Clustering Pulmonary Hypertension Patients Using the Plasma Proteome

Athénaïs Boucly, MD, PhD[1,2], Shanshan Song, PhD[1], Merve Keles, PhD[1], Dennis Wang, PhD[1,3], Luke S. Howard, MD, DPhil, FRCP [4], Marc Humbert, MD, PhD[2], Olivier Sitbon, MD, PhD[2], Allan Lawrie, PhD[1], A A Roger Thompson, PhD[5], Philipp Frank, PhD[6], Mika Kivimaki, FMedSci[6,7], Christopher J. Rhodes, PhD[1*], Martin R. Wilkins, DSc FMedSci[1*]

[1] National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, United Kingdom

[2] Université Paris-Saclay, AP-HP, INSERM UMR_S 999, Service de Pneumologie et Soins Intensifs Respiratoires, hôpital Bicêtre, Le Kremlin-Bicêtre, France

[3] Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore

[4] National Pulmonary Hypertension Service, Imperial College Healthcare NHS Trust, Hammersmith Hospital, London, United Kingdom

[5] Division of Clinical Medicine, The University of Sheffield, Sheffield, UK

[6] Brain Sciences, University College London, London, United Kingdom

[7] Clinicum, University of Helsinki, Helsinki, Finland

*Shared senior authorship

Corresponding author: Martin Wilkins (m.wilkins@imperial.ac.uk)

ORCID:

AB: 0000-0001-6246-5557

SS: 0000-0002-0851-8387

MK: 0000-0002-5212-3753

DW: 0000-0003-0068-1005

LH: 0000-0003-2822-210X

MH: 0000-0003-0703-2892

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

OS: 0000-0002-1942-1951

AL: 0000-0003-4192-9505

AT: 0000-0002-0717-4551

MiK: 0000-0002-4699-5627

CR: 0000-0002-4962-3204

MW: 0000-0003-3926-11771

Author contributions:

AB, CR, MRW designed the initial concept. AB and PF performed the statistical analysis. AB, CR and MRW wrote the manuscript. AB, SS, MK, DW, LH, MH, OS, AL, AART, PF, MK, CR and MRW participated in interpretation of data. All the authors reviewed the manuscript.

At a Glance Commentary

Current Scientific Knowledge on the Subject: It is recognised that the plasma proteome (by acting as a "liquid biopsy") has the potential to provide a deep molecular phenotype in pulmonary hypertension and enable personalised medicine. Studies to date have been largely confined to patients with pulmonary arterial hypertension and focused on prognostic markers for risk assessment rather than their use as theragnostics.

What This Study Adds to the Field: Through unsupervised clustering of the plasma proteome in a broad population of patients with clinically defined pulmonary hypertension, this study identified 4 patient groups linked to underlying molecular pathways, independent of the current clinical classification. The differential expression of PDGF and TGF-β pathways across the proteome clusters offers the opportunity for plasma proteomic profiling to select patients for studies of drugs targeting these pathways. The findings lay the foundation for the precise targeting of patents with tailored therapeutics according to molecular data.

Artificial Intelligence Disclaimer: No artificial intelligence tools were used in writing this manuscript.

This article has an online data supplement, which is accessible at the Supplements tab.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Abstract

Introduction

Patients with pulmonary hypertension are classified according to clinical criteria to inform treatment decisions. Knowledge of the molecular drivers of pulmonary hypertension might better inform treatment choice.

Methods

Between 2013 and 2021, 470 patients with pulmonary hypertension, 136 disease controls and 59 healthy controls were enrolled as a discovery cohort. Plasma levels of 7288 proteins were assayed (SomaScan 7K platform). Proteins that distinguished pulmonary hypertension from both control groups were selected for unsupervised clustering (k-means clustering of UMAP dimensions). Clinical characteristics and outcomes were compared across clusters. Separate cohorts of serially sampled patients from pulmonary hypertension centers in the United Kingdom (n=229) and France (n=79) provided independent validation.

Results

156 plasma proteins that distinguished pulmonary hypertension from disease and healthy controls formed 4 clusters with diverse 5-year survival rates: 78% (cluster 4), 62% (cluster 2), 44% (cluster 3), and 33% (cluster 1). The distinction and clinical relevance of the clusters were confirmed in validation cohorts by their association with survival. To further characterise the therapeutic relevance of the clusters we investigated 2 experimental drug targets: the Platelet-Derived Growth Factor (PDGF) pathway was up-regulated in cluster 3 compared to other clusters and the Transforming Growth Factor-β (TGF-β) pathway was up-regulated in cluster 1.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Conclusion

Plasma proteomic profiling of patients with pulmonary hypertension distinguishes 4 clusters, independent of the clinical classification. These groups, based on differential plasma protein levels, could act as theragnostic biomarkers for new therapies targeting PDGF and TGF-β pathways.

Word count 241/250

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Introduction

Pulmonary hypertension (PH) can present in relative isolation or as a comorbidity in left heart failure, chronic lung disease and other conditions.(1, 2) It causes death from right heart failure and remains a formidable challenge for therapeutic drug development.(1, 2) The first step in management is the classification of a patient into one of five clinical groups, which guides treatment strategy.(1, 2) Classification into a single group can be problematic as up to 40% of patients show mixed etiology.(3) Moreover, relying on clinical characteristics and measurements does little to define critical drug targets and aid new drug development.

Proteomics is a powerful tool for unravelling the intricate molecular landscape of diseases.(4) The plasma proteome comprises several thousand circulating proteins secreted or leaked from tissues.(5, 6) To date, the focus of high-throughput plasma proteomics in PH has been to identify key circulating markers of disease progression or treatment response in Group 1 patients with pulmonary arterial hypertension (PAH, precapillary PH that may be idiopathic, heritable, associated with drug exposure, connective tissue disease and congenital heart disease).(7, 8) (9)(10) However, this focus on Group 1, and the assignment on clinical criteria of some patients with PAH to other PH groups, particularly Group 2 (left heart failure) and Group 3 (lung disease)(3), may undermine the insights the plasma proteome can provide into finding new drug targets and therapeutic options. We argue that in-depth molecular profiling applied to the broader population of patients with a clinical diagnosis of PH is a better approach to developing targeted treatments for PH.(11)

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Here we use unsupervised clustering of plasma proteins from patients with clinically defined PH to identify robust protein signatures independent of the clinical classification, with the overarching goal of paving the way for more personalized and targeted therapeutic strategies.

Methods

Discovery cohort

The discovery study population comprised patients with suspected PH who attended Imperial College NHS Trust between 2013 and 2021. Patients with PH were classified in Group 1 (PAH), Group 2 (PH associated with left heart disease, PH-LHD), Group 3 (PH associated with lung disease, PH-lung) or Group 4 (chronic thrombo-embolic PH, CTEPH), using ESC/ERS guidelines (12, 13). Patients referred with suspected PH but with a mean pulmonary artery pressure (mPAP) <25mmHg on right heart catheterisation were classified as NoPH (symptomatic disease) controls. Contemporaneous plasma samples were obtained from volunteers without cardiovascular or respiratory diseases (healthy controls). All patients were recruited with informed written consent and local research ethics committee approval (11/LO/0395 and 17/LO/0563). Sample collection and processing are detailed in the supplemental methods.

Validation cohorts

Separate cohorts of PH patients with serial plasma samples collected over the same time period were used for independent validation: the UK National Cohort Study (NCT01907295);

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

the French EFORT study: Evaluation of Prognostic Factors and Therapeutic Targets in PAH (NCT01185730); and the Sheffield Teaching Hospitals Observational Study of patients with PH, Cardiovascular or Respiratory Disease (18/YH/0441). The Whitehall II study(14) provided a dataset on samples collected from a large cohort that were healthy at baseline.

Selecting relevant proteins

Patients from the discovery cohort were randomized into training (80%) and replication groups (20%). Proteins levels were compared between PH patients and both healthy and NoPH controls by logistic regression models, correcting for age, sex and principal component outliers (**Figure S1**). All comparisons were corrected for multiple testing using Benjamini-Hochberg false discovery rate (FDR). A threshold of $q<0.05$ was considered statistically significant.

To identify the combination of proteins that best predicted PH diagnosis, a least absolute shrinkage and selection operator (LASSO) modelling approach was applied(9), with the regularization parameter determined by the lowest error plus 1 standard error using the *glmnet* R-package.(15) Similar analyses were performed for proteins that distinguished PH patients and controls to identify the combination of proteins that best reflected PH pathology. Performance of these models was tested in the replication group by Receiver Operating Characteristic (ROC) analyses using the *pROC* R-package.

Clustering of PH patients using proteins

Proteins that distinguished PH patients from both healthy controls and NoPH controls (in models corrected for age, sex, principal component outliers, haemolysis, coagulation Factor

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

X and cystatin C) were taken forward for dimensional reduction using the UMAP R-package, followed by cluster analysis of protein-derived UMAP dimensions using the *NbClust* R-package. Demographic and clinical differences between the clusters were assessed by non-paired ANOVA, Kruskal-Wallis and chi-squared tests. We compared survival in the different clusters by log-rank test, from plasma sampling to death or censoring. We trained a Random forest model to classify new samples for cluster membership to validate our findings in independent cohorts. The classifier can be downloaded from https://doi.org/10.5281/zenodo.14509735

Pathway enrichment

Molecular enrichment analysis was performed using the *WebGestaltR* R-package. Heatmaps of proteins within pathways of PAH drugs in development were performed using *gplots* and *pheatmap* R-packages. The relative fluorescence of proteins of interest in the clusters were compared by non-paired ANOVA tests with Dunnett's multiple pairwise comparisons.

Statistical analysis was performed in R (version 4.3.1) and SPSS (version 29; IBM). An overview of the methodology is displayed in **Figure 1 and in the Supplemental Methods**.

Results

Study populations

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Our discovery study population comprised 470 PH patients, classified as PAH (n=131), PH-LHD (n=122), PH-lung (n=93) and CTEPH (n=124), along with 136 NoPH (symptomatic disease controls) and 59 healthy controls (**Figure 1, Table 1, Table S1**). Among patients with PH, 379 (81%) were newly diagnosed with PH (i.e. incident patients). The mean age was 64 $\pm$ 16 years; 56% were female and 74% were in functional class III. All individuals were randomised into training (80%) and replication (20%) subgroups (**Table S2**) for initial analysis. Validation was conducted on 2 independent PAH cohorts: one prevalent PAH cohort from the UK (n=165 patients including 125 with serial samples) and one incident PAH cohort from France with serial samples (n=79), and a separate UK PH-LHD group (n=64, **Table S3**). The Whitehall II study(14) (**Table S4**) provided an independent healthy control population (n=6196) as a negative control.

Plasma proteome profiles

Principal component analysis was performed to evaluate variation in protein expression profiles and identify patterns across the samples. The percentage of variance explained by each principal component is provided in **Table S5**. Using standard supervised analysis comparing PH patients and controls (detailed in the Supplemental Material), plasma proteins that differed by circulating level between PH and both healthy and NoPH controls were used to construct models among patients with PH to distinguish the main clinical PH groups (**Tables S6-S8, Figures S1-S6**). There was significant overlap in the proteins associated with each PH subgroup (all pairs $p<0.001$, Fisher's exact test, **Figure S1**), suggesting important molecular clusters across these clinical groups.

Unsupervised cluster analysis of all PH patient proteomes

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

To identify novel proteomic clusters, we focussed on proteins associated with PH irrespective of clinical group and robust to potential confounders. Plasma levels of 165 SOMAmers (targeting 156 proteins) differed significantly between PH (of any aetiology) and both NoPH and healthy controls, after correction for age, sex, principal component outliers, haemolysis, coagulation Factor X and cystatin C (**Figure 2**). The dimensions of this dataset of proteins were reduced by UMAP. Unsupervised K-means clustering analysis of the proteomic UMAP dimensions of all 470 PH patients revealed that, with a substantial stability rate of 89%, the optimal number of clusters was 4 (**Figure 3A**), supporting a robust and consistent clustering of patients which was visually apparent (**Figure S7**).

Patients in cluster 4 were younger and had fewer comorbidities than the others, while patients in cluster 1 had more severe PH (**Table 2**). After a median of 3.2 years (interquartile range 1.8-5.3) from plasma sampling, 188 (40%) patients had died. Events occurred in 65% of cluster 1, 59% of cluster 3, 33% of cluster 2 and 23% of cluster 4. At 5 years, the Kaplan-Meier survival rate was divergent (log rank test, $p<0.001$; **Figure 3B**), highest in cluster 4 (78%), lowest in cluster 1 (33%) and intermediate for cluster 2 (62%) and 3 (44%).

In the subset of 131 patients with PAH, patients in cluster 4 had the best survival, patients in cluster 1 had the worse survival while patients in cluster 2 and 3 had a similar survival (log-rank $p=0.73$, **Figure S8**). Hence, in the subsequent survival analyses in PAH-only independent cohorts, clusters 2 and 3 were combined.

Cross check with known prognostic biomarkers

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

To ‘sense check’ our clusters, we compared the plasma levels of previously identified prognostic protein biomarkers(7, 9, 10, 16–20) across the clusters (**Figure S9**). Many, such as BNP, NT-proBNP, Beta-Nerve Growth Factor (ß-NGF), C-X-C Motif Chemokine Ligand 9 (CXCL9), Activin A, follistatin-like 3 (FSTL3), renin, matrix metalloproteinase 2 (MMP2), inhibitors of metalloproteinases 1 and 2 (TIMP1/TIMP2), thrombospondin 2 (TSP2), insulin-like growth factor-binding protein-1 (IGFBP1), interleukin 1 like-receptor-4 (IL1-R4), interleukin 18 (IL-18), peroxidasin (PXDN) or polydom (SVEP1), were significantly increased in cluster 1 (poorest survival) compared to the other clusters, the direction of change consistent with previously published observations for these proteins.

Enrichment of biological pathways

To further understand the proteins that characterise each of the clusters (**Figure S10**), we conducted an enrichment analysis of the top 100 up- and down-regulated proteins from each cluster. This highlighted significant biological pathways, revealing a diverse array of enriched terms that provide valuable insights into the underlying molecular mechanisms (**Table S9 and Figure S11**). For example, extracellular matrix organization proteins were down-regulated in cluster 4 (associated with best survival) but up-regulated in cluster 1 (worse survival, **Figure S11**).

Validation of proteomic clustering in two independent PAH cohorts and one cohort of PH-LHD

We trained a Random forest classifier on a combination of 61 proteins, selected by LASSO regression, to assign new samples to one of 4 clusters. LASSO scores for each cluster (**Table S10**) clearly distinguished cluster membership (**Figure S12**). Using these scores as input, we

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

trained a random forest classifier in the discovery cohort and applied this to predict clusters in the independent cohorts (**Figure S13**). To confirm the robustness and clinical relevance of our clusters, we then assessed risk and outcomes. In separate UK and French PAH cohorts, 97% and 88% respectively of patients classified as cluster 4 were either at low risk or intermediate-low risk of death according to the ESC/ERS 4 strata risk tool,(1, 2) while 75% and 92% of patients classified as cluster 1 were at intermediate-high or high risk of death (log rank test, $p<0.001$ in both cohorts).

Consistent with our findings in the discovery cohort, 5-year survival was better in cluster 4, worse in cluster 1 and intermediate in clusters 2 and 3 (log rank test $p<0.001$ in both PAH validation cohorts at each time, **Figure 3C-F**). This was also observed in a PH-LHD patient group (log rank test, $p=0.022$, **Figure S14A**) and in a mixed cohort combining the UK PAH and PH-LHD patients (log rank test, $p<0.001$, **Figure S14B**).

Patient migration between clusters over time and survival

To assess the dynamic nature of our clusters, we assessed serial samples from the UK PAH Cohort and the French EFORT validation cohorts. In the UK and French cohorts, 36% and 38%, respectively, changed cluster over time (**Figure 4 A&B**). Patients who switched from cluster 2 or 3 to cluster 1 (n=8) had a poorer survival than those who remained in the same cluster or switched to cluster 4 (n=58, log rank test, $p<0.001$, **Figure 4C**), while changes from cluster 1 to another (n=8) were associated with a significant improvement in survival (log rank test, $p=0.006$, **Figure 4D**).

Identification of potential theragnostic biomarkers

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

To investigate the therapeutic relevance of the protein clusters, we investigated two potential disease modifying drug targets: the Platelet-Derived Growth Factor (PDGF) and Transforming Growth Factor-β (TGF-β) pathways (**Figure 5A** and **Figure 6A**). The PDGF pathway was upregulated in cluster 3 compared to other clusters (**Figure 5A**). In particular, levels of PDGF-BB were higher in cluster 3 than in other clusters in both discovery and validation cohorts (Dunnett's pairwise comparisons, p<0.001, **Figure 5B&C**).

The TGF-β pathway was downregulated in cluster 3 and upregulated in cluster 1 (**Figure 6A**). Levels of Activin A were higher in cluster 1 than in clusters 3 and 4 (Dunnett's pairwise comparisons, q<0.001) in discovery (**Figure 6B**) and higher than in cluster 4 in the validation cohort (Dunnett's pairwise comparisons, q=0.009, **Figure 6C**), while levels of follistatin were significantly higher in cluster 1 than in other clusters in both cohorts (ANOVA, p<0.001, Dunnett's pairwise comparisons q<0.001, **Figure 6D&E**).

Distribution of cluster proteins in a population cohort

The proportion of patients assigned to cluster 1 (highest mortality) fell with decreasing mPAP (**Table S11**). The Whitehall II study provided the opportunity to investigate the distribution of the proteins in the general population. We hypothesised that the clusters associated with intermediate-high risk PH would be poorly detected in this cohort. Of the 6196 Whitehall II participants with valid protein data, only 2 (0.032% vs 22.4% in PH) belonged to cluster 1 while clusters 2 (n = 213, 3.4% vs 30% in PH), and 3 (n = 527, 8.5% vs 12.6% in PH) were uncommon and cluster 4 represented the majority (n = 5454, 88% vs 35% in PH, **Figure S15**).

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Discussion

Here, a comprehensive analysis of the circulating proteome, involving 470 PH (Groups 1-4) patients and 195 controls, dissected the clinical presentation of PH into distinct molecular subsets. We identified plasma proteins that distinguish PH from both healthy and NoPH (disease) controls and, through unsupervised clustering independent of the clinical classification, revealed 4 PH patient clusters linked biologically to underlying pathways manifesting significant differences in survival. In doing this, we identified patients where the underlying pathology may plausibly be driven by pathways targeted by drugs currently under investigation. These patients could be prioritised for targeted clinical studies.

It is well recognised that PH is a convergent phenotype that presents significant challenges for diagnosis, treatment, and prognosis. The widely used clinical classification acknowledges that PH may arise alone or as a co-morbidity but does not inform the underlying pathology. The plasma protein profile can help to differentiate PAH from healthy controls(10) and inform prognosis for PAH patients(9, 10, 16) but has also emerged as a molecular instrument for unravelling the pathophysiological diversity of PH.(10) Sweatt et al used a multiplex immunoassay and machine learning to identify immune endotypes in PAH.(7) Here we broaden the proteomic net and examine differences in circulating levels of approximately 7,000 proteins across the clinical spectrum. We were able to identify protein signatures associated with the clinically-defined PH groups, but there was significant overlap across these groups. In short, the clinical groups did not distinguish patients based on disturbed

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

biological pathways that would inform treatment. We therefore turned to advanced unsupervised bioinformatics to classify PH patients based on plasma protein distribution.

