fibrosis	HTN, hypothyroidism	DM type 2	CAD, DM type 2	No	HTN, hyperlipidemia,	HTN, hypothyroidism, OA	Obesity, CAD, DM type 2, hypothyroidism
Family history	Yes, daughter	Yes, father	No	No	No	No	No	No	No
Status	dead	alive	alive	alive	dead	alive	alive	alive	alive

Table 4. Comparison of high impact likely loss-of-function variants in *KDR* in the Human large-scale sequencing reference populations gnomAD and TOPMed with the NBR non-PAH controls and PAH cases. Abbreviations: gnomAD - The Genome Aggregation Database, TOPMed - The Trans-Omics for Precision Medicine program, *KDR* - kinase insert domain receptor, NBR - NIHR BioResource - Rare Diseases, LoF - loss of function.

Large-scale sequencing population	**High impact LoF variants in *KDR***	**Individuals**	**Alleles**	**Frequency**
gnomAD (v2.1)*	20	141,456	282,912	0.000071
gnomAD (v3)	10	71,702	143,404	0.000070
TOPMed	5	62,784	125,568	0.000040
NBR non-PAH controls	1	11,889	23,778	0.000042
NBR PAH cases	4	1,122	2,244	0.001783

**KDR* constraint metrics: pLI = 1; o/e = 0.15 (0.09 - 0.25); exp(LoF) = 73; obs(LoF) = 11

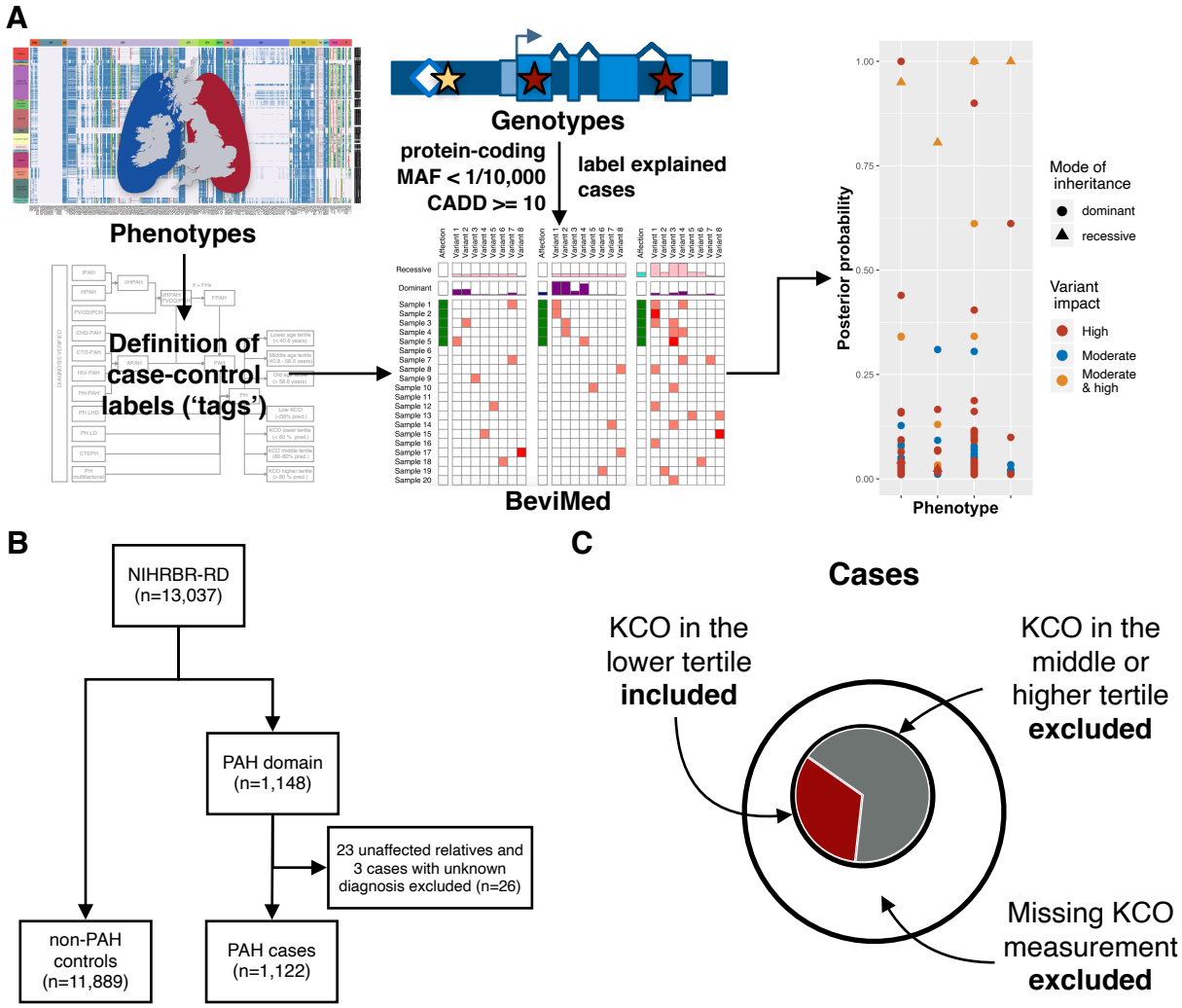


Figure 1. Design of the genetic association study. A, Overview of the analytical approach. Using deep phenotyping, data tags were assigned to patients who shared phenotypic features (for more details see Figure II in the Data Supplement). Rare sequence variants, called from whole-genome sequencing data, were filtered and explained cases were labeled. BeviMed was applied to a set of unrelated individuals to estimate 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 for carbon monoxide, MAF - minor allele frequency.

