Genetic investigations of causality in disease applied to iron deficiency in pulmonary arterial hypertension

Thesis submitted for the degree of

Doctor of Philosophy

Imperial College London By

Anna Ulrich

2020

National Heart and Lung Institute, Medicine

Imperial College London

Hammersmith Hospital

London W12 0NN

I declare that this thesis was conducted and written by myself and the work included within is my own unless otherwise stated.

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC).

Acknowledgements

First of all, I would like to thank my supervisors, Dr Christopher Rhodes, Prof Inga Prokopenko and Prof Martin Wilkins for the patient guidance, encouragement and advice they have provided throughout my time as their PhD student.

I have been extremely lucky to have a supervisor like Chris who cared so much about my work, responded to my questions tirelessly and helped in putting my project and thesis into perspective. I owe a special thank you to Inga who provided encouragement and guidance in more than just the project itself and who was always enthusiastic to assist in any way she could throughout my studentship. I would also like to thank Martin for providing the opportunity to undertake this PhD, for guiding me in publishing my work and for involving me in the analysis of the Ferinject clinical trial.

I would also like to acknowledge Dr John Wharton for his help with writing my first paper, for sending me all those useful publications on my subject and for helping me find my way around in the hospital`s patient database. Completing this work would have been all the more difficult were it not for the support and friendship of all the PhD students and researchers of the Rhodes-, Prokopenko- and Wilkins labs.

To conclude, I cannot forget to thank my parents and my brother for all the unconditional support and patience in these challenging last months of completing my thesis.

Abstract

Idiopathic pulmonary arterial hypertension (IPAH) is a rare and devastating disease with poor prognosis and high mortality despite novel therapeutic options. Iron deficiency is a therapeutic target of interest in IPAH, given its prevalence and association with outcomes, the promising results of open label studies of iron replacement in IPAH patients and the physiological link between iron availability and the response to hypoxia in healthy individuals. To date, evidence of a causal relationship between iron deficiency and IPAH is lacking.

Epidemiological investigation into the prevalence and prognostic value of iron availability in a etiologically heterogenous cohort of pulmonary hypertension (PH) patients including IPAH showed that iron deficiency is common in all PH etiologies as well as disease controls referred to the PH clinic at Hammersmith Hospital. Distinct definitions of iron deficiency associated with varying outcomes in patients highlight the difficulty in correctly identifying iron deficiency in chronic diseases and calls for the evaluation of iron status biomarkers in PH.

Mendelian Randomisation (MR) analysis of iron availability via red cell distribution width (RDW) and IPAH susceptibility did not provide evidence for a causal relationship between iron status and IPAH susceptibility. Secondary MR analysis using instrumental RDW loci acting through systemic iron status also did not support a causal association. Both analyses were well powered to detect a causal effect estimated by the observational study although a more modest causal effect, similar to MR estimates previously reported between iron status and other diseases, cannot be ruled out.

The MR analysis of iron status and IPAH is in line with the results of the first cross-over randomized controlled trial (RCT) of iron supplementation in IPAH patients with iron deficiency. Single infusion of ferric carboxymaltose (1000 mg) was well tolerated and restored iron stores to iron replete levels in 62% of IPAH patients. No effect of iron treatment was found on either primary or secondary outcomes. Secondary analyses found no effect of the magnitude of improvement in iron availability on change in disease severity and exercise measures from baseline to 12 weeks. The conclusions from these two approaches (MR and RCT) suggest that iron deficiency is not a powerful therapeutic target in PAH and therefore not a priority for future clinical studies.

Genome- and transcriptome-wide eQTL-mapping in a cohort of IPAH patients identified hundreds of potentially novel expression quantitative trait loci (eQTL) colocalizing with twice the proportion of GWAS loci for lung-related phenotypes than eQTL validated by population-based studies. Apart from pulmonary conditions, these novel eQTL specific to or active in PAH could be useful in understanding

genetic risk factors for other diseases that share common mechanisms with PAH, such as those with immune dysregulation.