We identified 4 distinct clusters of patients based on their proteomes. The biological importance of these is evident in that they stratified patients with different clinical severity and outcomes. Validation analyses performed on two PAH-only independent cohorts and one cohort of PH-LHD confirmed the link between clusters and survival, emphasising their clinical relevance, and showed that dynamic changes in clusters over time were associated with significant changes in survival. The distribution of recognised prognostic biomarkers in PAH across the clusters was consistent with previous studies and further underscores their biological significance.(7, 9, 10, 16–20) For example, circulating levels of BNP, NT-proBNP, renin, cytokines, Activin A, FSTL3, and proteins involved in extracellular matrix organisation were increased in cluster 1 (the cluster with the poorest survival) and lowest in cluster 4 (the cluster with the best survival). This makes biological sense; circulating BNP and NT-proBNP report on cardiac workload, while circulating levels of extracellular matrix organization proteins may link to ongoing vascular remodelling.(21, 22)

The real clinical opportunity in the 4 protein clusters is not their use as prognostic markers but in their potential to guide therapeutic decision making through the prism of personalized medicine. As proof of principle, we investigated known drug targets: the PDGF and TGF-β pathways.(23) The PDGF pathway was upregulated in cluster 3. This pathway has long been implicated in the pathogenesis of PH, due to its role in mediating vascular remodelling and proliferation of pulmonary artery smooth muscle cells.(24) Oral imatinib, a tyrosine kinase inhibitor, has been shown to improve haemodynamics and exercise capacity in PAH, although

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

with concerns about safety in this patient group.(25) The PDGF pathway remains of active interest as a therapeutic target (26) and cluster 3 could be exploited to identify a subset of patients where the benefits of tyrosine kinase inhibition outweigh the potential side effects.

Likewise, upregulation of the TGF-β pathway in cluster 1 might signal a group of patients most likely to benefit from drugs such as the activin ligand trap, sotatercept, that target this pathway. Genetic and now pharmacological studies with sotatercept underscore the importance of the TGF-β pathway in PAH. Its dysregulation has been linked to endothelial dysfunction, inflammation, and fibrosis in the pulmonary vasculature.(27, 28) Sotatercept, derived from the activin receptor type IIA, is thought to rebalance bone morphogenetic protein (BMP)-TGF-β signalling in PAH.(28) A recent proteomic study of a small number of patients has reported the effect of sotatercept on a panel of circulating biomarkers, including reducing BMP9 and BMP10 levels and changes in inflammatory mediators. (29) The Phase II PULSAR and the phase III STELLAR trials have provided evidence that sotatercept, when added to standard therapy, significantly improves haemodynamics and exercise capacity in patients with PAH, although not without safety concerns.(30–32) Utilising the proteomic signature from cluster 1 may permit better targeting of the drug to patients that will benefit.

This introduces the concept of theragnostics to PH medicine; the use of a test to inform and direct drug therapy. Currently, drug selection is based on the clinical subgroup to which a patient is assigned and their ‘risk score’, an assessment of the severity of their PH. (1,2) By identifying patients with upregulated PDGF or TGF-β pathways, clinicians could tailor PAH management when considering drugs that act on these pathways. Treatments could be directed towards the specific molecular drivers perturbed in each patient and improve the

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

benefit-harm balance that accompanies every drug. The protein clusters may also identify patients assigned to other clinical PH groups (i.e. outside Group 1) that might benefit from these drugs and deserve inclusion in clinical trials. Integrating these clusters, dervived from proteomic profiling, into future clinical studies is the next step towards validating their translational value and assessing their potential clinical impact.

A significant strength of our study lies in the large patient cohort recruited in PH expert centres and the validation of our findings across 2 independent PAH cohorts, with serial samples, and one cohort of PH-LHD. While generated in a cohort of patients with largely prevalent PH (UK cohort), the 4 protein clusters were reproduced in newly diagnosed, treatment-naïve patients (French cohort) and were not affected by duration of illness; the median duration of PH in the discovery cohort was similar across the clusters and so not a major factor in determining protein distribution. Conversely, the risk-associated clusters were not prevalent in the general population (Whitehall II study). This observation speaks to the importance of using the 4 clusters in context; refining the management of patients with a clinical diagnosis of PH.

This study used a mPAP ≥25mmHg rather than >20mmHg to define PH, in line with the license for currently approved drugs; excluding the small number (n=25) patients with a mPAP >20 to ≤24mmHg from the analysis did not affect the clusters. There are limitations to the SomaScan assay. While the platform has a large number of proteins, there remain many more measurable proteins in plasma not included in this analysis. Furthermore, the assay provides measurements as RFUs (Relative Fluorescent Units), rather than absolute concentrations. These values can be used to compare patients and changes over time, but they are not

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

suitable for use in clinical applications that require absolute concentration to inform treatment decisions. Previous studies showed a good correlation between SomaScan measurements and ELISA(9, 33–35) and mass spectrometry(36), giving this assay a high degree of confidence. Blood samples were collected alongside routine clinical plasma samples, showing the practical deployment of this protein panel in a clinical setting. However, for the panel to be routinely useful and at a reasonable cost, the rapid automated testing of the panel of proteins needed to identify clusters 1 and 3 on a widely available platform would be required.

Conclusion

Through an unsupervised analysis of the plasma proteome, we have identified molecular signatures that may redefine the classification and management of PH, echoing precision medicine approaches adopted in other fields, such as oncology. We described 4 PH patient groups linked to underlying pathways, independent of the current clinical classification of PH. The differential expression of PDGF and TGF-β pathways across the proteomic clusters signposts a new era of personalized therapy in PH. These findings advocate for the inclusion of plasma protein profiling in routine clinical assessment to enable the precise targeting of molecular pathways with tailored therapeutics, ultimately improving patient outcomes and advancing the field towards truly personalized medicine.

Word count (excluding abstract) 3497

ACKNOWLEDGMENT

The authors thank all volunteers and patients for their participation and acknowledge the NIHR BioResource centres, the Imperial NIHR Clinical Research Facility and staff for their contributions. We thank the NIHR BioResource-Rare Diseases Consortium and UK PAH Cohort Study Consortium.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

SOURCES OF FUNDING

AB is supported by ERS/EU RESPIRE4 Marie Skłodowska-Curie Postdoctoral Research Fellowship (R4202205-00947). DW is supported by the Wellcome Trust (217068/Z/19/Z), the UK Academy of Medical Sciences (APR7_1002), UK Engineering and Physical Sciences Research Council (EP/V029045/1), Singapore A*STAR (H24P2M0005). MiK and PF are supported by the Wellcome Trust, UK (221854/Z/20/Z). MiK is also supported by National Institute on Aging (NIH), US (R01AG056477), Medical Research Council, UK (MR/Y014154/1), Academy of Finland (350426) and Finnish Foundation for Cardiovascular Research (a86898). The UK National Cohort of Idiopathic and Heritable PAH was supported by; the NIHR BioResource; the British Heart Foundation (BHF SP/12/12/29836) and the UK Medical Research Council (MR/K020919/1). CJR is supported by BHF Basic Science Research fellowship (FS/SBSRF/22/31025). The funder had no role in study design, data collection, data analysis, data interpretation, or writing of this article. The views expressed are those of the authors.

DISCLOSURES

BOUCLY Athénaïs reports a relationship with Grupo Ferrer Internacional SA that includes: board membership, consulting or advisory, and speaking and lecture fees. BOUCLY Athénaïs reports a relationship with Merck Sharp & Dohme UK Ltd that includes: board membership, consulting or advisory, funding grants, and speaking and lecture fees. BOUCLY Athénaïs reports a relationship with Astra Zeneca that includes: speaking and lecture fees. BOUCLY Athénaïs reports a relationship with AOP Orphan that includes: board membership, consulting or advisory, and speaking and lecture fees. BOUCLY Athénaïs reports a relationship with Janssen Pharmaceuticals Inc that includes: speaking and lecture fees.

SONG Shanshan reports no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

KELES Merve reports no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

WANG Dennis reports Wellcome Trust (217068/Z/19/Z), Academy of Medical Sciences (APR7_1002), UK Engineering and Physical Sciences Research Council (EP/V029045/1), A*STAR (H24P2M0005).

HOWARD Luke reports a relationship with Janssen, Aerovate, MSD, Apollo, Gossamer Bio that includes: board membership, consulting or advisory, and speaking and lecture fees. Howard Luke reports ATXA Therapeutics stock options.

HUMBERT Marc reports a relationship with 35 Pharma that includes: consulting or advisory fees. Humbert Marc reports a relationship with Aerovate that includes: board membership, consulting or advisory fees. Humbert Marc reports a relationship with AOP Pharma that includes: board membership, consulting or advisory fees, speaking and lecture fees, funding grants. Humbert Marc reports a relationship with Bayer that includes: speaking and lecture fees. Humbert Marc reports a relationship with Chiesi Pharmaceuticals Inc that includes: board membership, consulting or advisory fees. Humbert Marc reports a relationship with Grupo Ferrer Internacional SA that includes: board membership, consulting or advisory, and speaking and lecture fees. Humbert Marc reports a relationship with Janssen that includes: board membership, consulting or advisory, and speaking and lecture fees. Humbert Marc reports a relationship with Keros that includes: board membership, consulting or advisory, funding grants, and speaking and lecture fees. Humbert Marc reports a relationship with United Therapeutics that includes: board membership, consulting or advisory.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

SITBON Oliver reports relationships with pharmaceutical companies including Aerovate, Altavant, AOP Orphan, Gossamer Bio, Ferrer, Janssen, Liquidia, Merck, Respira Therapeutics, Roivant and United Therapeutics. In addition to being investigator in clinical trials involving these companies, relationships include board membership, consulting or advisory, and speaking and lecture fees. In addition, Dr Sitbon's institution received funding grants from AOP Orphan, Gossamer Bio, Ferrer, Janssen and Merck.

LAWRIE Allan reports grants or contracts with British Heart Foundation, Alexion Pharmaceuticals, Janssen Pharmamaceuticals and MSD.

THOMPSON reports support from Janssen-Cilag Ltd

FRANK Philipp reports a grant: Wellcome Trust (221854/Z/20/Z).

KIVIMAKI Mika reports grants from Wellcome Trust (221854/Z/20/Z), National Institute on Aging (R01AG056477), Medical Research Council (MR/Y014154/1), Academy of Finland (360426), Finnish Foundation for Cardiovascular Research (a86898).

RHODES Christopher reports a BHF Senior Basic Science fellowship (FS/SBSRF/22/31025) and a Cardiovascular Theme grant from NIHR Imperial Biomedical Research Centre (BRC).

WILKINS Martin reports a relationship with MorphogenIX, VIVUS, Janssen, Chiesi, Aerami, Benevolent AI, Pennington Marches, Sprigings, Apollo Therapeutics, GSK, Acceleron, Novartis that includes: board membership, consulting or advisory, and speaking and lecture fees. Wilkins Martin reports W12 Therapeutics stock options.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

References

1. Humbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, *et al.* 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Respir J* 2023;61:2200879.
2. Humbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, *et al.* 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J* 2022;43:3618–3731.
3. Hemnes AR, Leopold JA, Radeva MK, Beck GJ, Abidov A, Aldred MA, *et al.* Clinical Characteristics and Transplant-Free Survival Across the Spectrum of Pulmonary Vascular Disease. *J Am Coll Cardiol* 2022;80:697–718.
4. Jiang L, Wang M, Lin S, Jian R, Li X, Chan J, *et al.* A Quantitative Proteome Map of the Human Body. *Cell* 2020;183:269-283.e19.
5. Williams SA, Kivimaki M, Langenberg C, Hingorani AD, Casas JP, Bouchard C, *et al.* Plasma protein patterns as comprehensive indicators of health. *Nat Med* 2019;25:1851–1857.
6. Williams SA, Ostroff R, Hinterberg MA, Coresh J, Ballantyne CM, Matsushita K, *et al.* A proteomic surrogate for cardiovascular outcomes that is sensitive to multiple mechanisms of change in risk. *Sci Transl Med* 2022;14:eabj9625.
7. Sweatt AJ, Hedlin HK, Balasubramanian V, Hsi A, Blum LK, Robinson WH, *et al.* Discovery of Distinct Immune Phenotypes Using Machine Learning in Pulmonary Arterial Hypertension. *Circ Res* 2019;124:904–919.
8. Amsallem M, Sweatt AJ, Ataam JA, Guihaire J, Lecerf F, Lambert M, *et al.* Targeted proteomics of right heart adaptation to pulmonary arterial hypertension. *Eur Respir J* 2021;57:2002428.
9. Rhodes CJ, Wharton J, Swietlik EM, Harbaum L, Girerd B, Coghlan JG, *et al.* Using the Plasma Proteome for Risk Stratifying Patients with Pulmonary Arterial Hypertension. *Am J Respir Crit Care Med* 2022;205:1102–1111.
10. Harbaum L, Rhodes CJ, Wharton J, Lawrie A, Karnes JH, Desai AA, *et al.* Mining the Plasma Proteome for Insights into the Molecular Pathology of Pulmonary Arterial Hypertension. *Am J Respir Crit Care Med* 2022;doi:10.1164/rccm.202109-2106OC.
11. Wilkins MR. Personalized Medicine for Pulmonary Hypertension:: The Future Management of Pulmonary Hypertension Requires a New Taxonomy. *Clin Chest Med* 2021;42:207–216.
12. Galiè N, Humbert M, Vachiery J-L, Gibbs S, Lang I, Torbicki A, *et al.* 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Respir J* 2015;46:903–975.
13. Galiè N, Humbert M, Vachiery J-L, Gibbs S, Lang I, Torbicki A, *et al.* 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Heart J* 2016;37:67–119.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

14. Marmot M, Brunner E. Cohort Profile: the Whitehall II study. *Int J Epidemiol* 2005;34:251–256.
15. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B Methodol* 1996;58:267–288.
16. Rhodes CJ, Ghataorhe P, Wharton J, Rue-Albrecht KC, Hadinnapola C, Watson G, *et al.* Plasma Metabolomics Implicates Modified Transfer RNAs and Altered Bioenergetics in the Outcomes of Pulmonary Arterial Hypertension. *Circulation* 2017;135:460–475.
17. Boucly A, Tu L, Guignabert C, Rhodes C, De Groote P, Prévot G, *et al.* Cytokines as prognostic biomarkers in pulmonary arterial hypertension. *Eur Respir J* 2023;61:2201232.
18. Guignabert C, Savale L, Boucly A, Thuillet R, Tu L, Ottaviani M, *et al.* Serum and Pulmonary Expression Profiles of the Activin Signaling System in Pulmonary Arterial Hypertension. *Circulation* 2023;147:1809–1822.
19. Tiede SL, Wassenberg M, Christ K, Schermuly RT, Seeger W, Grimminger F, *et al.* Biomarkers of tissue remodeling predict survival in patients with pulmonary hypertension. *Int J Cardiol* 2016;223:821–826.
20. Arvidsson M, Ahmed A, Säleby J, Hesselstrand R, Rådegran G. Plasma matrix metalloproteinase 2 is associated with severity and mortality in pulmonary arterial hypertension. *Pulm Circ* 2022;12:e12041.
21. Jandl K, Radic N, Zeder K, Kovacs G, Kwapiszewska G. Pulmonary vascular fibrosis in pulmonary hypertension - The role of the extracellular matrix as a therapeutic target. *Pharmacol Ther* 2023;247:108438.
22. Jandl K, Marsh LM, Hoffmann J, Mutgan AC, Baum O, Bloch W, *et al.* Basement Membrane Remodeling Controls Endothelial Function in Idiopathic Pulmonary Arterial Hypertension. *Am J Respir Cell Mol Biol* 2020;63:104–117.
23. Humbert M. Viewpoint: activin signalling inhibitors for the treatment of pulmonary arterial hypertension. *Eur Respir J* 2023;62:2301726.
24. Perros F, Montani D, Dorfmüller P, Durand-Gasselin I, Tcherakian C, Le Pavec J, *et al.* Platelet-derived growth factor expression and function in idiopathic pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2008;178:81–88.
25. Hoeper MM, Barst RJ, Bourge RC, Feldman J, Frost AE, Galié N, *et al.* Imatinib mesylate as add-on therapy for pulmonary arterial hypertension: results of the randomized IMPRES study. *Circulation* 2013;127:1128–1138.
26. Frantz RP, McLaughlin VV, Sahay S, Escribano Subías P, Zolty RL, Benza RL, *et al.* Seralutinib in adults with pulmonary arterial hypertension (TORREY): a randomised, double-blind, placebo-controlled phase 2 trial. *Lancet Respir Med* 2024;12:523–534.
27. Humbert M, Guignabert C, Bonnet S, Dorfmüller P, Klinger JR, Nicolls MR, *et al.* Pathology and pathobiology of pulmonary hypertension: state of the art and research perspectives. *Eur Respir J* 2019;53:1801887.
28. Guignabert C, Humbert M. Targeting transforming growth factor-β receptors in pulmonary hypertension. *Eur Respir J* 2021;57:2002341.
29. Savale L, Tu L, Normand C, Boucly A, Sitbon O, Montani D, *et al.* Effect of Sotatercept on Circulating Proteomics in Pulmonary Arterial Hypertension. *Eur Respir J* 2024;2401483.doi:10.1183/13993003.01483-2024.
30. Humbert M, McLaughlin V, Gibbs JSR, Gomberg-Maitland M, Hoeper MM, Preston IR, *et al.* Sotatercept for the Treatment of Pulmonary Arterial Hypertension. *N Engl J Med* 2021;384:1204–1215.
31. Hoeper MM, Badesch DB, Ghofrani HA, Gibbs JSR, Gomberg-Maitland M, McLaughlin

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

VV, *et al.* Phase 3 Trial of Sotatercept for Treatment of Pulmonary Arterial Hypertension. *N Engl J Med* 2023;388:1478–1490.
32. Souza R, Badesch DB, Ghofrani HA, Gibbs JSR, Gomberg-Maitland M, McLaughlin VV, *et al.* Effects of sotatercept on haemodynamics and right heart function: analysis of the STELLAR trial. *Eur Respir J* 2023;62:2301107.
33. Sanges S, Rice L, Tu L, Valenzi E, Cracowski J-L, Montani D, *et al.* Biomarkers of Hemodynamic Severity of Systemic Sclerosis-Associated Pulmonary Arterial Hypertension by Serum Proteome Analysis. *Ann Rheum Dis* 2022;ard-2022-223237.doi:10.1136/ard-2022-223237.
34. Berrone E, Chiorino G, Guana F, Benedetti V, Palmitessa C, Gallo M, *et al.* SOMAscan Proteomics Identifies Novel Plasma Proteins in Amyotrophic Lateral Sclerosis Patients. *Int J Mol Sci* 2023;24:1899.
35. Ciampa E, Li Y, Dillon S, Lecarpentier E, Sorabella L, Libermann TA, *et al.* Cerebrospinal Fluid Protein Changes in Preeclampsia. *Hypertension* 2018;72:219–226.
36. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* 2018;361:769–773.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure Legends

Figure 1: Overview of study design. Circulating levels of 7288 proteins were assayed (SomaScan 7K platform) in plasma samples from 470 PH patients, irrespective of clinical pulmonary hypertension subgroup, 136 disease controls and 59 healthy controls enrolled as a discovery cohort. Proteins that distinguished pulmonary hypertension from both control groups were selected for unsupervised clustering (k-means clustering of UMAP dimensions). Separate cohorts of serially sampled patients from the United Kingdom (n=229) and France (n=79) provided independent validation of the clusters and dynamic association with clinical status. Enrichment analysis was used to identify key molecular pathways in each cluster. PH: pulmonary hypertension; PAH: pulmonary arterial hypertension; PH-LHD: PH associated with left heart disease; PH-lung: PH associated with lung disease; CTEPH: chronic thrombo-embolic PH; HC: healthy controls; No PH: symptomatic disease controls without PH.

Figure 2: Methodology used to select somamers used for clustering analysis.
Proteins that distinguished PH patients (in at least one etiological diagnostic group) from both healthy controls and NoPH controls (in models corrected for age, sex, principal component outliers, haemolysis, coagulation Factor X and cystatin C) were used for clustering analysis. Venn diagrams indicate the overlap of proteins identified in each analysis run and the final selection of 165 SOMAmers measuring 156 unique proteins.

Figure 3: Proteomic clusters. Uniform Manifold Approximation and Projection (UMAP) of the 156 proteins used for clustering analysis (A) and Kaplan-Meier survival curves according to clusters in the discovery cohort (B) and UK (C = baseline and D = first follow-up visits) and French PAH (E = baseline and F = first follow-up visits) validation cohorts. **A:** Each color corresponds to a cluster identified by k-means clustering analysis. **B:** survival curves of patients classified in cluster 1 (purple), 2 (green), 3 (red), 4 (blue). Log rank test, $p<0.001$. **C, D, E, F:** survival curves of patients classified in cluster 1 (purple), 2 or 3 (dark blue), 4 (blue). Log rank test, $p<0.001$ for each analysis (C, D, E, F).