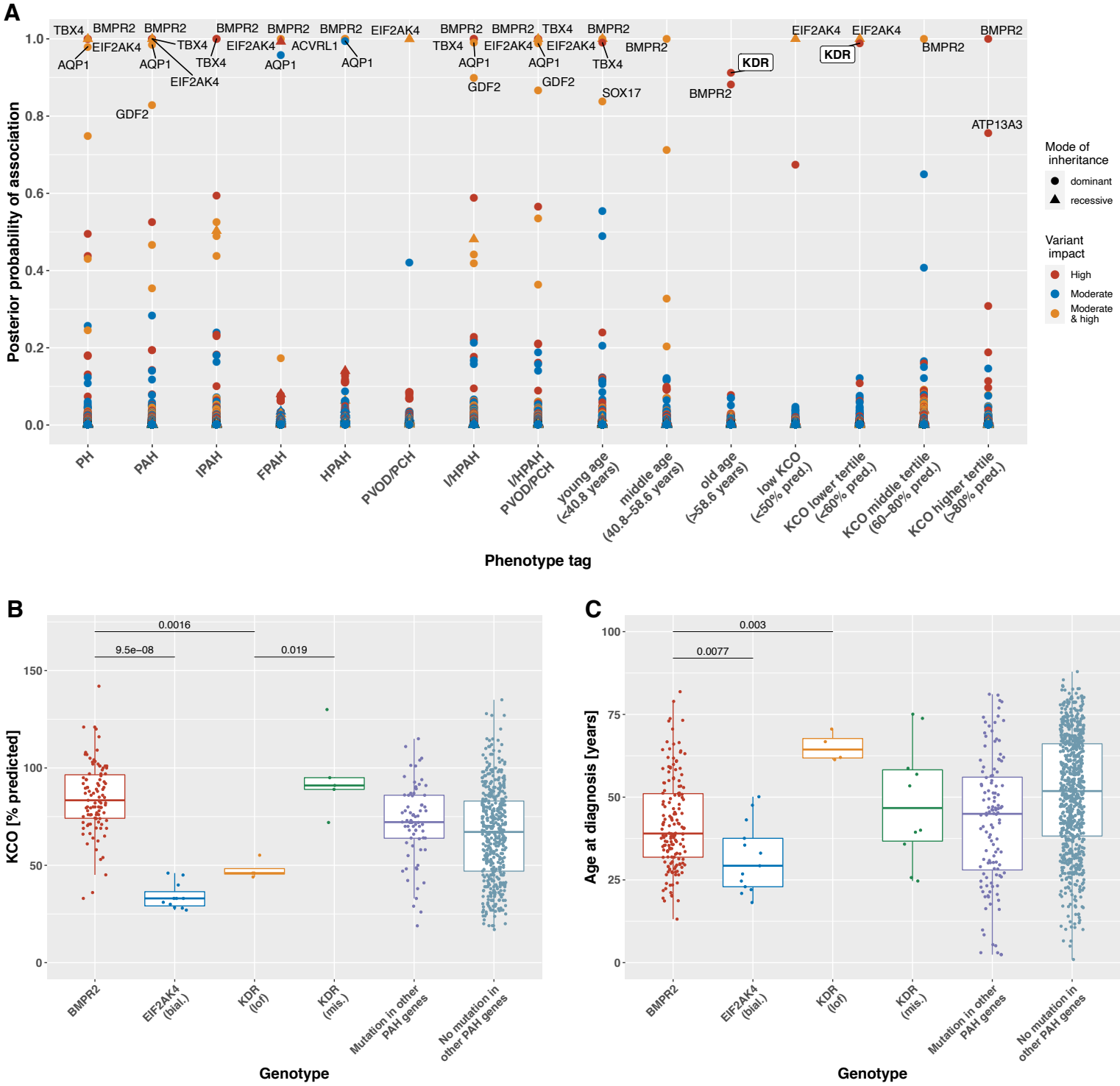
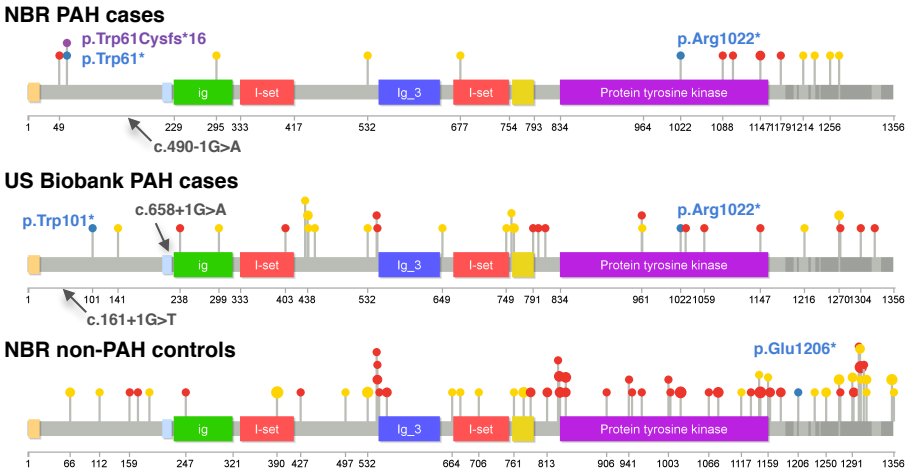


Figure 2. Rare variant association study results revealing established and novel genotype-phenotype links. A, Figure showing phenotype tags on the x-axis and corresponding posterior probability of genotype-phenotype association on the y-axis, as calculated by BeviMed. The definitions of the tags are listed in Table 1. Shape and colour of points indicate the mode of inheritance and impact/consequence type of variants driving the association. Box-and-whisker plots showing the distribution of (**B**) the transfer coefficient for carbon monoxide (KCO) and (**C**) the age at diagnosis 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. Abbreviations: bial. - biallelic, lof - loss-of-function, mis. - missense.