The use of genetic epidemiology to investigate causal relationships and assess the genetic architecture shared between PAH and other candidate phenotypes could lead to improved understanding of PAH mechanisms and identify PAH risk factors and potential therapeutic targets. Such analyses will be made more powerful by increasing the accuracy of PAH genetic effect estimates through extensive collaboration, and by improving QTL definitions of phenotypes relevant to PAH.

Table of Contents

1 Introduction 15
1.1 Pulmonary Hypertension 15
1.2 Epidemiology of PAH 15
1.3 Mechanisms of PAH 18
1.4 Genetic background of PAH 20
1.5 Therapeutic options for PAH 24
1.5.1 Endothelin 1 (ET-1) 24
1.5.2 Phosphodiesterase 5 24
1.5.3 Soluble guanylate cyclase 25
1.5.4 Prostanoids 25
1.5.5 Prostacyclin IP receptor agonist 25
1.6 Iron homeostasis and iron biomarkers 26
1.7 Iron deficiency in PAH 29
1.8 Causal inference in epidemiology 34
1.9 Mendelian Randomisation 37
1.10 Expression quantitative trait loci 39
1.11 Summary of background 42
1.12 Hypothesis 42
1.13 Aims 42
2 Materials and methods 44
2.1 Definition of patient groups 44
2.2 Definition of iron deficiency in the National Pulmonary Hypertension Service 44
2.2.1 Iron biomarkers in chronic disease 45
2.3 Whole-genome DNA sequencing in the NIHRBR study 47
2.3.1 Next generation sequencing 47
2.3.2 The NIHRBR study 48
2.4 RNA sequencing in the PAH Cohort study 48

2.4.1 RNA sequencing 48
2.4.2 PAH Cohort study 49
2.5 Expression quantitative trait loci mapping in PAH 50
2.6 Mendelian Randomisation: one-sample design and assumptions. 51
2.7 Mendelian Randomisation: two-sample design. 53
2.8 Genetic effects for red cell distribution width and pulmonary arterial hypertension 57
2.9 Genetic risk score calculation in the VUMC study 58
2.9.1 The VUMC study 58
2.9.2 Genetic risk score calculation 58
2.10 Clinical studies of intravenous iron in PAH 60
3 Iron deficiency and anaemia in pulmonary hypertension diagnostic groups: a cross-sectional study in the National Pulmonary Hypertension Service, Hammersmith Hospital 61
3.1 Introduction 61
3.2 Methods 64
3.2.1 Data 64
3.2.2 Statistical analyses 65
3.3 Results 68
3.3.1 Iron deficiency is common in all patient groups attending the PH clinic. 68
3.3.2 Haemoglobin levels are higher in H/IPAH, PAH CHD and CTEPH than in controls. 70
3.3.3 The ratio of soluble transferrin receptors and ferritin is not associated with inflammation, while soluble transferrin receptor levels are in white individuals. 73
3.3.4 Hepcidin levels are similar between H/IPAH, PAH associated with LHD and disease controls. 78
3.3.5 sTfR levels do not prognosticate, but belonging to the sTfR-only group predicts mortality. 81
3.4 Discussion 87
3.5 Conclusions 92
4 Mendelian randomisation analysis of red cell distribution width in pulmonary arterial hypertension 93

4.1 Introduction 93
4.2 Methods 95
4.2.1 Definition of pulmonary arterial hypertension (PAH) 95
4.2.2 Genetic and phenotype data in contributing studies 95
4.2.3 Statistical analyses 97
4.3 Results 101
4.3.1 Defining the association of RDW and PAH 101
4.3.2 Genetic risk score for RDW 101
4.3.3 Statistical power to detect causal effect 102
4.3.4 RDW-PAH causal relationship 103
4.4 Discussion 105
4.5 Conclusions 108
5 Supplementation of iron in idiopathic pulmonary hypertension: a randomized controlled crossover trial 109
5.1 Introduction 109
5.2 Methods 111
5.2.1 Participants 112
5.2.2 Study design 115
5.2.3 Statistical analyses 115
5.3 Results 117
5.3.1 Baseline iron deficiency of participants 117
5.3.2 Primary and secondary outcomes by treatment 117
5.3.3 Primary and secondary outcomes by change in iron status 123
5.4 Discussion 129
5.5 Conclusions 132
6 Expression quantitative trait locus mapping in pulmonary arterial hypertension 133
6.1 Introduction 133
6.2 Methods 135