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure 4: Sankey diagrams showing cluster changes over time in UK (A) and French (B) cohorts and association with survival (C, D). A: In the UK cohort, 36% patients changed cluster over time. **B**: In the French cohort 38% patients changed cluster over time. **C**: Survival of UK patients in clusters 2 or 3 according to cluster changes over time (stable or improvement in dark blue, worsening to cluster 1 in red). Log rank test, p<0.001. **D**: Survival of UK patients in cluster 1 according to cluster changes over time (improvement in light blue, stable in purple). Log rank test, p=0.006.

Figure 5: Enrichment of Platelet-Derived Growth Factor (**PDGF) pathway in cluster 3.** Heatmap (**A**) and levels of PDGF-BB according to clusters in discovery cohort (**B**) and UK validation PAH cohort (**C**). *Abbreviations:* K: cluster; RFU: Relative Fluorescence Unit**.** *Statistics:* (B) non-paired ANOVA test, p<0.001. All Dunnett's pairwise comparisons vs cluster 3 (K3), q<0.001. (C) non-paired ANOVA test, p<0.001. All Dunnett's pairwise comparisons vs cluster 3 (K3), q<0.001.

Figure 6: Enrichment of TGF beta pathway cluster 1. Heatmap (**A**) and levels of Activin A (**B, C**) and follistatin (**D, E**) according to clusters in discovery and UK validation PAH cohorts, respectvely. *Abbreviations:* FSTL1: follistatin; K: cluster; RFU: Relative Fluorescence Unit. *Statistics:* (B) non-paired ANOVA test, p<0.001. Dunnett's pairwise comparisons K1 vs K3 and K1 vs K4, q<0.001. (C) non-paired ANOVA test, p=0.019. Dunnett's pairwise comparison K1 vs K4, q=0.009. (D) non-paired ANOVA test, p<0.001. All Dunnett's pairwise comparisons vs K1, q<0.001. (E) non-paired ANOVA test, p<0.001. All Dunnett's pairwise comparisons vs K1, q<0.001.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Table 1: Demographics, clinical and hemodynamic characteristics of the study population

	Healthy controls, N=59	**No PH controls, n=136**	**Pulmonary hypertension, N=470**
Sex Female / Male, n (%)	41(69) / 18(31)	87(64) / 49(36)	262(56) / 208(44)
Age, years	46 ± 12	61 ± 16	64 ± 16
Race, n (%) Caucasian African Asian No data	 38 (64) 2 (3) 8 (14) 11 (19)	 92 (68) 15 (11) 8 (6) 21 (15)	 348 (74) 28 (6) 30 (6) 64 (14)
Treatment naïve patients, n (%)	59 (100)	136 (100)	379 (81)
Systemic hypertension, n (%)	0	70 (51)	147 (31)
Diabetes mellitus, n (%)	0	16 (12)	68 (14)
Ischaemic heart disease, n (%)	0	6 (4)	22 (5)
Atrial fibrillation permanent, n (%)	0	15 (11)	68 (14)
Thyroid disease, n (%)	0	1 (1)	18 (4)
COPD, n (%)	0	9 (7)	45 (10)
No comorbidity, n (%)	59 (100)	34 (25)	139 (30)
Time between diagnosis and sample, years (IQR)	na	0 (0-0)	0 (0-0)
NYHA FC I-II / III / IV, n (%)	na	47 (35) / 85 (63) / 3 (2)	80 (17) / 347 (74) / 41 (9)
6MWD, m	na	312 ± 139	240 ± 147
BNP, ng/L	na	48 (16-141)	166 (57-440)
RAP, mmHg	na	8 ± 4	10 ± 5
mPAP, mmHg	na	22 ± 9	43 ± 10
PAWP, mmHg	na	12 ± 4	12 ± 6
Cardiac output, L/min	na	6.4 ± 2.7	4.4 ± 1.8

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Cardiac index, L/min/m^2	na	3.3 ± 1.5	2.4 ± 0.9
PVR, WU	na	1.6 ± 0.9	8 ± 5
SvO2, %	na	76 ± 7	74 ± 13

Abbreviations: COPD: chronic obstructive pulmonary disease; NYHA FC: New York Heart Association functional class; 6MWD: 6-min walk distance; BNP: brain natriuretic peptide; RAP: right atrial pressure; mPAP: mean pulmonary arterial pressure; PAWP: pulmonary arterial wedge pressure; PVR: pulmonary vascular resistance; SvO2: mixed venous oxygen saturation; na: not applicable.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Table 2: Demographic, functional, exercise and hemodynamic characteristics of pulmonary hypertension according to clusters

	Cluster 1 N=105	Cluster 2 N=141	Cluster 3 N=59	Cluster 4 N=165	P-value
Sex Female / Male, n (%)	45 (43) / 60 (57)	81 (57) / 60 (43)	35 (59) / 24 (41)	101 (61) / 64 (39)	0.023
Age, years	69 ± 12	64 ± 16	67 ± 16	59 ± 16	<0.001
Race, n (%) Caucasian African Asian No data	 83 (79) 1 (1) 7 (7) 14 (13)	 97 (69) 11 (8) 13 (9) 20 (14)	 45 (76) 2 (3.5) 2 (3.5) 10 (17)	 123 (75) 14 (8) 8 (5) 20 (12)	0.18
Pulmonary arterial hypertension, n (%)	24 (23)	48 (34)	12 (20)	47 (28)	<0.001
PH associated with LHD, n (%)	38 (36)	43 (31)	14 (24)	27 (16)	
PH associated with lung disease, n (%)	26 (25)	16 (11)	17 (29)	34 (21)	
Chronic thromboembolic PH, n (%)	17 (16)	34 (24)	16 (27)	57 (35)	
Systemic hypertension, n (%)	32 (30)	57 (40)	17 (29)	41 (25)	0.031
Diabetes mellitus, n (%)	16 (15)	22 (16)	7 (12)	23 (14)	0.906
Ischaemic heart disease, n (%)	7 (7)	7 (5)	3 (5)	5 (3)	0.613
Atrial fibrillation permanent, n (%)	26 (25)	23 (16)	9 (15)	10 (6)	<0.001
Thyroid disease, n (%)	5 (5)	4 (3)	3 (5)	6 (4)	0.827
COPD, n (%)	16 (15)	12 (9)	6 (10)	11 (7)	0.127
No comorbidity, n (%)	20 (19)	41 (29)	17 (29)	61 (37)	0.019
Time between diagnosis and sample, years (IQR)	0 (0-0)	0 (0-0.2)	0 (0-0.1)	0 (0-0.2)	0.233
NYHA FC, n (%) I-II / III / IV	13 (12.5) 76 (72.5)/ 16 (15)	22 (15.5) 108 (76.5)/ 11 (8)	11 (19) 43 (73)/ 5 (8)	34 (21) 122 (74)/ 9 (5)	0.114 *0.010
6MWD, m	144 (48-288)	240 (96-337)	216 (96-342)	323 (144-408)	<0.001
BNP, ng/L (IQR)	713 (381-1177)	210 (134-356)	227 (63-571)	47 (19-95)	<0.001
RAP, mmHg	13 ± 5	10 ± 5	11 ± 5	8 ± 4	<0.001
mPAP, mmHg	44 ± 9	45 ± 12	43 ± 10	41 ± 13	0.011
PAWP, mmHg	14 ± 7	12 ± 5	14 ± 6	12 ± 5	0.030

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

CI, L/min/m^2	2.0 ± 0.7	2.4 ± 0.9	2.3 ± 0.9	2.6 ± 0.9	<0.001
PVR, WU	10 ± 5	9 ± 7	9 ± 6	7 ± 4	<0.001
SvO2, %	61 ± 12	66 ± 9	64 ± 14	70 ± 10	<0.001

Abbreviations: PH: pulmonary hypertension; CTEPH: chronic thromboembolic PH; COPD: chronic obstructive pulmonary disease**;** NYHA FC: New York Heart Association functional class; 6MWD: 6-min walk distance; BNP: brain natriuretic peptide; RAP: right atrial pressure; mPAP: mean pulmonary arterial pressure; PAWP: pulmonary arterial wedge pressure; CI: cardiac index; PVR: pulmonary vascular resistance; SvO2: mixed venous oxygen saturation.

* cluster 1 vs cluster 4

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

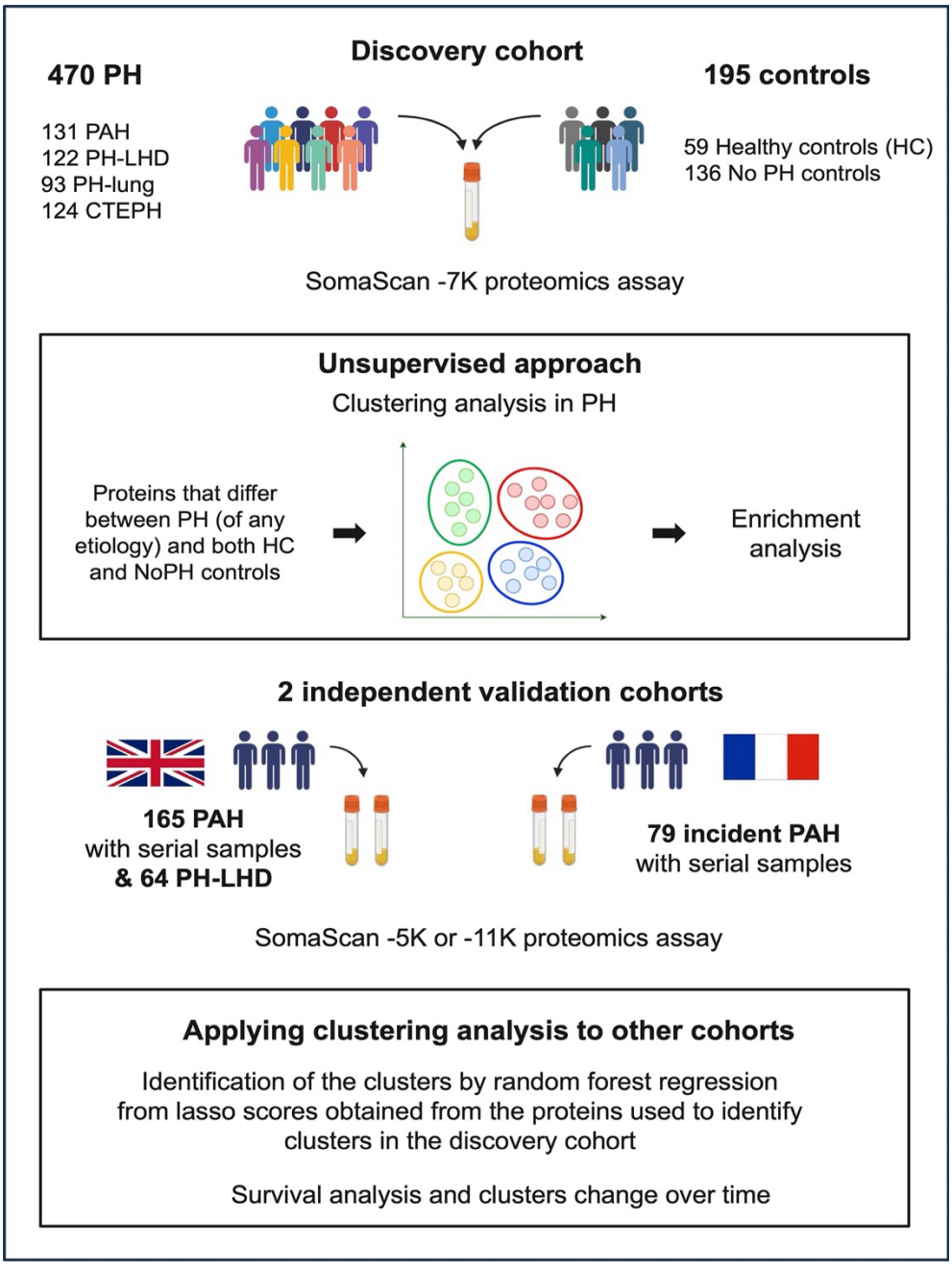


Figure 1: Overview of study design. Circulating levels of 7288 proteins were assayed (SomaScan 7K platform) in plasma samples from 470 PH patients, irrespective of clinical pulmonary hypertension subgroup, 136 disease controls and 59 healthy controls enrolled as a discovery cohort. Proteins that distinguished pulmonary hypertension from both control groups were selected for unsupervised clustering (k-means clustering of UMAP dimensions). Separate cohorts of serially sampled patients from the United Kingdom (n=229) and France (n=79) provided independent validation of the clusters and dynamic association with clinical status. Enrichment analysis was used to identify key molecular pathways in each cluster. PH: pulmonary hypertension; PAH: pulmonary arterial hypertension; PH-LHD: PH associated with left heart disease; PH-lung: PH associated with lung disease; CTEPH: chronic thrombo-embolic PH; HC: healthy controls; No PH: symptomatic disease controls without PH.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

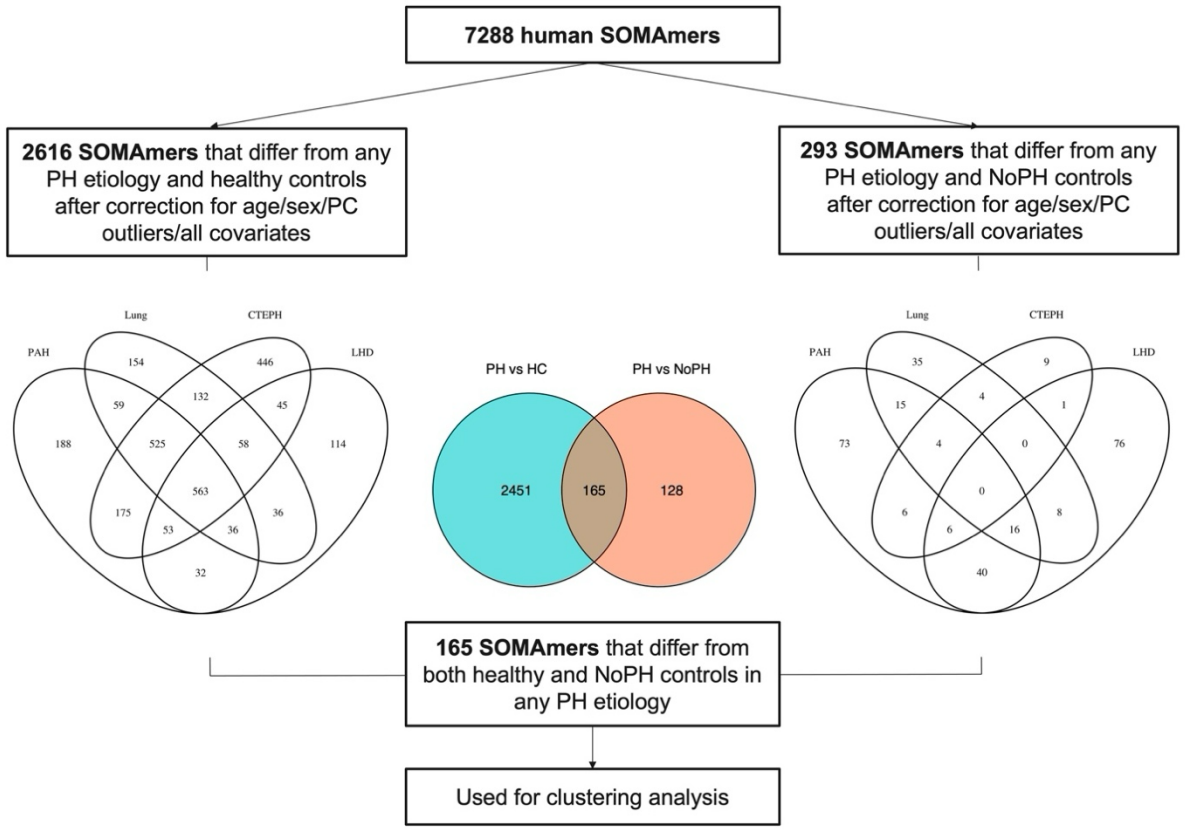


Figure 2: Methodology used to select somamers used for clustering analysis. Proteins that distinguished PH patients (in at least one etiological diagnostic group) from both healthy controls and NoPH controls (in models corrected for age, sex, principal component outliers, haemolysis, coagulation Factor X and cystatin C) were used for clustering analysis. Venn diagrams indicate the overlap of proteins identified in each analysis run and the final selection of 165 SOMAmers measuring 156 unique proteins.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

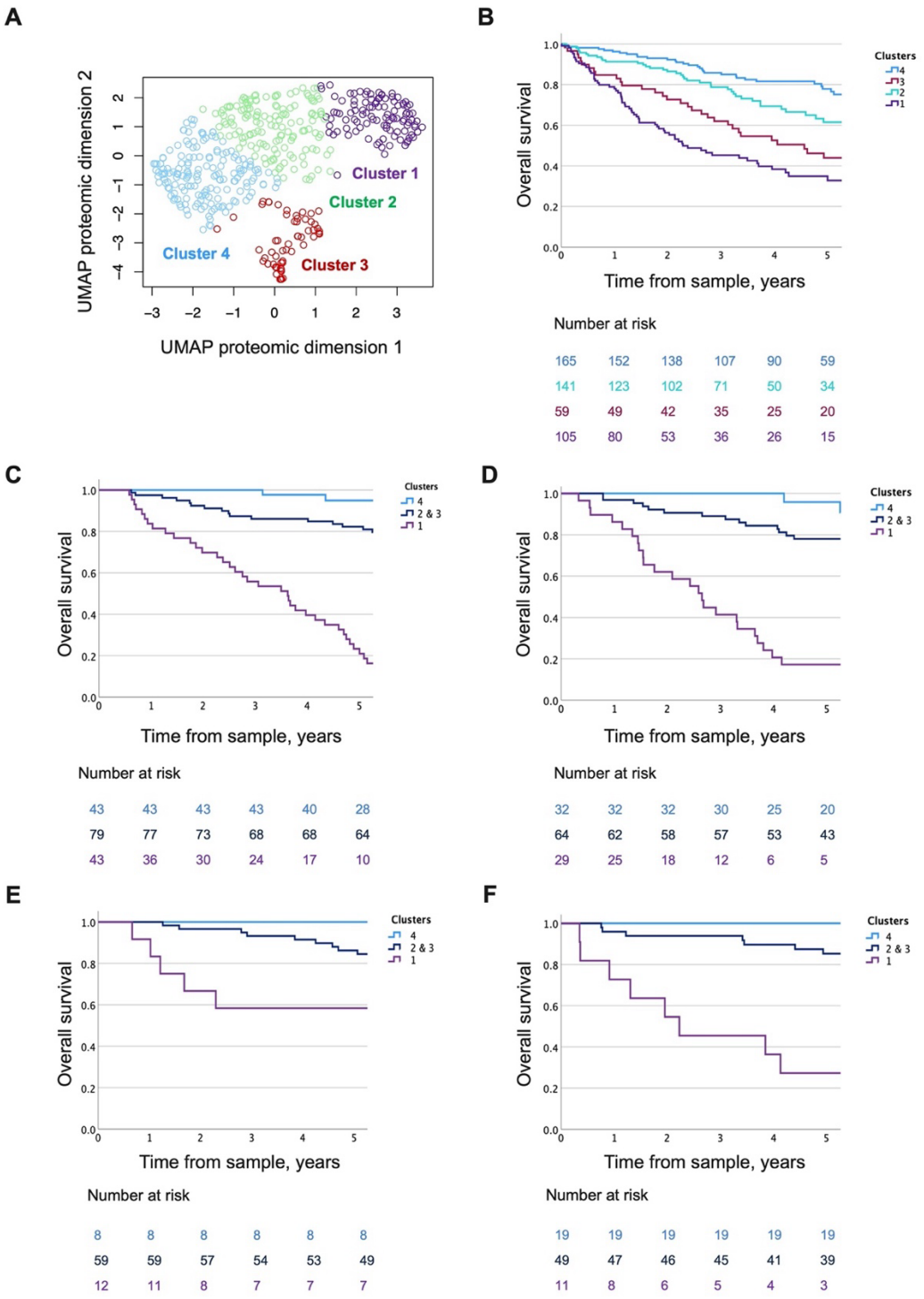


Figure 3: Proteomic clusters. Uniform Manifold Approximation and Projection (UMAP) of the 156 proteins used for clustering analysis (A) and Kaplan-Meier survival curves according to clusters in the discovery cohort (B) and UK (C = baseline and D = first follow-up visits) and French PAH (E = baseline and F = first follow-up visits) validation cohorts. A: Each color corresponds to a cluster identified by k-means clustering analysis. B: survival curves of patients classified in cluster 1 (purple), 2 (green), 3 (red), 4 (blue). Log rank test, $p<0.001$. C, D, E, F: survival curves of patients classified in cluster 1 (purple), 2 or 3 (dark blue), 4 (blue). Log rank test, $p<0.001$ for each analysis (C, D, E, F).