A

KDR SNVs and indels



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

KDR variant consequence	NBR		USBB
	Cases	Controls	Cases
frameshift	1	0	0
nonsense	2	1	2
splice acceptor	1	0	0
splice donor	0	0	2
missense	13	107	31

B

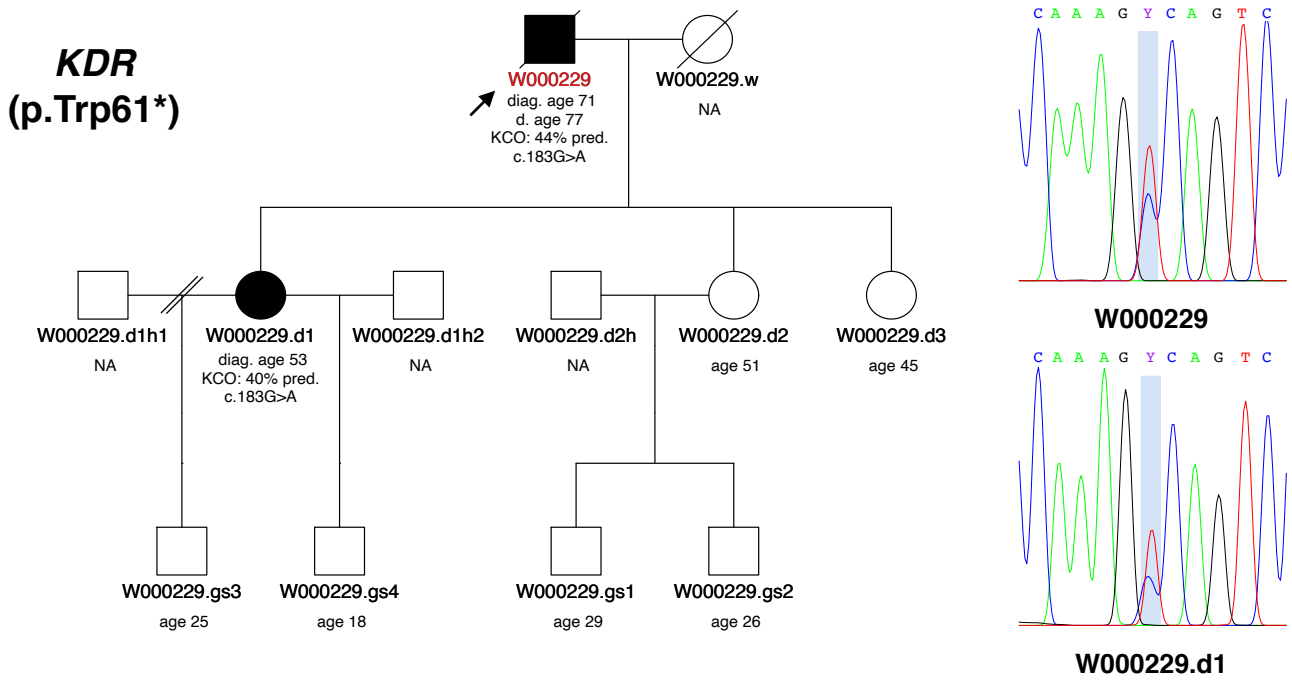


Figure 3. Summary of rare single nucleotide variants (SNVs) and small insertions and deletions (indels) identified in the novel PAH candidate gene *KDR*. A, Only rare predicted deleterious variants in *KDR* are shown (MAF<1/10,000 and CADD≥10). SNVs and indels are represented by colored lollipops on top of the protein sequence. The domain annotations were retrieved from Uniprot (accession number P35968). Lollipop colors indicate the consequence type and sizes represent the variant frequency within a cohort. Missense variants that are predicted to be deleterious (SIFT) and damaging (PolyPhen-2) are coloured in red, otherwise in yellow (i.e. SIFT and PolyPhen-2 disagree). High impact variants are labelled with the respective HGVS notation. The number of variants by predicted consequence type and cohort is provided in the table. **B**, Familial segregation of *KDR* nonsense variant c.183G>A (p.Trp61*) with PAH (i.e. reduced KCO and late onset) from father (W000229) to daughter (W000229.d). Sanger sequencing results are shown in the chromatograms.