6.2.1 Study participants and sample processing 135
6.2.2 Gene expression data 135
6.2.3 Genetic data 135
6.2.4 eQTL effect estimation 136
6.2.5 Validation rate 137
6.2.6 Studies used for validation 139
6.2.7 Overlap with common variants affecting PAH susceptibility 140
6.2.8 Overlap with variants from the GWAS Catalog 140
6.2.9 Functional enrichment analysis of genes with novel eQTL 141
6.3 Results 142
6.3.1 eQTL discovery 142
6.3.2 Validation rate 144
6.3.3 Overlap with common variants affecting PAH susceptibility 144
6.3.4 Overlap with variants from the GWAS Catalog 146
6.3.5 Functional enrichment analysis of genes with novel eQTL 148
6.4 Discussion 149
6.5 Conclusions 154
7 Conclusions and future work 155
8 Appendix 160
8.1 Iron deficiency and anaemia in pulmonary hypertension diagnostic groups: a cross-sectional study in the National Pulmonary Hypertension Service, Hammersmith Hospital 160
8.2 Mendelian randomisation analysis of red cell distribution width in pulmonary arterial hypertension 167
8.3 Supplementation of iron in idiopathic pulmonary hypertension: a randomized controlled crossover trial 171
8.4 Expression quantitative trait locus mapping in pulmonary arterial hypertension 178

List of figures

Figure 1.1 Remodeled pulmonary artery. 18
Figure 1.2 Regulation of hepcidin production. 28
Figure 1.3 Causal pie model. 34
Figure 1.4 Schematic diagram of confounding in epidemiology. 36
Figure 1.5 Mendelian Randomisation analysis of homocysteine and coronary heart disease. 38
Figure 1.6 NHGRI-EBI Catalog of published genome-wide association studies (GWAS). 41
Figure 2.1 Two-stage least squares regression for estimating causal effect in the one-sample setting. 51
Figure 2.2 Assumptions of Mendelian Randomization. 52
Figure 2.3 Mendelian Randomisation using summary-level data: low-density lipoprotein and coronary heart disease. 55
Figure 3.1 Proportions of iron status groups in disease controls and five PH diagnostic groups. 69
Figure 3.2 Predictors of `Ferr-TSAT+sTfR` iron deficiency. 70
Figure 3.3 Haemoglobin levels in disease controls and five PH diagnostic groups in white patients. . 71
Figure 3.4 Haemoglobin levels by ethnicity in different diagnostic groups. 72
Figure 3.5 Predictors of soluble transferrin receptor levels in patients by ethnicity. 74
Figure 3.6 Ratio of soluble transferrin receptor and ferritin levels by iron deficiency group. 76
Figure 3.7 Predictors of hepcidin and its levels in disease controls and five PH diagnostic groups..... 79
Figure 3.8 Hepcidin levels by iron deficiency status. 80
Figure 3.9 Prognostic value of soluble transferrin receptor (sTfR) levels. 82
Figure 3.10 Prognostic value of soluble transferrin receptor and ferritin ratio (sTfR/ferritin). 84
Figure 3.11 Prognostic value of iron deficiency groups. 86
Figure 4.1 Flow diagram of the exclusion steps in the UK centers and the Vanderbilt University Medical Centre (VUMC) participating in the RDW and PAH association analysis. 96
Figure 4.2 Mendelian Randomisation (MR) analyses of red cell distribution width (RDW) and pulmonary arterial hypertension (PAH). 99
Figure 4.3 RDW levels in the merged cohort of PAH cases (N=642) and the disease control cohort (N=15,889). 101
Figure 4.4 Power to detect a causal association in the RDW-PAH Mendelian Randomisation analyses. 103
Figure 4.5 RDW-PAH Mendelian Randomisation analyses. 104
Figure 5.1 Study design of the Ferinject trial. 115