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

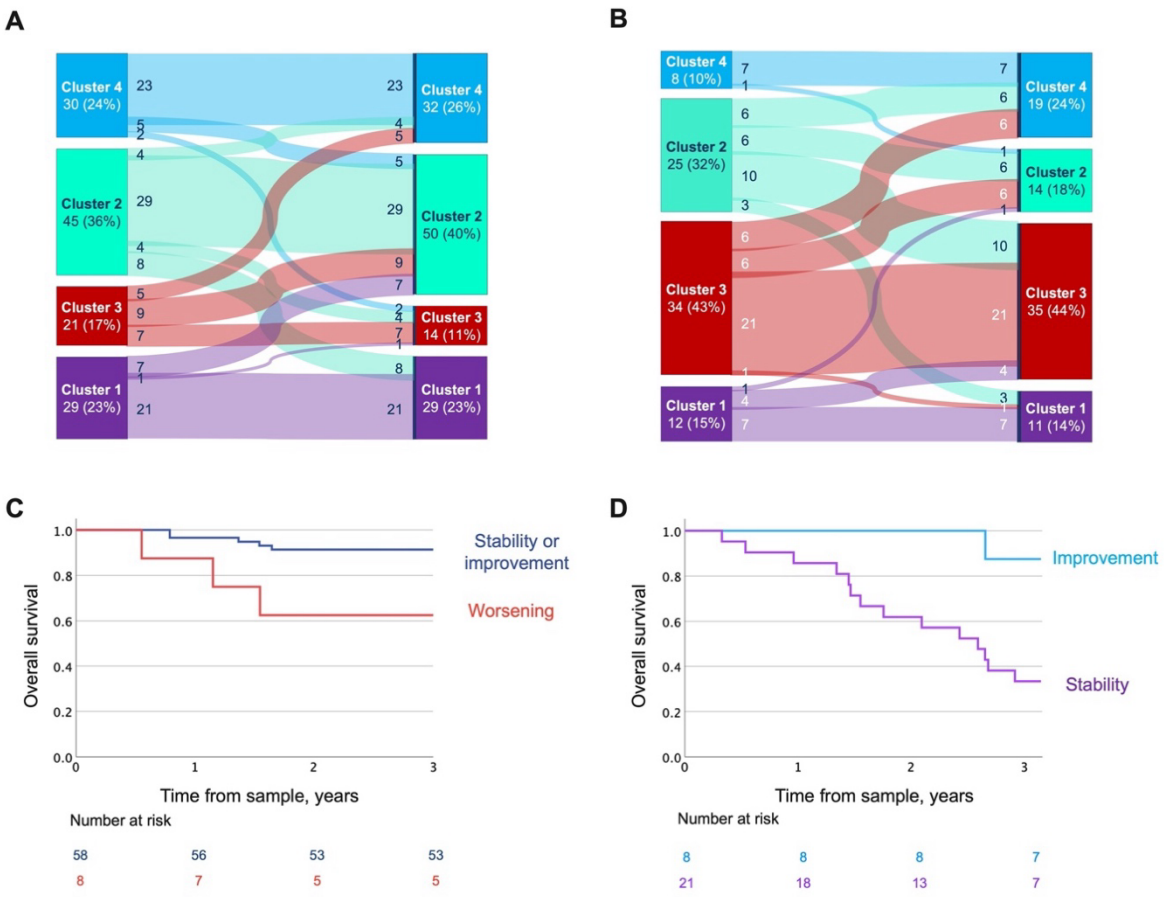


Figure 4: Sankey diagrams showing cluster changes over time in UK (A) and French (B) cohorts and association with survival (C, D). A: In the UK cohort, 36% patients changed cluster over time. **B**: In the French cohort 38% patients changed cluster over time. **C**: Survival of UK patients in clusters 2 or 3 according to cluster changes over time (stable or improvement in dark blue, worsening to cluster 1 in red). Log rank test, p<0.001. **D**: Survival of UK patients in cluster 1 according to cluster changes over time (improvement in light blue, stable in purple). Log rank test, p=0.006.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

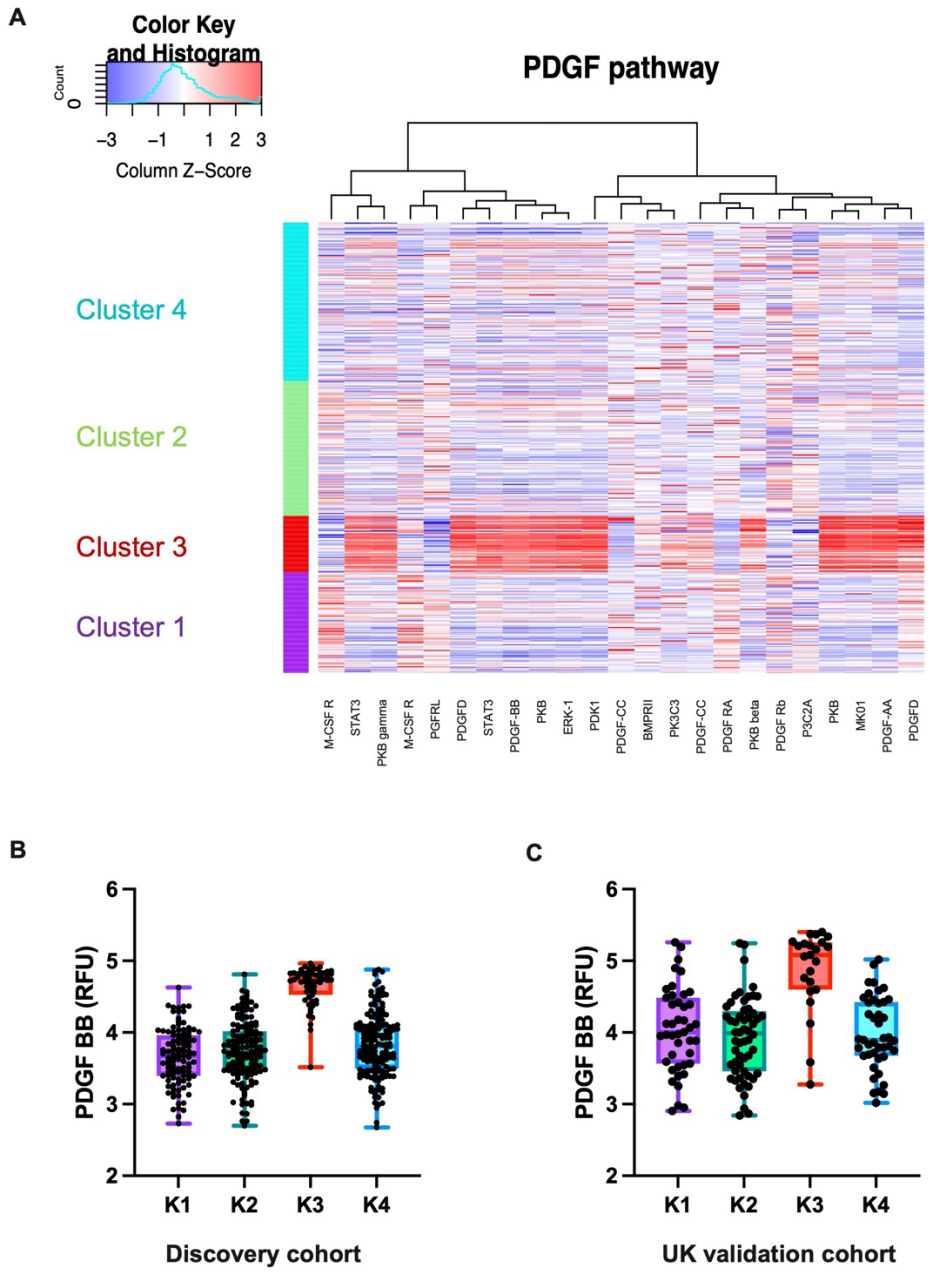


Figure 5: Enrichment of Platelet-Derived Growth Factor (PDGF) pathway in cluster 3. Heatmap (**A**) and levels of PDGF-BB according to clusters in discovery cohort (**B**) and UK validation PAH cohort (**C**). *Abbreviations:* K: cluster; RFU: Relative Fluorescence Unit**.** *Statistics:* (B) non-paired ANOVA test, p<0.001. All Dunnett's pairwise comparisons vs cluster 3 (K3), q<0.001. (C) non-paired ANOVA test, p<0.001. All Dunnett's pairwise comparisons vs cluster 3 (K3), q<0.001.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

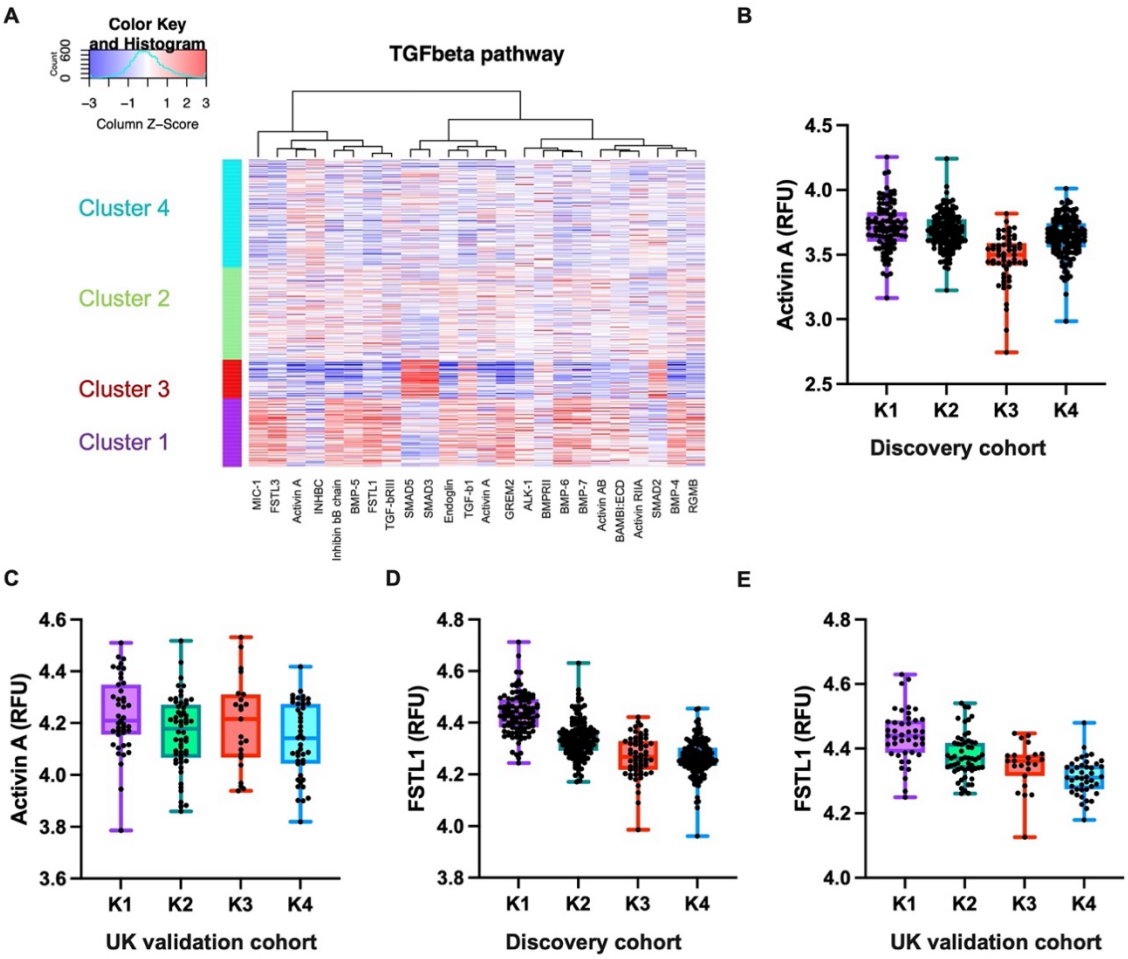


Figure 6: Enrichment of TGF beta pathway cluster 1. Heatmap (**A**) and levels of Activin A (**B, C**) and follistatin (**D, E**) according to clusters in discovery and UK validation PAH cohorts, respectvely. *Abbreviations:* FSTL1: follistatin; K: cluster; RFU: Relative Fluorescence Unit. *Statistics:* (B) non-paired ANOVA test, p<0.001. Dunnett's pairwise comparisons K1 vs K3 and K1 vs K4, q<0.001. (C) non-paired ANOVA test, p=0.019. Dunnett's pairwise comparison K1 vs K4, q=0.009. (D) non-paired ANOVA test, p<0.001. All Dunnett's pairwise comparisons vs K1, q<0.001. (E) non-paired ANOVA test, p<0.001. All Dunnett's pairwise comparisons vs K1, q<0.001.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Clustering Pulmonary Hypertension Patients Using the Plasma Proteome

Athénaïs Boucly, MD, PhD, Shanshan Song, PhD, Merve Keles, PhD, Dennis Wang, PhD, Luke S. Howard, MD, DPhil, FRCP, Marc Humbert, MD, PhD, Olivier Sitbon, MD, PhD, Allan Lawrie, PhD, A A Roger Thompson, Philipp Frank, PhD, Mika Kivimaki, FMedSci, Christopher J. Rhodes, PhD, Martin R. Wilkins, DSc FMedSci

ONLINE DATA SUPPLEMENT

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

I. Extended Methods

Participants

The discovery study population comprised patients with suspected PH who attended a specialist clinic at Imperial College NHS Trust between 2013 and 2021. All patients were managed according to the ESC/ERS guidelines.[12,13] Patients with PH, defined by a mean pulmonary artery pressure ≥25mmHg, were classified in Group 1 (PAH), Group 2 (PH associated with left heart disease, PH-LHD), Group 3 (PH associated with lung disease, PH-lung) or Group 4 (chronic thrombo-embolic PH, CTEPH).[12,13] Patients referred with suspected PH but with a mean pulmonary artery pressure <25mmHg on right heart catheterisation were classified as symptomatic disease controls. Contemporaneous plasma samples were obtained from volunteers without cardiovascular or respiratory diseases who acted as health controls. All patients were recruited with informed written consent and local research ethics committee approval (11/LO/0395 and 17/LO/0563).

Separate cohorts of PH patients with serial plasma samples collected over the same time period were used for independent validation: the UK National Cohort Study (NCT01907295); the French EFORT study: Evaluation of Prognostic Factors and Therapeutic Targets in PAH (NCT01185730); and the Sheffield Teaching Hospitals Observational Study of patients with PH, Cardiovascular or Respiratory Disease (18/YH/0441). The Whitehall II study[14] provided a dataset based on samples collected from a large cohort that were healthy at baseline to understand the behaviour of PH-associated clusters in a population cohort.

Sample collection and processing

With the exception of the Whitehall II cohort, patients were sampled non-fasted at their routine clinical appointment visits. All samples were taken from peripheral veins. Serial samples were available in 125 patients from the UK PAH Cohort and 79 patients from the EFORT cohort. The median follow-up time between the first and the second sample was 12.1 (11.0 – 13.5) months in the UK PAH cohort and 4.6 (3.9 – 7.3) months in the EFORT cohort. The EFORT cohort included only newly diagnosed patients with PAH. Patients were therefore treatment naïve at time of first sample. Patients were treated as follows: Calcium channel blockers n=5, oral monotherapy n=33, oral dual therapy n=33, and initial triple combination

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

therapy n=8. Among the 125 patients from the UK PAH Cohort with serial samples, there were only 17 treatment escalation (11 to dual therapy and 6 to triple therapy).

Plasma EDTA samples were stored at -80^{o}C and shipped to SomaLogic (Boulder, CO, USA) for SomaScan proteomic analysis. Samples from the discovery cohort were assayed using the 7K platform (comprising 7335 Somamers targeting 7288 human proteins). Proteomic analysis of the 2 independent validation cohorts used the SomaScan version 4 assay (which measures 4979 human Somamers). In Whitehall II, both 7k and 4.0 assays were used. In all studies, technicians were blinded to patient status. Relative fluorescence units were log-10 scale transformed to normalize protein levels prior to analysis.

Statistical analyses

A) Supervised approach to identify PH specific proteins

Patients and controls from the UK discovery cohort were randomized into training (80%) and replication groups (20%) to adequately power discovery analysis of all proteins and replication of proteins meeting statistical significance. To ensure the reproducibility of the random analyses, the random seed value was fixed using the *set.seed(123)* function.

Principal component analysis was performed to evaluate the variation in protein expression profiles and to identify patterns of variation across the samples. Proteins levels were compared between PH patients and (healthy and No-PH) controls by logistic regression models, correcting for age, sex and principal component outliers (**Figure S1**). Sensitivity analyses were performed to confirm that protein differences were independent of haemolysis (cell-free haemoglobin as a covariate), coagulation factor X, renal function (cystatin C). All comparisons were corrected for multiple testing using Benjamini-Hochberg false discovery rate (FDR). A threshold of $q<0.05$ was considered statistically significant.

A LASSO approach was applied to all PH-specific proteins (previously identified) to reduce the number of proteins of interest and identify the optimal combination for predicting PH diagnosis. This modeling approach used 10-fold cross-validation, with the regularization parameter (lambda) determined by the lowest error plus 1 standard error (to minimize overfitting), implemented with the *glmnet* R-package.[15] Similar analyses were performed in

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

the dataset of proteins statistically different between patients with PH and controls to identify the combination of proteins that best reflected PH pathology. Receiver operating characteristic (ROC) analyses of the different protein combinations were performed using the *pROC* R-package in the replication group of our dataset, then compared to the performance of N-terminal pro-brain natriuretic peptide (NT-proBNP) using DeLong test.

B) Unsupervised approach:

B.1: clustering analysis based on proteomic profile

Proteins able to identify PH patients from both healthy controls and No-PH controls (in models corrected for age, sex, principal component outliers, haemolysis, coagulation Factor X and cystatin C) were taken forward for clustering analysis (**Figures 1 and 2**). The dimensions of the dataset (comprising the previously identified proteins) were reduced via the Uniform Manifold Approximation and Projection (UMAP) method using *UMAP* R-package, and the derived UMAP dimensions were then used for clustering. We used the *NbClust* R-package which determines the optimal number of clusters (based on the proteomic profile) with the highest stability by varying all combinations of number of clusters (from 2 to 10), distance measures, and clustering methods.

B.2: Classifying samples based on cluster membership

We classified samples based on the proteome-based clusters. LASSO regression was first performed to reduce the number of proteins needed to define the clusters. A Random forest classifier (*caret* and *randomForest* R-packages) from LASSO scores was trained to predict the cluster membership of new samples and used to classify samples from other cohorts.

B.3: Clinical differences between clusters

Demographic and clinical differences between the different clusters were assessed by non-paired ANOVA or Kruskal-Wallis tests according to the data distribution and chi-squared tests. We compared survival of the different clusters by log-rank test, from plasma sampling to death or censoring. Survival status for PH patients was censored on December 31, 2022. Overall survival was represented using the Kaplan–Meier method. To check whether our results were consistent with previous studies[7,9,10,14–18], we identified biomarkers known to be associated

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

with prognosis in PAH on a volcano plot showing plasma levels of proteins in cluster with the worst survival.

B.4: Enrichment analysis

Molecular enrichment analysis was performed using the *WebGestaltR* R-package to identify up-and down-regulated pathways of each cluster. Heatmaps of proteins within pathways of PAH drugs in development were performed using *gplots* and *pheatmap* R-packages. The relative fluorescence of proteins of interest in the different clusters were compared by non-paired ANOVA tests with Dunnett's multiple pairwise comparisons.