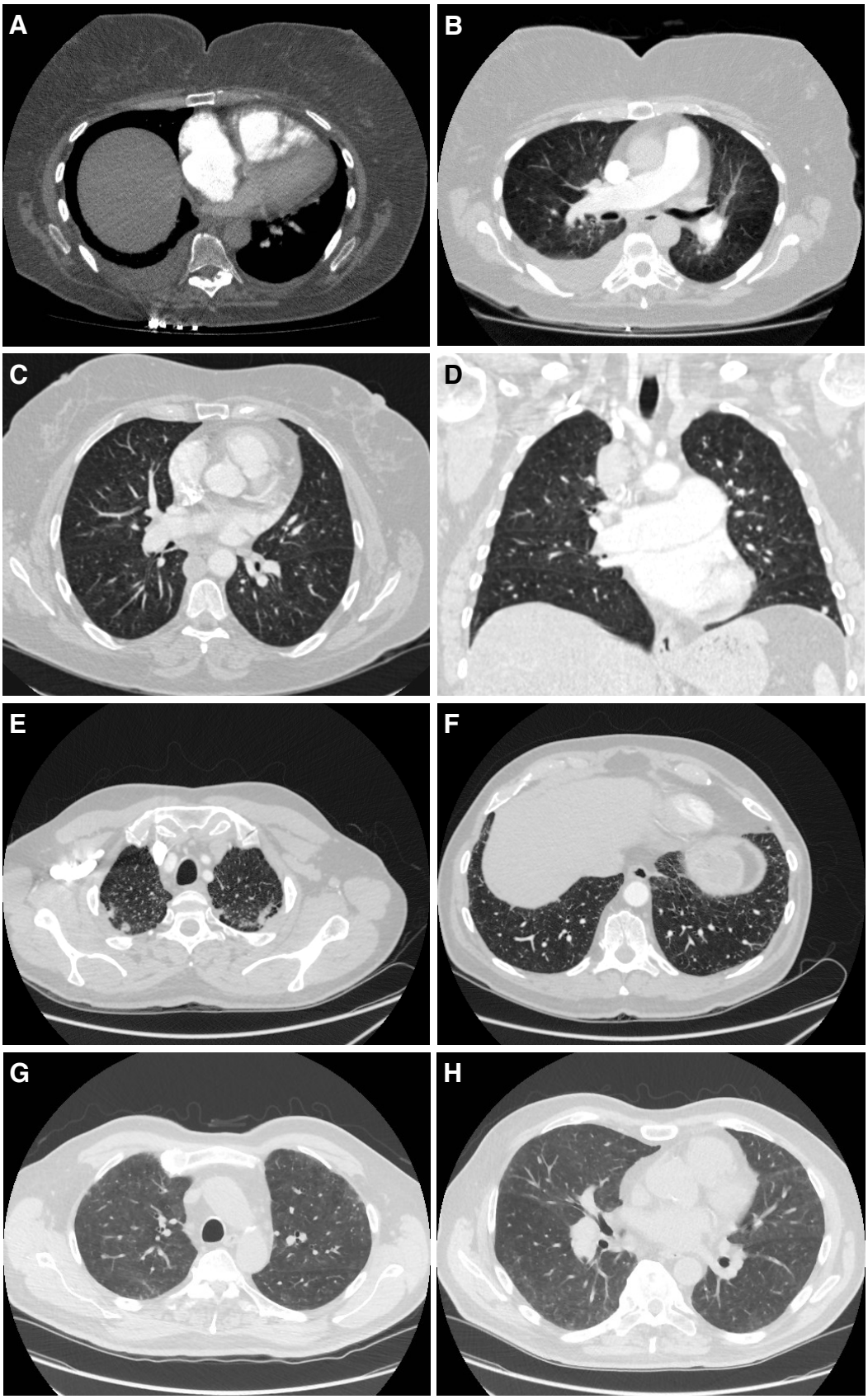


Figure 4. Chest computerized tomography (CT) scans of patients carrying high impact *KDR* mutations. A, Axial image of pulmonary CT angiogram at the level of the right ventricle (RV) moderator band, showing flattening of interventricular septum, leftwards bowing of the interatrial septum and the enlargement 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**, **H**).