Figure 5.2 Phosphate levels in the Ferinject trial by randomization. 117
Figure 5.3 Changes in primary outcomes and iron status in the Ferinject trial. 119
Figure 5.4 Ratio of soluble transferrin receptor and ferritin levels at baseline and twelve weeks after iron administration in the treatment-only analyses. 123
Figure 5.5 Ratio of soluble transferrin receptor and ferritin levels at baseline and twelve weeks after iron administration in the treatment-only analyses in patients with severe iron deficiency at baseline. 124
Figure 5.6 Results of the treatment-only analyses: effect of iron status improvement as estimated by sTfR/Ferritin ratio on change in endpoint. 126
Figure 5.7 Peak O2 pulse and peak VO2 in the treatment-only analyses. 127
Figure 6.1 Population substructure in the PAH Cohort study. 137
Figure 6.2 Results of the PAH Cohort eQTL discovery. 142
Figure 6.3 PAH susceptibility variant chr7:142495156 (T/G) is also an eQTL for TRBJ2-7. 146
Figure 6.4 Reported GWAS traits for eQTL in the PAH Cohort. 147

List of tables

Table 1.1 Clinical Classification of Pulmonary Hypertension. 17
Table 1.2 Genes hosting rare pathogenic mutations in pulmonary arterial hypertension. 22
Table 1.3 Pathways of disease mechanism in PAH targeted by current pharmacotherapy, drug classes and generic names. 24
Table 1.4 Summary table of intervention studies to date investigating the effects of iron supplementation on outcomes in pulmonary arterial hypertension and studies investigating the role of iron availability in the pulmonary vasoconstrictive response to hypoxia in healthy individuals. 33
Table 3.1 Demographic characteristics and iron indices of patients referred to the National Pulmonary Hypertension Service at Hammersmith Hospital. 67
Table 4.1 Characteristics of the study population used for estimating the association between RDW and PAH, stratified by sex. 95
Table 5.1 Baseline patient characteristics by randomization in Europe and China. 114
Table 5.2 Effect of iron treatment on primary and secondary outcomes in Europe and China. 122
Table 6.1 Patient characteristics and white blood cell fractions in the PAH Cohort study. 136
Table 6.2 Characteristics of the PAH Cohort eQTL mapping study and the eQTL studies used for validation. 140
Table 6.3 Lead variants for the top 50 expression quantitative trait loci in the PAH Cohort ordered by variance explained in gene expression. 143
Table 6.4 Pathways enriched for genes with novel eQTL in the PAH Cohort. 148

List of abbreviations

5-HT - 5-hydroxytryptamine, serotonin

BMP - bone morphogenetic protein

BMPR - bone morphogenetic protein receptor

CAD – coronary artery disease

CHD – coronary heart disease

CHF – chronic heart failure

CI – confidence interval

COPD – chronic obstructive pulmonary disease

CPET – cardiopulmonary exercise test

CTEPH – chronic thromboembolic pulmonary hypertension

DNA - deoxyribonucleic acid

ECPET - endurance bicycle cardiopulmonary exercise testing

EPO - erythropoietin

eQTL – expression quantitative trait locus

ET - endothelin

FDR – false discovery rate

GRS – genetic risk score

GWAS – genome-wide association study

HIF - hypoxia-inducible factors

HIV – human immunodeficiency virus

HPAH – heritable pulmonary arterial hypertension

HRC – The Haplotype Reference Consortium

ICPET - incremental bicycle cardiopulmonary exercise testing

IL-6 – interleukin-6

IPAH – idiopathic pulmonary arterial hypertension

IQR – interquartile range

IVW – inverse variance weighted regression

KEGG – Kyoto Encyclopedia of Genes and Genomes

LD – linkage disequilibrium

LDL – low-density lipoprotein

MAF – minor allele frequency

MHC – major histocompatibility complex

MR – Mendelian Randomisation

NGS – next generation sequencing

NIHRBR - UK National Institute for Health Research BioResource

NO – nitric oxide

NPHS - National Pulmonary Hypertension Service

NYHA – New York Heart Association

ORA – overrepresentation analysis

PAH – pulmonary arterial hypertension

PAH CHD - pulmonary arterial hypertension associated with congenital heart disease