C) Cluster performance in the general population

To evaluate the ability of the clusters to identify participants who would develop PH in an initially healthy population, we assessed the cumulative incidence for participants in each cluster during follow-up. After confirming the proportional hazards assumption, we computed hazard ratios and 95% confidence intervals for membership in a cluster compared to absence at baseline and incident PH at follow-up using Cox proportional hazards models adjusted for age, sex, and ethnicity. To quantify the predictive performance of clusters associated with incident PH, we calculated conventional predictive statistics, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Statistical analysis was performed in R (version 4.3.1) and SPSS (version 29; IBM). Continuous variables are expressed as mean with standard deviation or median (interquartile range (IQR)) according to the data distribution.

An overview of the full methodology is displayed in **Figure 1**.

II. Supplemental Results

Plasma proteome differences between PH and controls

First we used logistic regression modelling to find 2616 SOMAmers where circulating levels distinguished PH from healthy controls and 293 that distinguished PH from No-PH (FDR $q<0.05$, **Figure S2A**). Similar analyses were applied to each clinical PH subgroup; specifically,

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

(i) PAH and healthy controls (2637 SOMAmers) or No-PH (1083); (ii) PH-LHD and healthy controls (1971) or No-PH (830); (iii) PH-lung and healthy controls (2010) or No-PH (717); and (iv) CTEPH and healthy controls (2696) or No-PH (711) (**Figure S3**). In sum, one thousand and eight unique SOMAmers were differentially expressed between both healthy and No-PH controls and at least one subgroup of PH, in models corrected for age, sex and principal component outliers (**Figure S1**, **Table S6**). We applied Fisher's exact test to evaluate the statistical significance of the shared proteins between multiple group comparisons (e.g., PAH vs PH-LHD, PAH vs PH-lung, PAH vs CTEPH, PH-LHD vs PH-lung, etc.), with all p-values found to be < 0.001.

Reducing 1008 proteins to concise sets associated with PH and clinical PH subgroups

Next we used lasso regression to identify a more concise combination of 25 proteins that differentiated PH from healthy controls and 40 proteins distinguishing PH from No-PH patients in a training group and demonstrated good sensitivity and specificity in recognising PH when applied to the replication cohort (**Table S7**, **Figure S2B and S2C**); AUC: 0.997 (0.989-1.000), p<0.001 vs healthy controls, 0.722 (0.621-0.823), p=0.001 vs No-PH. The diagnostic performance of these protein combinations outperformed NT-proBNP in distinguishing PH from healthy controls (0.913 [0.856-0.970], DeLong test p=0.006, **Figure S2B**) and performed at least as well as NT-proBNP in distinguishing PH from No-PH (AUC NT-proBNP: 0.658 [0.546-0.770], p=0.013; DeLong test=0.206, **Figure S2C**).

A similar analysis was performed using the 1008 SOMAmers differentially expressed between controls and any PH aetiology to identify the main clinical PH groups (Group 1, 2, 3 or 4, **Figure S4**). Lasso regressions to predict PAH, PH-LHD, PH-lung and CTEPH produced models comprised of 17, 35, 40 and 29 SOMAmers, respectively. These models performed well in identifying PH-LHD, PH-lung or CTEPH among patients with PH in the replication cohort: AUC PH-LHD 0.747 (0.609-0.885), p=0.001; AUC PH-lung 0.745 (0.633-0.857), p<0.001; AUC CTEPH 0.768 (0.663-0.872), p=0.005, respectively (**Table S8**, **Figure S5**). The combination of these 3 models was able to identify patients with PAH by elimination (**Figure S6**): AUC 0.684 (0.552-0.815), p=0.007.

PH prediction by cluster proteins in a population cohort

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

The Whitehall II study provided the opportunity to investigate the performance of the proteins used for cluster analysis in the general population. We hypothesised that the clusters associated with intermediate-high risk PH would be poorly detected in this cohort. Of the 6196 Whitehall II participants with valid protein data, only 2 (0.032% vs 22.3% in PH) belonged to cluster 1 while clusters 2 (n = 213, 3.4% vs 30% in PH), and 3 (n = 527, 8.5% vs 12.6% in PH) were uncommon and cluster 4 represented the majority (n = 5454, 88% vs 35% in PH, **Figure S15**). During the mean follow-up of 19.8 years, 57 (0.92%) participants were hospitalised with a diagnosis of PH (ICD10-code I27.0, I27.2, or I27.9). The cumulative hazard of developing PH was higher in cluster 2 than in clusters 3 and 4, with the separation in hazard curves between these groups beginning 7 years after baseline (**Figures S16 and S17**). The age-, sex- and ethnicity-adjusted hazard ratio for individuals in cluster 2 versus other participants was 2.35 (95% CI 0.93–5.93), but predictive capacity was poor (sensitivity 8.8%, specificity 96.6%, PPV 2.3%, NPV 99.1%, **Table S11**).

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

III. Supplemental tables

Table S1: Demographics, clinical and hemodynamic characteristics of patients with PH

	Pulmonary arterial hypertension, N=131	**PH associated with left heart disease, N=122**	**PH associated with lung disease, N=93**	**Chronic thromboembolic PH, N=124**
Sex Female / Male, n (%)	89 (68) / 42 (32)	71 (58) / 51 (42)	45 (48) / 48 (52)	57 (46) / 67 (54)
Age, years	58 ± 18	70 ± 11	65 ± 12	62 ± 17
Treatment naïve patients, n (%)	74 (56)	120 (98)	79 (85)	106 (85)
Systemic hypertension, n (%)	41 (31)	46 (38)	27 (29)	33 (27)
Diabetes mellitus, n (%)	10 (8)	25 (20)	24 (26)	9 (7)
Ischaemic heart disease, n (%)	8 (6)	3 (2)	6 (6)	5 (4)
Atrial fibrillation permanent, n (%)	7 (5)	41 (34)	11 (12)	9 (7)
Thyroid disease, n (%)	6 (5)	4 (3)	4 (4)	4 (3)
No comorbidity, n (%)	44 (34)	32 (26)	23 (25)	40 (32)
Subdiagnosis, n (%) Idiopathic PAH Heritable PAH Drugs associated PAH CTD CHD Portal hypertension Other	 39 (30) 4 (3) 1 (1) 42 (32) 23 (17) 12 (9) 10 (8)	na	na	na
Time between diagnosis and sample, years (IQR)	0 (0 – 0.9)	0 (0 – 0)	0 (0 – 0)	0 (0 – 0)
NYHA FC I-II / III / IV, n (%)	27 (21) / 89 (68) / 15 (11)	21 (17) / 93 (76) / 8 (7)	10 (11) / 69 (75) / 13 (14)	22 (18) / 96 (78) / 5 (4)
6MWD, m	306 (120 – 397)	192 (96 – 336)	144 (95 – 281)	288 (144 – 375)
BNP, ng/L	127 (46 – 370)	224 (112 – 468)	131 (44 – 596)	139 (51 – 350)
RAP, mmHg	9 ± 4	13 ± 5	10 ± 5	9 ± 5
mPAP, mmHg	47 ± 12	38 ± 9	42 ± 10	42 ± 12

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

PAWP, mmHg	10 ± 3	20 ± 6	12 ± 5	11 ± 3
Cardiac output, L/min	4.1 ± 1.8	4.5 ± 2.0	4.5 ± 1.5	4.6 ± 1.9
Cardiac index, L/min/m^2	2.3 ± 0.9	2.3 ± 0.9	2.4 ± 0.7	2.4 ± 0.9
PVR, WU	10 ± 6	5 ± 3	8 ± 4	8 ± 5
SvO2, %	77 ± 12	75 ± 11	68 ± 17	72 ± 11
PAH targeted therapies, n (%) CCB Oral monotherapy Oral dual therapy Dual therapy including PGI2 Triple therapy No data	 3 (2) 35 (27) 48 (37) 4 (3) 19 (14) 22 (17)	na	na	na

Abbreviations: CCB: calcium channel blockers; CTD: connective tissue disease; CHD: congenital heart disease; NYHA FC: New York Heart Association functional class; 6MWD: 6-min walk distance; BNP: brain natriuretic peptide; RAP: right atrial pressure; mPAP: mean pulmonary arterial pressure; PAWP: pulmonary arterial wedge pressure; PGI2: prostacyclin analog; PVR: pulmonary vascular resistance; SvO2: mixed venous oxygen saturation. na: not applicable.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Table S2: Demographics and clinical characteristics of the discovery cohort, subdivided into training (80%) and replication groups (20%)

	Training N=532			**Replication N=133**		
	Healthy controls, N=46	**NoPH controls, N=109**	**Patients with PH, N=377**	**Healthy controls, N=13**	**NoPH controls, N=27**	**Patients with PH, N=93**
Sex Female / Male, n (%)	32 (70) / 14 (30)	69 (63) /40 (37)	201 (53) / 176 (47)	9 (69) / 4 (31)	18 (67) / 9 (33)	61 (66) / 32 (34)
Age, years	47 ± 12	60 ± 16	64 ± 15	45 ± 13	65 ± 14	63 ± 17
Systemic hypertension, n (%)	0 (0)	53 (49)	117 (31)	0 (0)	17 (63)	30 (32)
Diabetes mellitus, n (%)	0 (0)	13 (12)	55 (15)	0 (0)	3 (11)	13 (14)
Ischaemic heart disease, n (%)	0 (0)	4 (4)	21 (6)	0 (0)	2 (7)	1 (1)
Atrial fibrillation permanent, n (%)	0 (0)	11 (10)	58 (15)	0 (0)	4 (15)	10 (11)
Thyroid disease, n (%)	0 (0)	1 (1)	13 (3)	0 (0)	0 (0)	5 (5)
COPD, n (%)	0 (0)	6 (6)	36 (10)	0 (0)	3 (11)	9 (10)
No comorbidity, n (%)	46 (100)	29 (27)	108 (29)	13 (100)	5 (19)	31 (33)
Aetiology of PH, n (%) **PAH** **PH-LHD** **PH-lung** **CTEPH**	na	na	 106 (28) 104 (28) 70 (18.5) 97 (25.5)	na	na	 25 (27) 18 (19) 23 (25) 27 (29)
NYHA FC I-II / III / IV, n (%)	na	37 (34)/ 69 (63) / 3 (3)	66 (18) / 277 (73) / 34 (9)	na	10 (37) / 17 (63) / 0 (0)	14 (15) / 72 (77) / 7 (8)
BNP, ng/L	na	48 (13 – 146)	183 (61 – 438)	na	44 (26 – 117)	134 (54 – 485)
RAP, mmHg	na	8 ± 5	10 ± 5	na	6 ± 3	11 ± 5
mPAP, mmHg	na	23 ± 9	43 ± 12	na	19 ± 3	42 ± 12
PAWP, mmHg	na	12 ± 4	12 ± 6	na	12 ± 3	12 ± 6
Cardiac output, L/min	na	6.6 ± 2.8	4.4 ± 1.9	na	5.2 ± 1.8	4.5 ± 1.8
Cardiac index, $L/min/m^2$	na	3.4 ± 1.5	2.4 ± 0.9	na	2.5 ± 0.9	2.4 ± 1.0
PVR, WU	na	1.6 ± 1.0	8.5 ± 6.0	na	1.4 ± 0.5	7.7 ± 5.0
SvO2, %	na	76 ± 8	65 ± 11	na	74 ± 4	68 ± 11

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Abbreviations: COPD: chronic obstructive pulmonary disease; NYHA FC: New York Heart Association functional class; 6MWD: 6-min walk distance; BNP: brain natriuretic peptide; RAP: right atrial pressure; mPAP: mean pulmonary arterial pressure; PAWP: pulmonary arterial wedge pressure; PVR: pulmonary vascular resistance; SvO2: mixed venous oxygen saturation; na: not applicable.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Table S3: Demographics and clinical characteristics of the validation cohorts

	UK validation cohort of patients with PAH, N=165	**French validation cohort of incident patients with PAH, N=79**	**UK validation cohort of patients with PH-LHD, N=64**
Sex Female / Male, n (%)	114 (69) / 51 (31)	56 (71) / 23 (29)	40 (62.5) / 24 (37.5)
Age, years	51 ± 16	51 ± 18	70 ± 11
Aetiology of PAH, n (%) **Idiopathic** **Heritable** **Anorexigen**	 138 (83.5) 26 (16) 1 (0.5)	 53 (67) 16 (20) 10 (13)	na
Time between diagnosis and sample, years (IQR)	3.5 (1.4 – 7.3)	0 (0 – 0)	0 (0 – 0)
NYHA FC **I-II / III / IV,** n (%)	70 (42) / 77 (47) / 18 (11)	30 (38) / 43 (54) / 6 (8)	12 (19) / 49 (76) / 3 (5)
6MWD, m	329 ± 164	345 ± 143	225 ± 159
BNP, ng/L	57 (25 – 157)	117 (47 – 290)	na
NT-proBNP, ng/L	na	na	1125 (500 – 2254)
RAP, mmHg	9 ± 6	8 ± 5	12 ± 6
mPAP, mmHg	50 ± 16	51 ± 12	37 ± 12
PAWP, mmHg	11 ± 4	9 ± 3	20 ± 5
Cardiac output, L/min	4.4 ± 1.8	4.4 ± 1.2	4.6 ± 1.7
Cardiac index, L/min/m^2	2.3 ± 0.9	2.5 ± 0.6	2.3 ± 0.8
PVR, WU	11 ± 6	10 ± 4	2 ± 1
ESC/ERS 4 strata risk status, n(%) Low Intermediate-low Intermediate-high High	 42 (25) 62 (38) 45 (27) 16 (10)	 15 (19) 30 (38) 27 (34) 7 (9)	na
PAH targeted therapies, n (%) Calcium channel blockers Oral monotherapy Oral dual therapy Dual therapy including PGI2 Triple therapy No data	 9 (5) 39 (24) 77 (47) 5 (3) 22 (13) 13 (8)	 5 (6) 33 (41) 33 (41) 0 8 (10) 0	na

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Abbreviations: NYHA FC: New York Heart Association functional class; 6MWD: 6-min walk distance; BNP: brain natriuretic peptide; NT-proBNP: N-terminal pro-brain natriuretic peptide; RAP: right atrial pressure; mPAP: mean pulmonary arterial pressure; PAWP: pulmonary arterial wedge pressure; PGI2: prostacyclin analog; PVR: pulmonary vascular resistance; na: not applicable.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Table S4. Characteristics of the Whitehall II cohort (n=6196)

Sex Female / Male, n (%)	1775 (28.6) / 4421 (71.4)
Age, years Mean (SD)	55.7 (6.0)
Ethnicity White / non-White, n (%)	5670 (91.5) / 526 (8.5)
Follow-up Time, years Mean±SD	19.8 (3.7)
Incidence of PH at follow-up, n (rate per 10,000 person-years)	57 (4.6)

Table S5: Percentage of variance explained by each principal component

Principal component (PC)	Explained variance, %
PC 1	16 %
PC 2	7 %
PC 3	4.4 %
PC 4	3.4 %
PC 5	2.6%
PC 6	2.0 %
PC 7	1.6 %
PC 8	1.3 %
PC 9	1.2 %
PC 10	0.9 %

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Table S6: Differentially expressed proteins between PH and controls

	PAH	PH-LHD	PH-lung	CTEPH	Any form of PH
Versus healthy controls	2637	1971	2010	2696	3505
Versus No-PH controls	1083	830	717	711	2049
Versus both healthy and No-PH controls	538	451	300	351	**1008**

Table S7: Area under curve of ROC analysis testing the performance in training and validation groups of the combination of proteins obtained by lasso regression to identify PH from healthy controls (A) and PH from symptomatic controls (B)

	AUC	Confidence interval	p-value
(A) PH versus healthy controls			
Training group	1.000	1.000 – 1.000	<0.001
Validation group	0.997	0.989 – 1.000	<0.001
(B) PH versus symptomatic controls			
Training group	0.918	0.889 – 0.947	<0.001
Validation group	0.722	0.621 – 0.823	0.001

Table S8: Area under curve of ROC analysis testing the performance in training and validation groups of the combination of proteins obtained by lasso regression to identify PAH from other PH (A), PH-LHD from other PH (B), PH-lung from other PH (C) and CTEPH from other PH (D)

	AUC	Confidence interval	p-value
(A) PAH versus other PH			
Training group	0.851	0.807 – 0.894	<0.001
Validation group	0.625	0.497 – 0.752	0.067
(B) PH-LHD versus other PH			
Training group	0.910	0.876 – 0.944	<0.001
Validation group	0.747	0.609 – 0.885	0.001
(C) PH-lung versus other PH			
Training group	0.961	0.942 – 0.980	<0.001
Validation group	0.745	0.633 – 0.857	<0.001
(D) CTEPH from other PH			
Training group	0.884	0.845 – 0.922	<0.001
Validation group	0.768	0.663 – 0.872	0.005

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Table S9: Enrichment analysis showing significantly up- or down-regulated pathways depending on clusters.

		Pathway	Proteins	Enrichment ratio	FDR (or p-value *)
CLUSTER 1 (vs other clusters)	UP REGULATION	BMP signalling pathway	BMP4, BMP5, BMP6, FSTL1, FSTL3, GDF15, GREM2, ROR2	6.6	0.024
		extracellular matrix organization	collagen, cystatin C, fibulin 5, FLRT2, GAS6, MFAP4, MMP2, PRSS2, PXDN, TIMP1, TIMP2, TNC, TNFRSF1A	5	<0.001
		Response to growth factor	ANGPT2, BMP4, BMP5, BMP6, EPHA2, FGF23, FLRT2, FSTL1, FSTL3, GAS1, GAS6, GDF15, GREM2, LTBP4, NRP1, ROR2, TNC, VEGFD	2.9	0.024
	DOWN REGULATION	cell-cell adhesion mediated by cadherin	cadherin 3, cadherin 7, plasminogen, serpin F2, WNT3A	12.7	0.042
		negative regulation of blood coagulation	Factor XI, kallikrein B1, plasminogen, protein kinase cGMP-dependent 1, SERPINF2, vitronectin	9.4	0.042
CLUSTER 4 (vs other clusters)	UP REGULATION	reverse cholesterol transport	APOA1, APOA5, APOM, LIPG	23.8	0.003
		negative regulation of blood coagulation	FII, FXI, KLKB1, KNG1, PLG, PROC, SERPINF2	11.9	<0.001
		protein activation cascade	APCS, C8G, CFHR5, CPN2, FXI, FXIIIB, FII, FVII, FCN2, FCN3, KLKB1, KNG1	11.8	<0.001
		blood coagulation	FXI, FXIIIB, FII, FVII, KLKB1, KNG1, PLG, PROC, SERPINA10, SERPIND1, SERPINF2, SHH, WNT3A	5	<0.001
	DOWN REGULATION	extracellular matrix organization	ADAMTSL2, CCDC80, collagen, cystatin C, fibulin, limican, MFAP4, MMP2, NID1, PXDN, TGF beta, TIMP1, TNC, VWF	5.8	<0.001
CLUSTER 2 (vs cluster 3)	UP REGULATION	cell-cell adhesion via plasma-membrane adhesion molecules	ADGRL3, AMIGO1, AMIGO2, CADM1, CDH5, EFNA5, L1CAM, PTPRD, ROBO2, SLITRK1	4.5	0.014
		cell morphogenesis involved in differentiation	AMIGO1, ANTXR1, collagen, ephrin, fibulin 1, FLRT2, ISLR2, L1CAM, MERTK, NEO1, NRXN3, NRTK2, PTPRD, ROBO2, SEMA4C, SEMA6B, SLITRK1	3.5	<0.001
		regulation of cell development	CDH5, EFNA5, ENG, FBLN1, FLRT2, HSPA5, IL6ST, ISLR2, JAG1, L1CAM, NOTCH3, NTRK2, PRTG, PTPRD, ROBO2, SEMA4C, SEMA6B, SLITRK1, TIMP2	3.0	0.003

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

	DOWN REGULATION	positive regulation of cellular protein catabolic process	CSNK2A1, IFNG, MAPK9, MDM2, METTL3, OAZ1, PAFAH1B2, PTEN, RNF41, TNFAIP3, UBE2V2	3.9	0.244 * p-value <0.001

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Table S10: Coefficients obtained by lasso regression to predict the 4 clusters

	proteins	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	(Intercept)	-70.12417677	-9.442547021	40.36921283	39.19751096
1	HXK4	-1.093944806	-2.619288397	4.961869077	-1.248635874
2	Carbonyl reductase 3	-1.000018282	-0.294442589	2.759713087	-1.465252216
3	PLOD3	0.197246903	-0.891189542	1.339825578	-0.645882939
4	C9	0.152032843	-0.833053864	0.409167888	0.271853133
5	SDF-1	0.314415507	0.225576128	0.159494638	-0.699486273
6	IL-1 R4	0.253897789	-0.174672787	0.079731719	-0.158956722
7	SAA2	0.067860448	-0.015931861	0.026830666	-0.078759253
8	P4R3A	0.166896106	0.017578666	0.023026026	-0.207500798
9	CRP	0.008523857	-0.006136064	-0.001318054	-0.00106974
10	PTGD2	0.000108883	0.000435364	-0.002415925	0.001871678
11	MCTS1	-0.005203563	-0.00012525	-0.005568738	0.010897551
12	PRS57	0.034110627	-0.036287109	-0.009138822	0.011315303
13	COLL1	0.417512564	-0.387378905	-0.012414183	-0.017719476
14	Pancreatic alpha-amylase	-0.04526405	-0.160487476	-0.01350491	0.219256436
15	SCUB3	0.028035874	-0.078500269	-0.021834819	0.072299214
16	MIC-1	0.070248575	-0.008041477	-0.030990883	-0.031216215
17	WIF-1	0.058596773	-0.069371699	-0.032656765	0.043431691
18	Pseudocholinesterase	-0.288041941	-0.258408037	-0.040651767	0.587101745
19	MFAP4	0.074109492	0.02324122	-0.040868833	-0.056481878
20	STX2	0.088011071	-0.034635721	-0.040873097	-0.012502253
21	ihh	0.00777366	-0.018522615	-0.051750724	0.062499679
22	sICAM-5...79	0.071545369	-0.012818977	-0.057420246	-0.001306146
23	SP-B	0.064153294	-0.08918026	-0.063727148	0.088754114
24	Trypsin 2	0.534067209	-0.357763949	-0.081978623	-0.094324637
25	IL27B	0.521925626	-0.047251124	-0.08732443	-0.387350071
26	PIGR	0.115752563	0.03050013	-0.094619794	-0.051632899
27	Carbonic anhydrase 6	-0.115774898	0.082524547	-0.099647494	0.132897846
28	SVEP1:EGF-like domains 4-6	0.778614194	-0.080413449	-0.113993999	-0.584206746
29	KERA	0.247380902	0.125440345	-0.114622735	-0.258198512
30	NOTUM	-0.09752195	-0.162506915	-0.127418277	0.387447141
31	sFRP-3	0.179402658	-0.055105606	-0.132530549	0.008233497
32	SVEP1:Sushi 15-18	1.631305193	-0.317140341	-0.185443215	-1.128721638
33	Carbohydrate sulfotransferase 9	0.595472791	0.262609246	-0.233050014	-0.625032023
34	SLPI	0.807227777	-0.14206239	-0.25105336	-0.414112027
35	fibulin 5	0.649853495	0.278963647	-0.301570431	-0.627246712
36	Sonic Hedgehog	0.066721589	-0.004252337	-0.311825488	0.249356236
37	EDIL3	0.202192361	0.122287736	-0.333107056	0.00862696
38	GRIA4	0.14895679	-0.208333971	-0.335373456	0.394750637
39	N-terminal pro-BNP	1.5558349	0.906237434	-0.342232952	-2.119839382
40	BNP	2.026677573	-0.068351798	-0.360443845	-1.59788193
41	Protein C	-0.609650247	0.445932532	-0.368240213	0.531957928
42	Periostin	0.427822973	-0.058337169	-0.385847047	0.016361242
43	CECR1	0.078223196	0.66946445	-0.390114611	-0.357573035
44	OLFL3	0.943687694	1.207822597	-0.396766527	-1.754743764
45	BMP-6	1.004819956	-0.419617583	-0.426007198	-0.159195174
46	NOE1	0.505541953	0.468634054	-0.428600362	-0.545575645

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

47	**sTREM-1**	0.347291105	-0.025559276	-0.494846282	0.173114453
48	**HE4**	0.828182948	-0.149793551	-0.502992997	-0.1753964
49	**Kininogen, HMW, Two Chain**	-0.233245218	-0.451978749	-0.509006116	1.194230083
50	**ADH4**	-0.429409046	0.953535753	-0.652890521	0.128763815
51	**ROBO2**	0.277919383	0.377592754	-0.659314277	0.003802139
52	**ADH1A**	-0.485196197	0.958908421	-0.699449911	0.225737687
53	**PTK7**	1.200335383	0.176785572	-0.737164355	-0.6399566
54	**IGFBP-7**	1.978137564	0.163065896	-0.75759166	-1.383611801
55	**ANTR1**	0.221243407	0.434781808	-0.923515748	0.267490532
56	**ST4S6**	1.168384486	-0.393775955	-0.996157803	0.221549272
57	**CILP2**	-0.142434885	0.523402066	-1.047874983	0.666907802
58	**Kininostatin**	-0.022912001	0.337024541	-1.19937035	0.88525781
59	**KREM1**	0.857073398	0.835584014	-1.694454484	0.001797072
60	**Cathepsin S**	1.118764166	1.167394595	-1.90898098	-0.377177781
61	**RIR2**	0.899176846	0.570903086	-2.059275513	0.589195581

61 Somamers with non-zero coefficients were selected from the 123 somamers entered in the lasso regression.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Table S11: Distribution of clusters according to mPAP threshold in discovery cohort

(A) mPAP ≥25 mmHg (n=470)*
Cluster 1 (poor survival): n=105 (22.4%)
Cluster 2: n=141 (30%)
Cluster 3: n=59 (12.6%)
Cluster 4 (best survival): n=165 (35%)

(B) NoPH controls with mPAP >20 to ≤24 mmHg (n=25)
Cluster 1 (poor survival): n=2 (8%)
Cluster 2: n=7 (28%)
Cluster 3: n=4 (16%)
Cluster 4 (best survival): n=12 (48%)

(C) NoPH controls with mPAP ≤20 mmHg (n=111)
Cluster 1 (poor survival): n=4 (3.5%)
Cluster 2: n=21 (19%)
Cluster 3: n=12 (11%)
Cluster 4 (best survival): n=74 (66.5%)

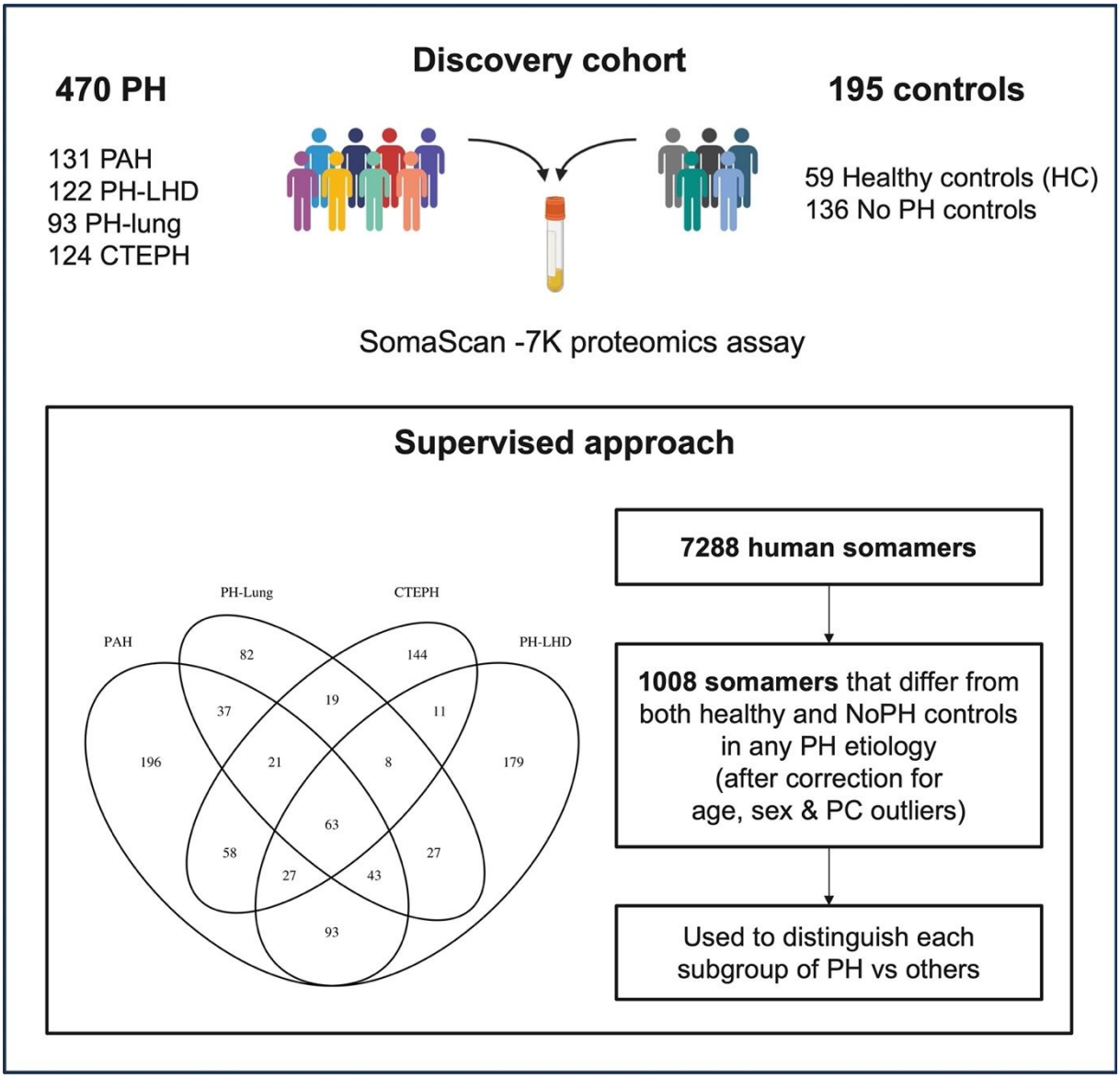
(D) Healthy controls (N=59)
Cluster 3: 1 (2%)
Cluster 4 (best survival): 58 (98%)

* No significant difference (Chi-squared test) between (A) mPAP ≥25mmHg vs (B) mPAP >20 to ≤24 mmHg, p= 0.31; (A) mPAP ≥25mmHg vs (C) NoPH controls with mPAP≤20mmHg, p<0.001.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

IV. Supplemental figures

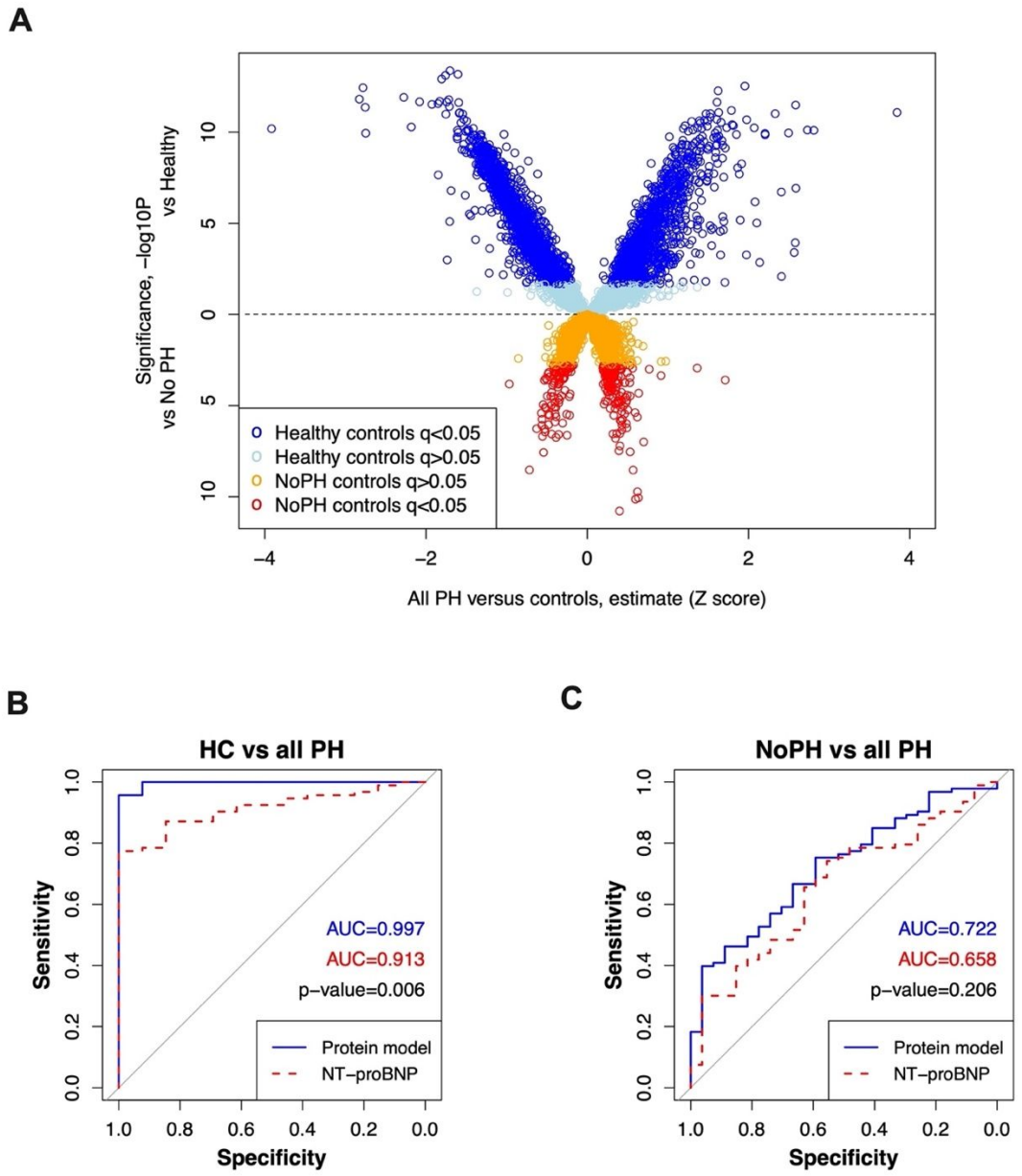
Figure S1: Methodology used to select somamers used to distinguish each PH subgroup from the others



Abbreviations: PH: pulmonary hypertension; PAH: pulmonary arterial hypertension; PH-LHD: PH associated with left heart disease; PH-lung: PH associated with lung disease; CTEPH: chronic thrombo-embolic PH; HC: healthy controls; No PH: symptomatic disease controls without PH; PC: principal component.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure S2: double Volcano plot showing the proteins differentially expressed in pulmonary hypertension and controls (healthy controls and symptomatic controls) (A) and ROC curves testing the performance in replication group of the combination of proteins obtained by lasso regression to identify PH vs healthy controls (B) and PH vs symptomatic controls (C)



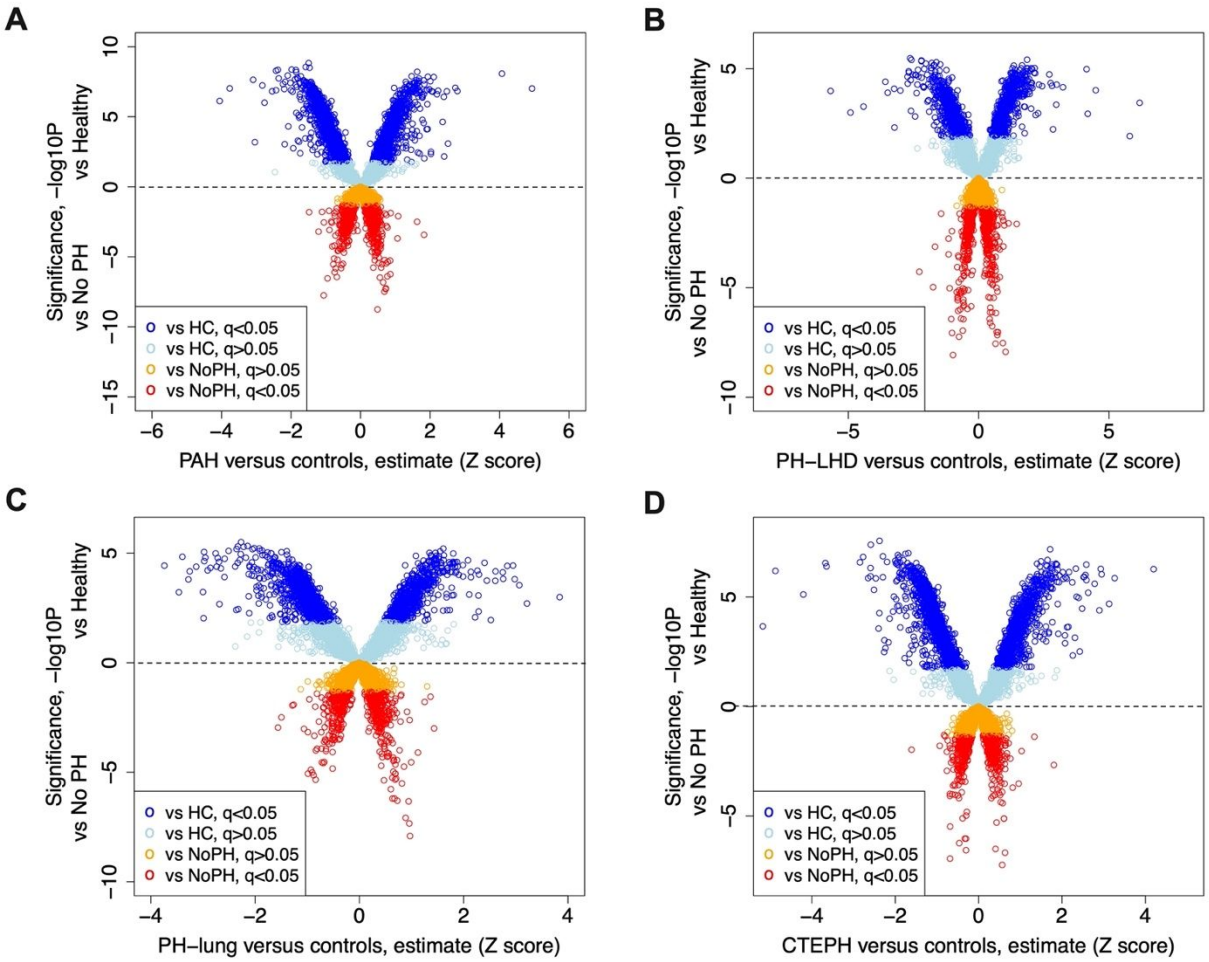
A/ -log10 p value derived from linear regression analysis

B/ AUC combination of 25 proteins: 0.997 (0.989-1.000), p<0.001
AUC NT-proBNP: 0.913 (0.856-0.970), p<0.001; Delong test = 0.006.

C/ AUC combination of 40 proteins: 0.722 (0.621-0.823), p=0.001
AUC NT-proBNP: 0.658 (0.546-0.770), p=0.013; Delong test = 0.206.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

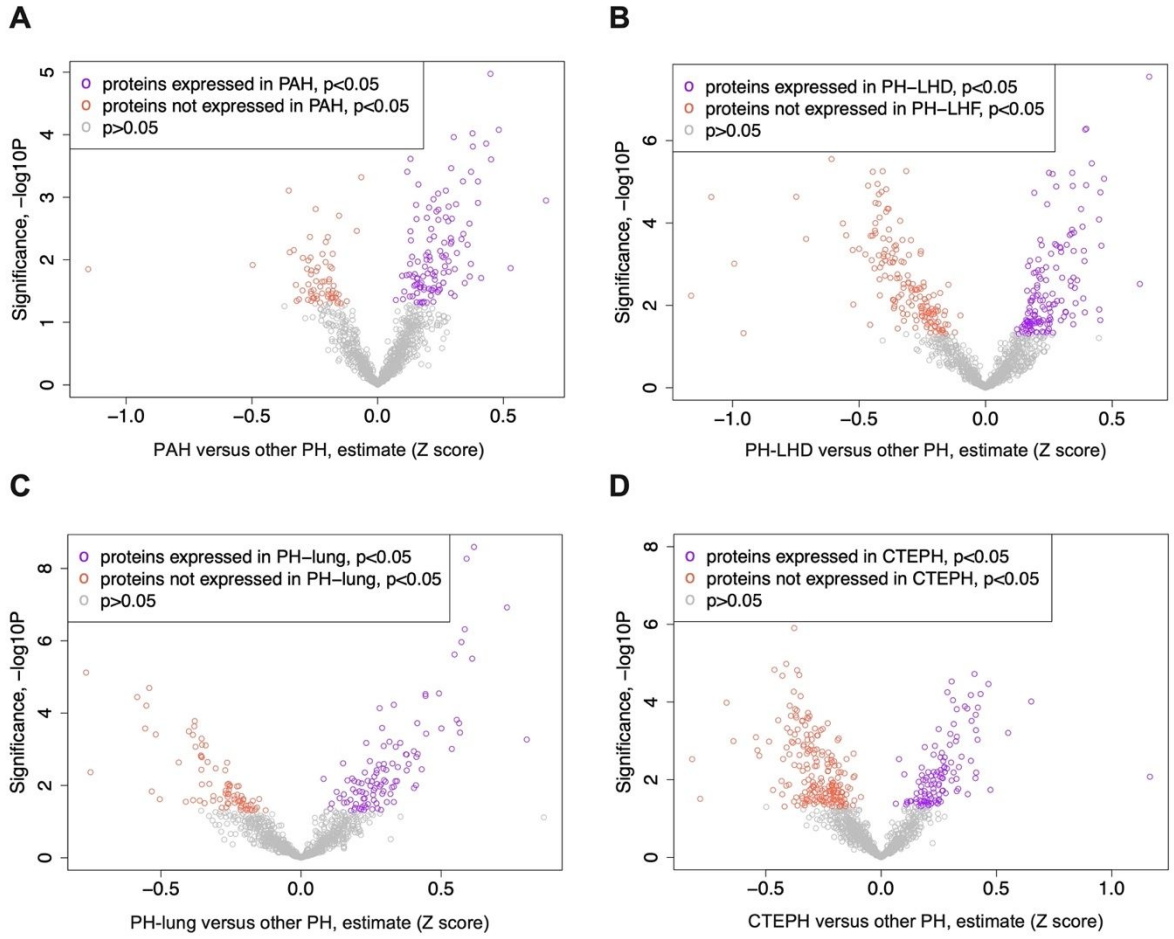
Figure S3: Double Volcano plot showing the proteins differentially expressed in each group of pulmonary hypertensions (A: PAH, B: PH-LHD, C: PH-lung, D: CTEPH) vs controls (healthy controls and symptomatic controls)



A, B, C, D/ -log10 p value derived from linear regression analysis

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

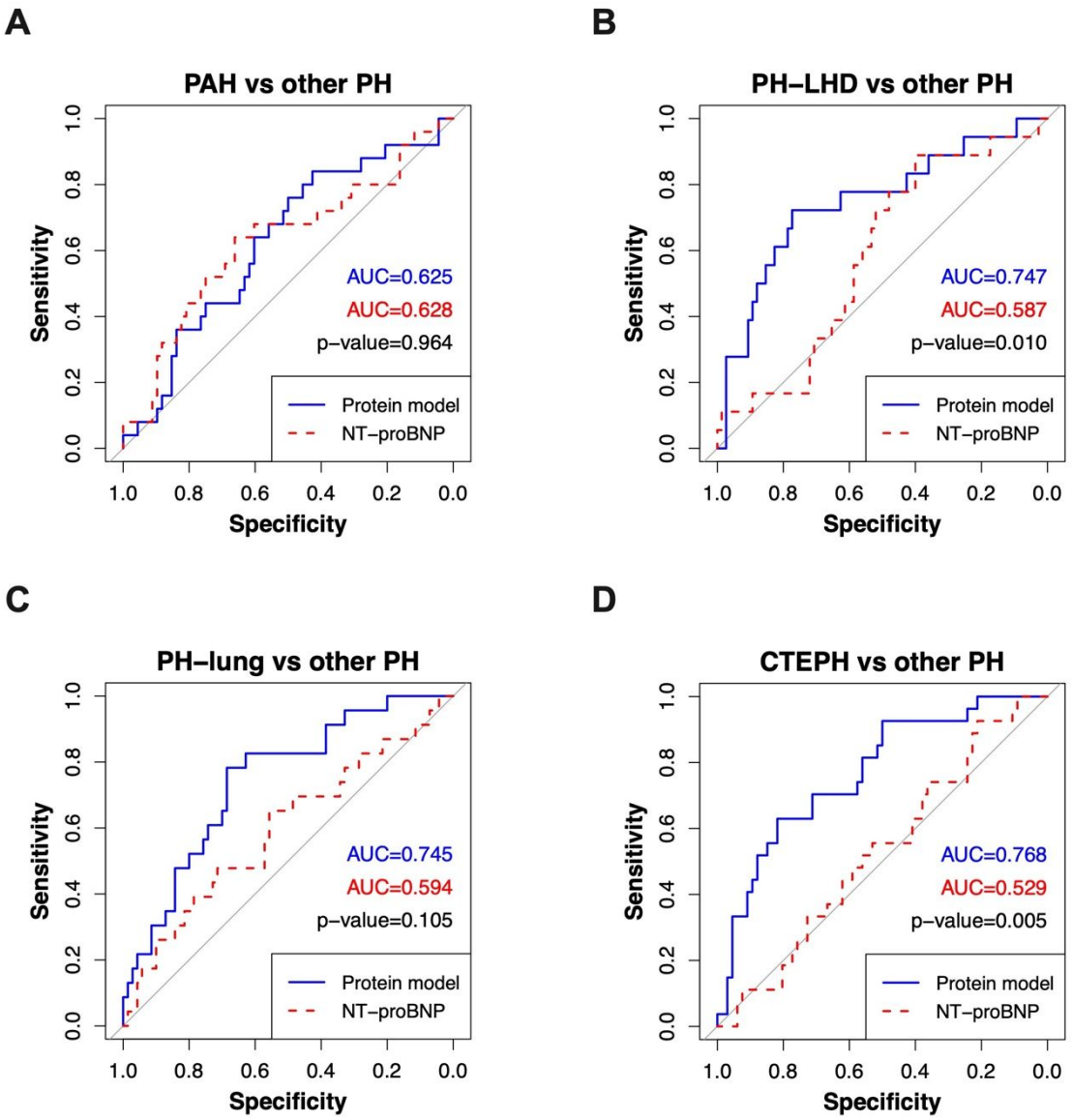
Figure S4: Volcano plot showing the proteins differentially expressed by each group of pulmonary hypertension (PH) (A: PAH, B: PH-LHD, C: PH-lung, D: CTEPH) compared to patients with another aetiology of PH



A, B, C, D/ -log10 p value derived from linear regression analysis

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure S5: ROC curves testing the performance in replication group of the combination of proteins obtained by lasso regression and NT-proBNP to identify PAH from other PH (A), PH-LHD from other PH (B), PH-lung from other PH (C) and CTEPH from other PH (D)



A/ AUC combination of 17 proteins (in blue): 0.625 (0.497-0.752), p=0.067
AUC NT-proBNP (in red): 0.628 (0.491-0.765), p=0.059; Delong test = 0.964.

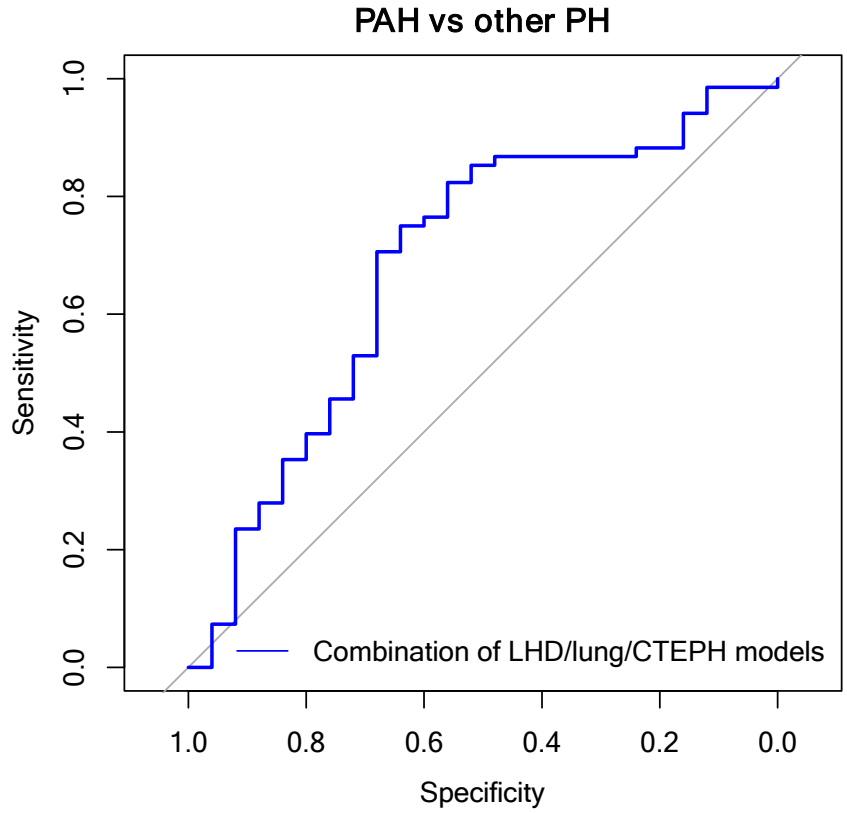
B/ AUC combination of 35 proteins (in blue): 0.747 (0.609-0.885), p=0.001
AUC NT-proBNP (in red): 0.587 (0.449-0.724), p=0.257; Delong test = 0.010.

C/ AUC combination of 40 proteins (in blue): 0.745 (0.633-0.857), p<0.001
AUC NT-proBNP (in red): 0.594 (0.454-0.735), p=0.177; Delong test = 0.105.

D/ AUC combination of 29 proteins (in blue): 0.768 (0.663-0.872), p=0.005
AUC NT-proBNP (in red): 0.529 (0.403-0.654), p=0.669; Delong test = 0.005.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

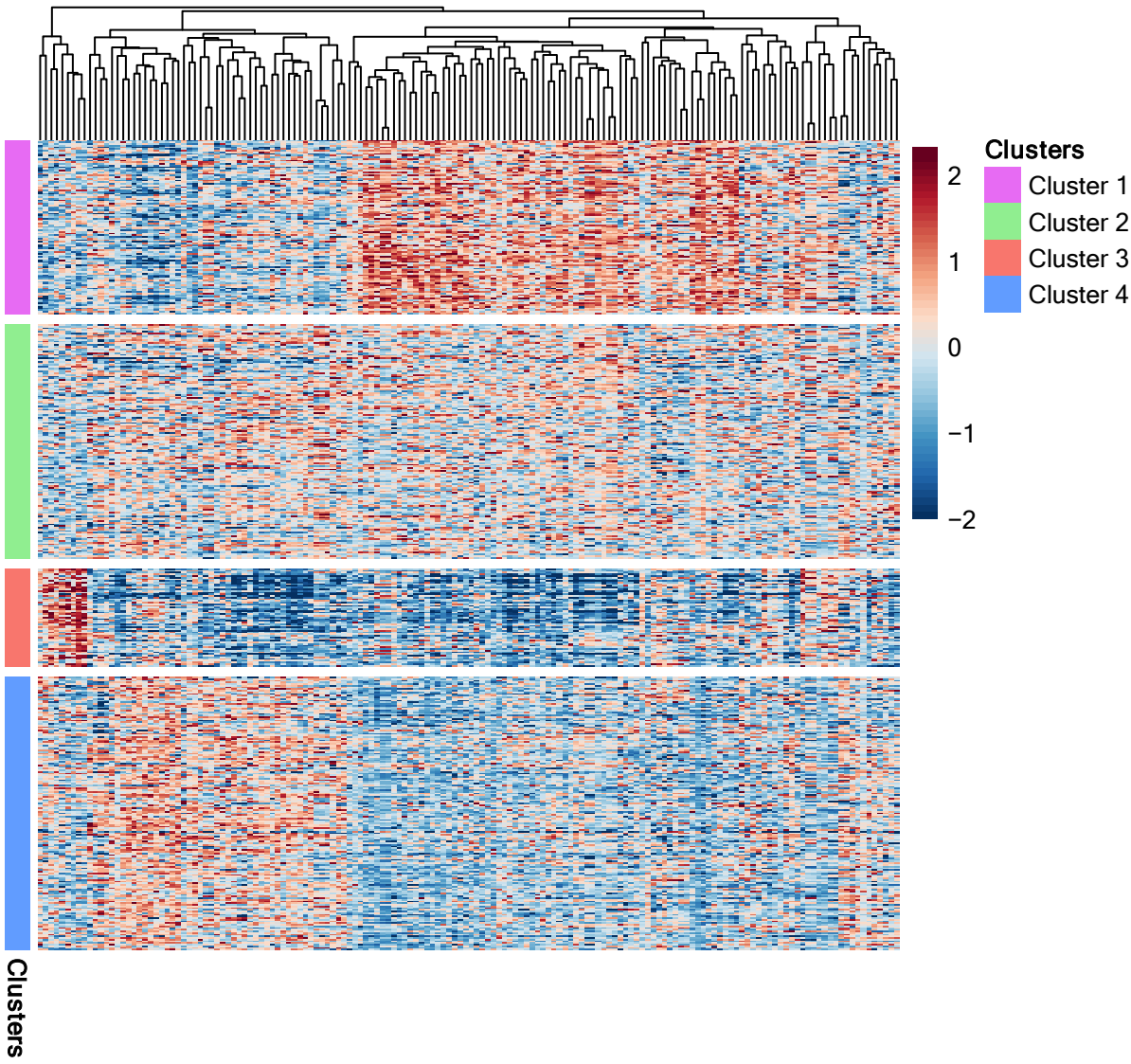
Fig S6: ROC curve of the logistic regression of the combination of proteins able to identify groups 2, 3 and 4 PH in order to test the ability to identify PAH from other PH groups



AUC combination of models: 0.684 (95% CI: 0.552 – 0.815), p=0.007

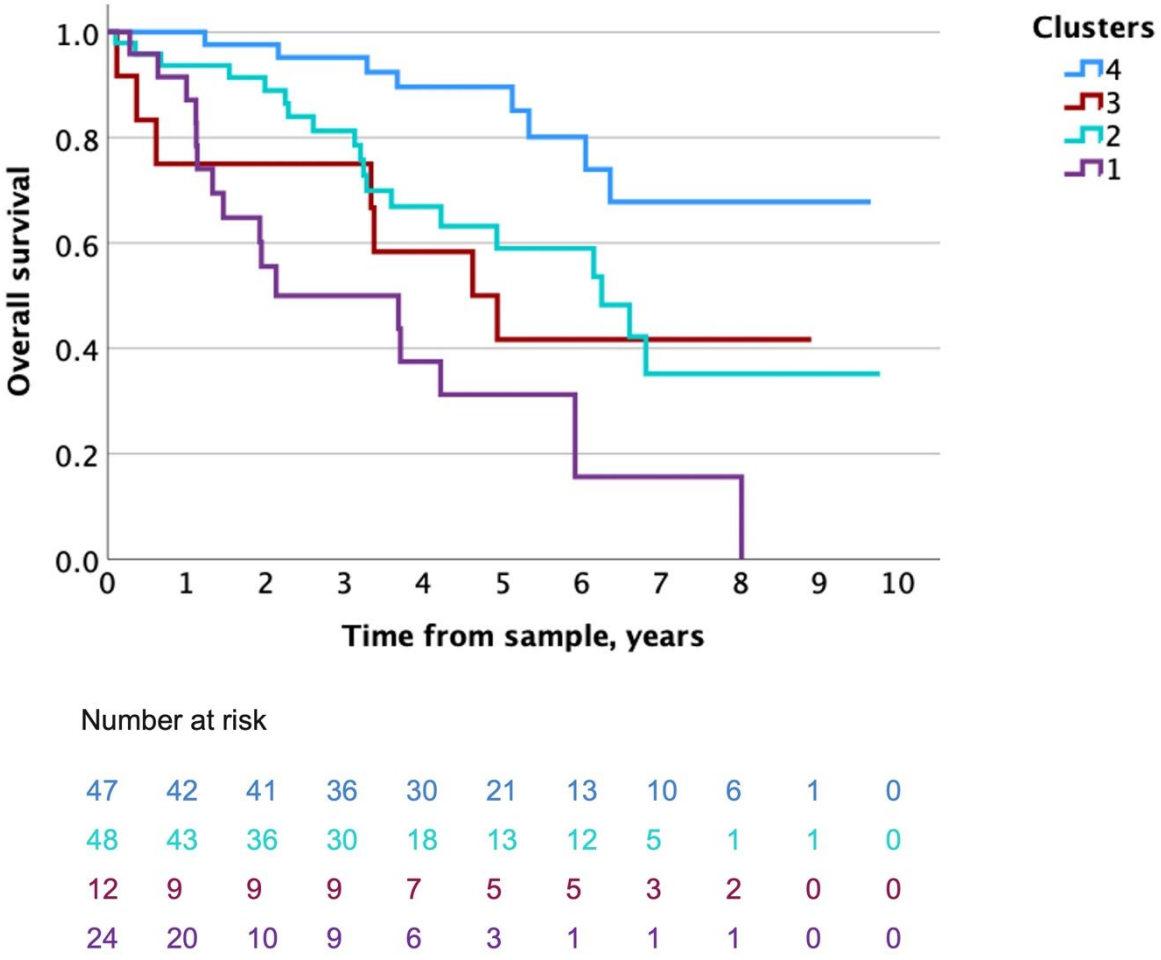
AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure S7: Heatmap of 165 somamers used to identify clusters in the discovery cohort



AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

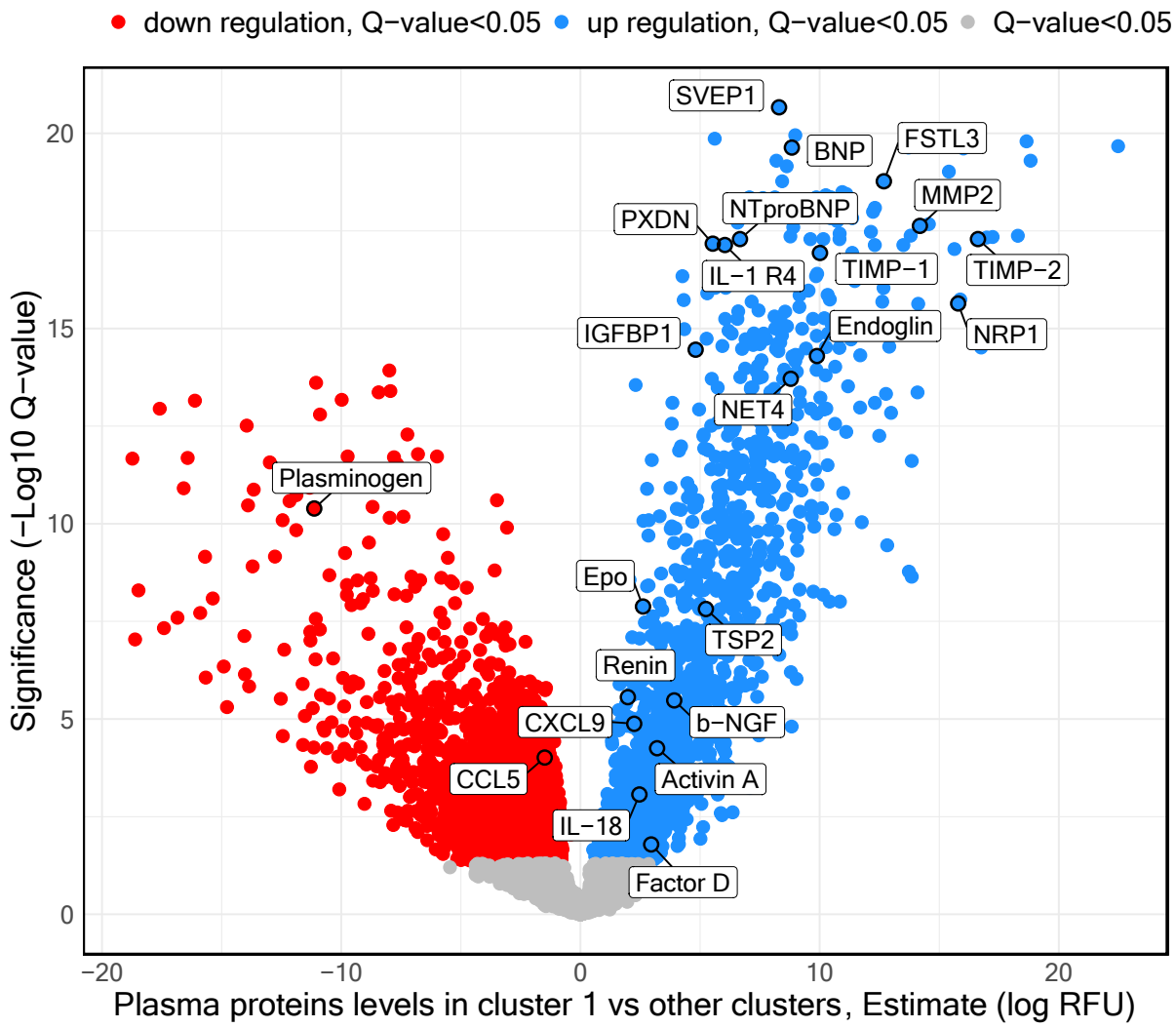
Figure S8: Kaplan-Meier survival curves according to clusters in patients with PAH in the discovery cohort



Log rank test clusters 2 versus 3, p=0.73.

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

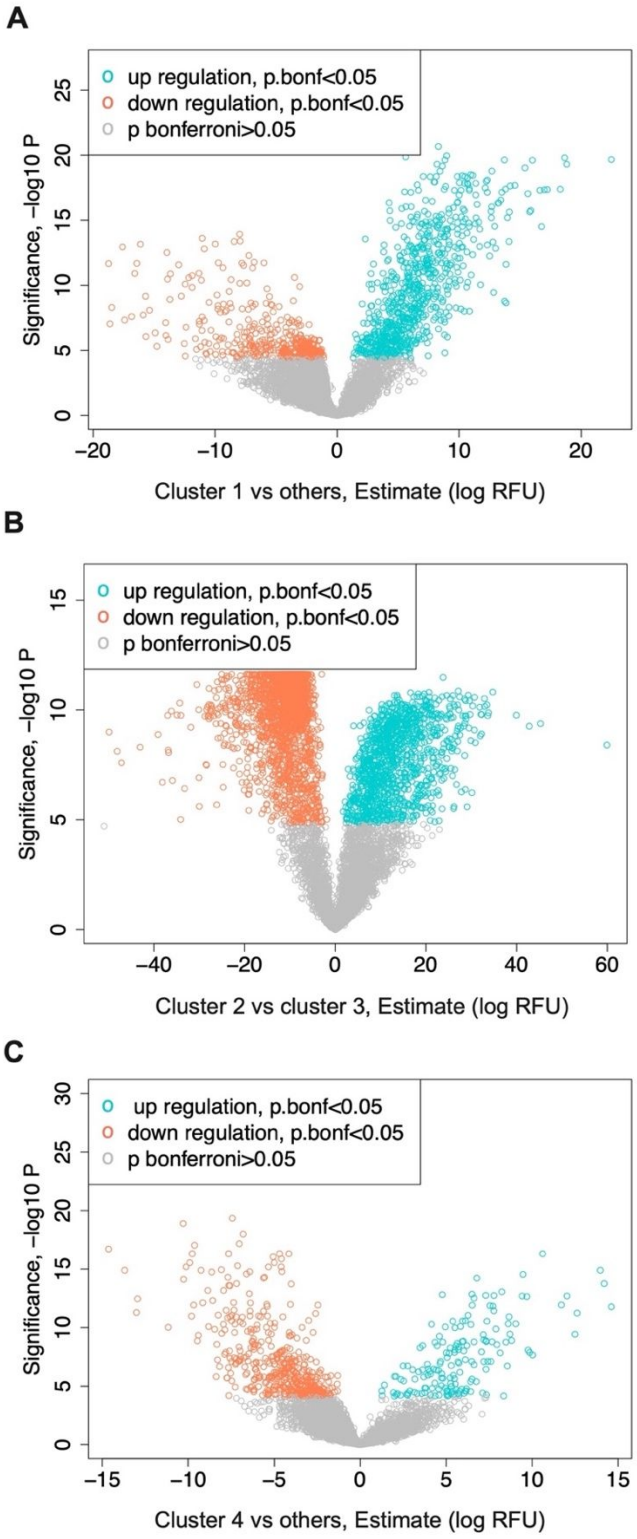
Figure S9: Volcano plot showing plasma levels of proteins, including known prognostic biomarkers, in cluster 1 compared to other clusters



A, B, C, D/ -log10 q value derived from linear regression analysis

AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

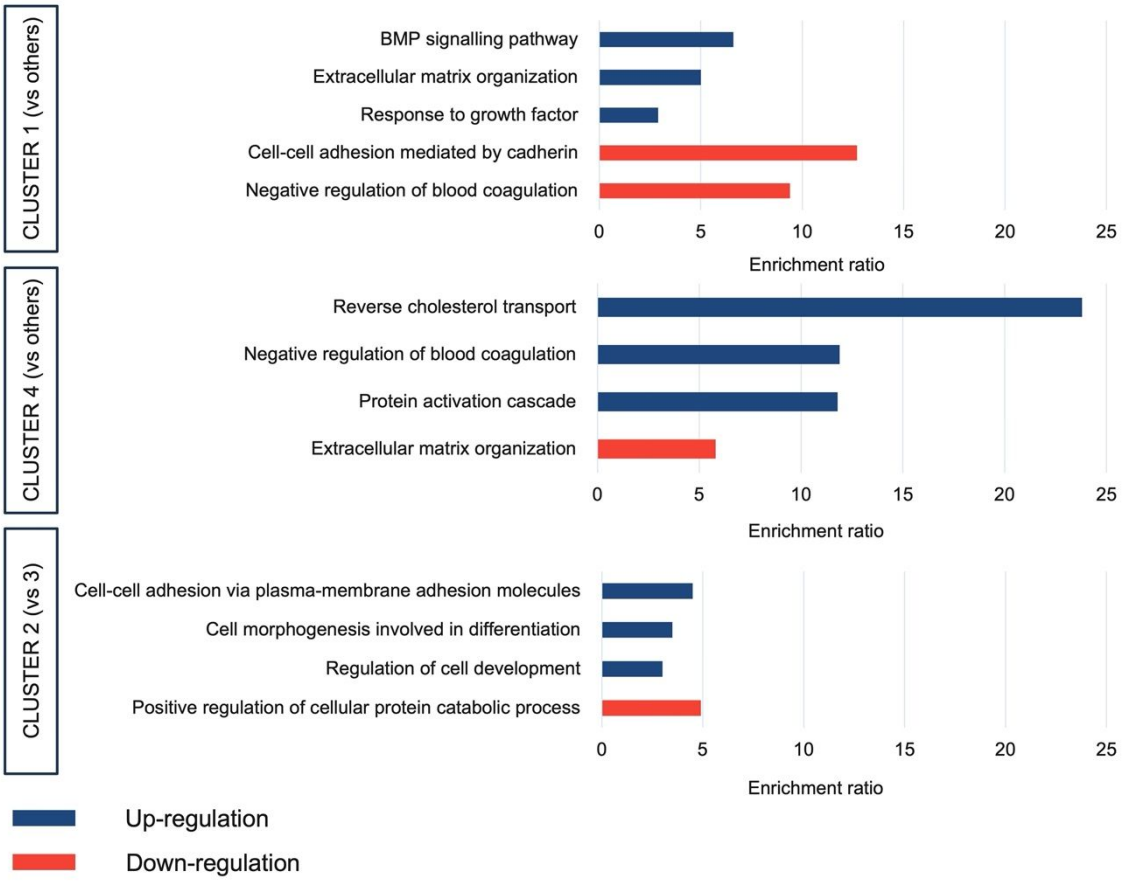
Figure S10: Volcano plot showing up and down-regulated proteins in cluster 1 (A) cluster 4 (B) and cluster 2 (C)



A, B, C/ -log10 p value derived from linear regression analysis

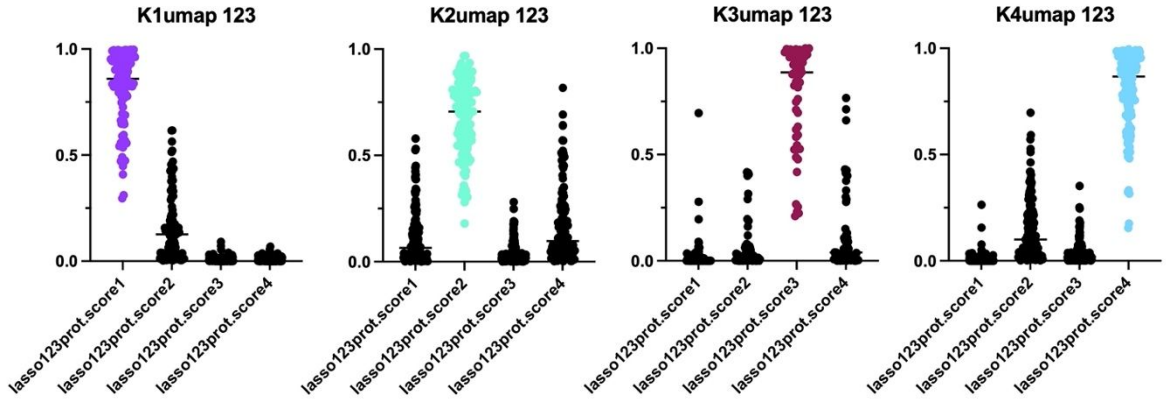
AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure S11: Enrichment analysis of the top 100 up- and down-regulated proteins in cluster



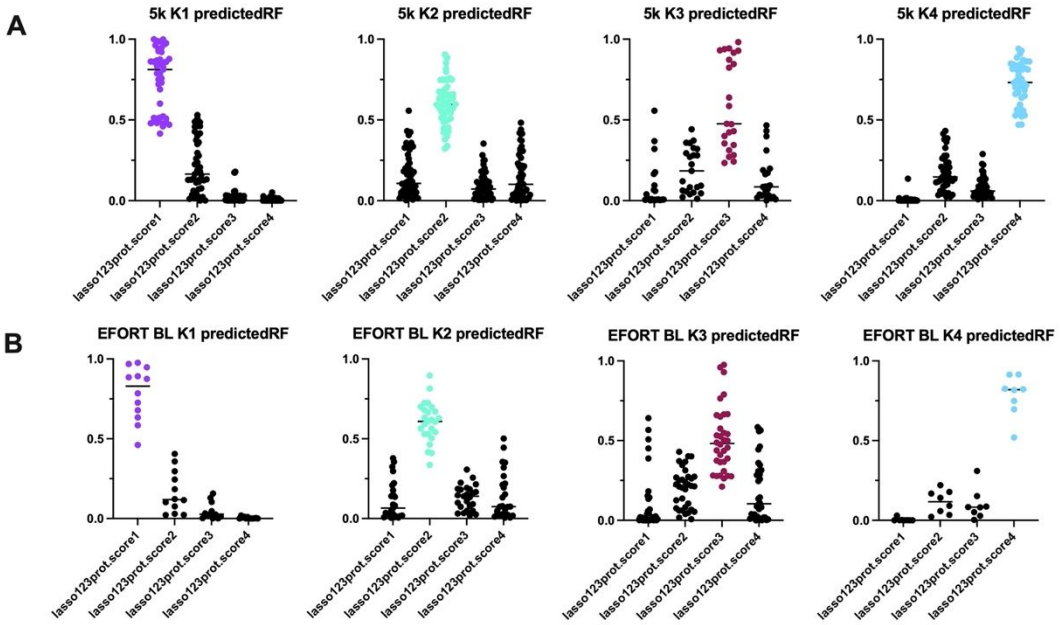
AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure S12: Lasso scores of each cluster in the discovery cohort of patients with pulmonary hypertension



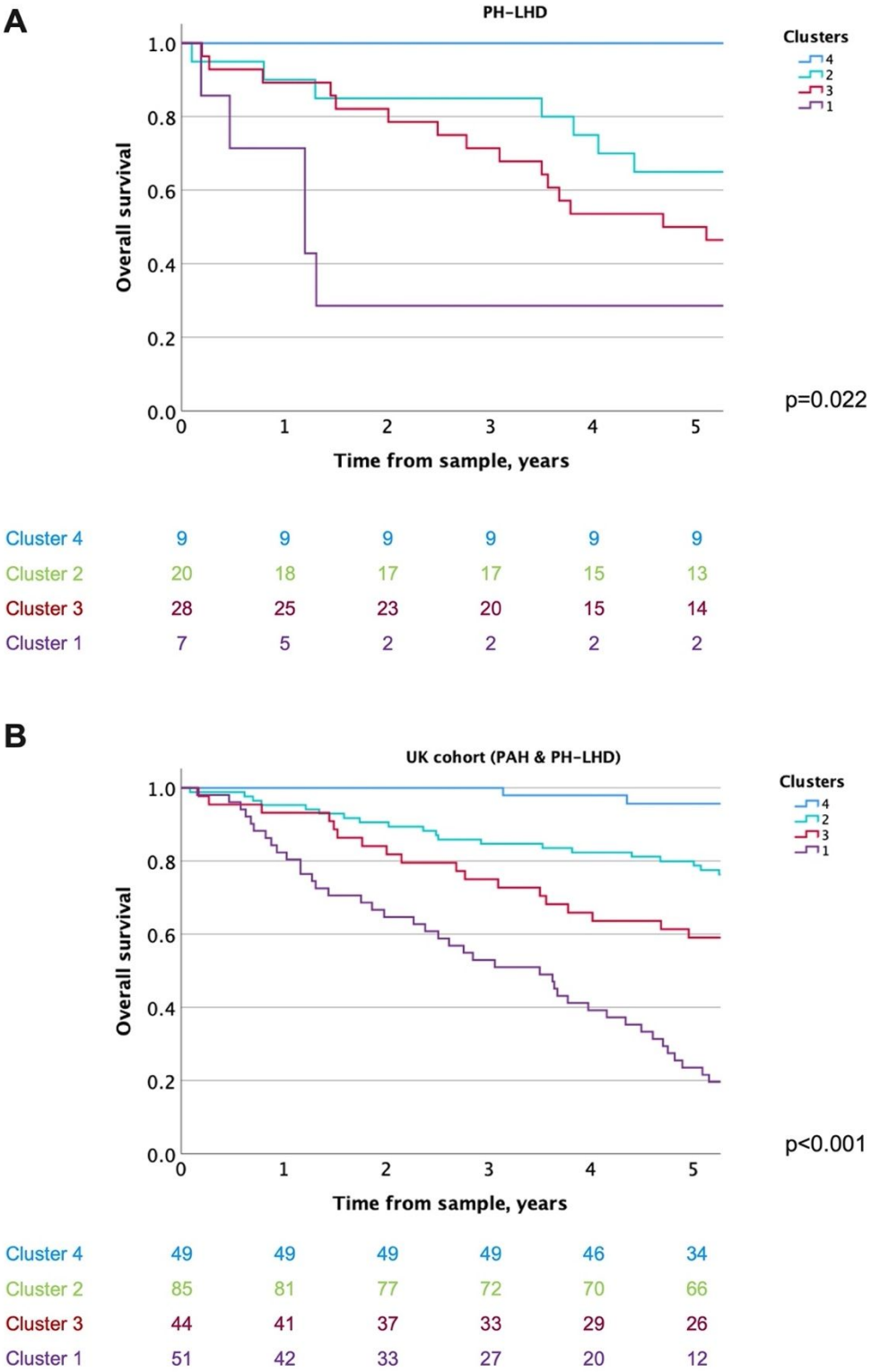
AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure S13: Lasso scores according to cluster identified by random forest in UK (A) and French (B) validation cohorts.



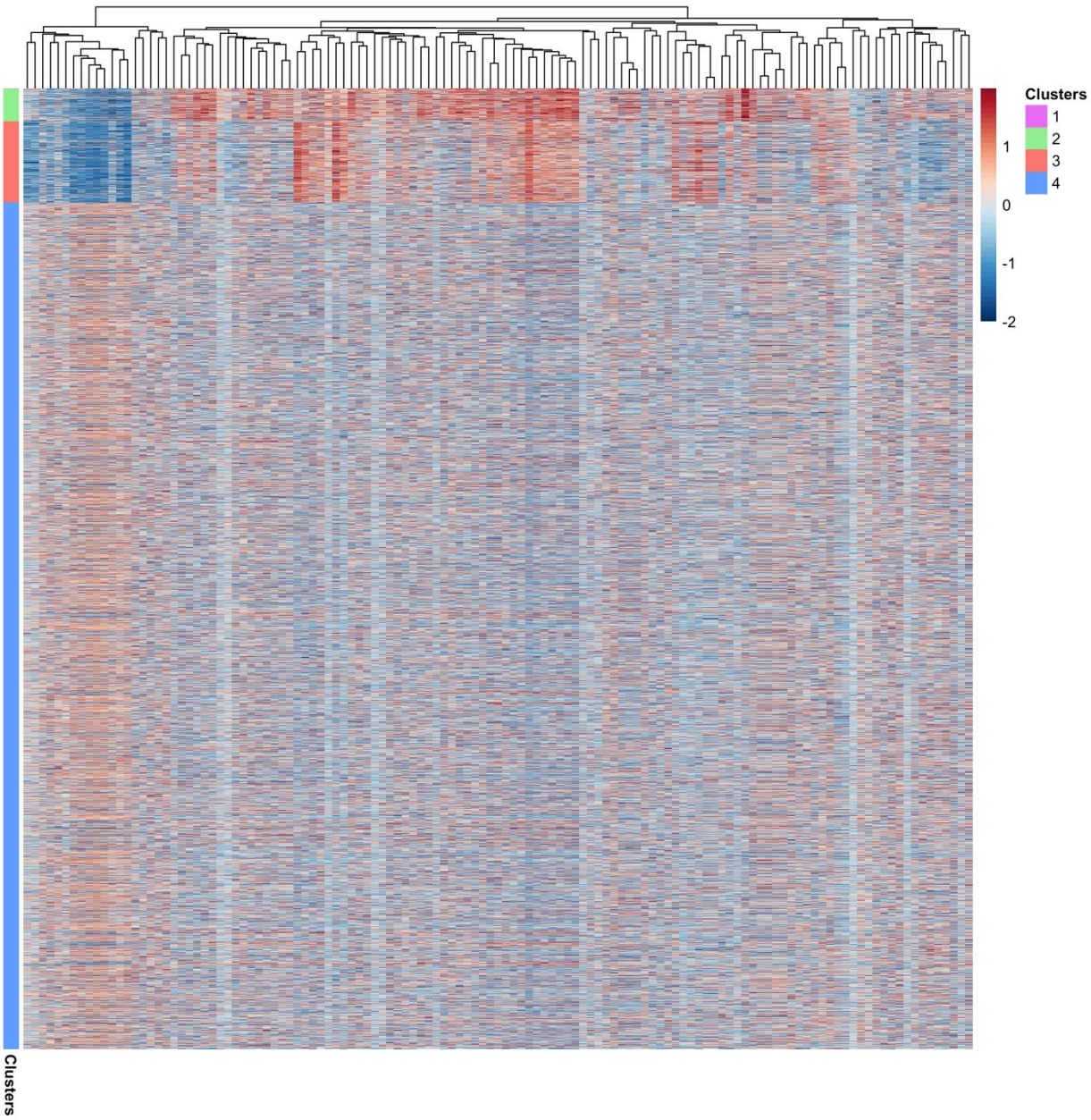
AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure S14: Kaplan-Meier survival curves according to clusters in the UK validation cohort of patients with PH-LHD (A) and UK PH cohort with both precapillary and postcapillary PH (B)



AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC

Figure S15: Heatmap of somamers used to identify clusters in the Whitehall II cohort



AJRCCM Articles in Press. Published May 09, 2025 as 10.1164/rccm.202408-1574OC