PAH CTD - pulmonary arterial hypertension associated with connective tissue disease

PAP - pulmonary arterial pressure

PASP - pulmonary artery systolic pressure

PAWP - pulmonary arterial wedge pressure

PCR - polymerase chain reaction

PH – pulmonary hypertension

PH LHD - pulmonary hypertension associated with left heart disease

PHD - prolyl-hydroxylase

PVR - pulmonary vascular resistance

QTL – quantitative trait locus

RCT – randomized controlled trial

RDW – red cell distribution width

RIN - RNA Integrity Number

RIT - randomised intervention trial

RNA - ribonucleic acid

SD – standard deviation

sGC - soluble guanylate cyclase

SMAD - mothers against decapentaplegic homolog

SMC – smooth muscle cell

sTfR – soluble transferrin receptor

TfR – transferrin receptor

TGF-β - transforming growth factor β

TPM - transcripts per million

TSAT – transferrin saturation

TSS – transcription start site

VEGF - vascular endothelial growth factor

WGS – whole genome sequencing

WHO – World Health Organization

WM – weighted median

1 Introduction

1.1 Pulmonary Hypertension

Pulmonary hypertension (PH) is a heterogenous group of diseases that affects an estimated 1% of the global population (Hoeper et al., 2016). It is more prevalent in those over the age of 65 and it can complicate common diseases such as lung and left-side hearted diseases. PH is defined as an increase in mean pulmonary arterial pressure (PAP, ≥20 mmHg) (Simonneau et al., 2019) at rest and is divided into five subgroups based on similarities in pathology and clinical presentation devised by the World Health Organization (WHO) (Table 1.1).

WHO group 1 PH, pulmonary arterial hypertension (PAH) (Table 1.1, WHO group 1.) is a devastating rare disease with an estimated prevalence of 7-26 cases/million in the developed world (Hoeper et al., 2016). Despite increased awareness and new therapeutic options, annual mortality remains around 10% (Hoeper et al., 2016). It is one of the pre-capillary forms of PH defined by an elevated mean PAP and a pulmonary arterial wedge pressure (PAWP) ≤15 mmHg, as diagnosed by right heart catheterization (Galie et al., 2016). PAWP estimates pressure in the left atrium with high values suggesting left ventricular failure. In addition to the increased PAP above 20 mmHg and PAWP below 15 mmHg, PAH diagnostic criteria includes increased pulmonary vascular resistance (PVR) above 3 Wood units. Increased PVR is caused by the vasoconstriction and structural remodeling of pulmonary arterioles that eventually leads to right ventricular hypertrophy and end-stage heart failure (Stacher et al., 2012). Despite increased awareness and new therapeutic options, PAH presents with substantial morbidity and mortality (3-year survival rate between 40% - 75%) (Hoeper et al., 2016).

1.2 Epidemiology of PAH

PAH can develop as a complication of connective tissue diseases (CTD), portal hypertension, congenital heart disease (CHD), infection with the human immunodeficiency virus or schistosomiasis and as a result of exposure to certain toxins. The phenotype of PAH has changed considerably over the last few decades. The mean age of patients diagnosed increased from 36 (±15) in the National institutes of Health registry in 1981 to 50-65 (±15) in present day registries (McGoon et al., 2013) whilst 5-year survival improved from 34% in the 1990s to 65% at present (Wilken et al., 2019) owing

to new treatment options being developed and applied in clinical practice. The diagnosis of PAH in the elderly could reflect improved awareness and case finding in the medical community rather than changing epidemiology. This explanation is supported by the mean age of diagnosis of registries in the developing world which resembles that of the Western registries three decades ago (McGoon et al., 2013). Historically, PAH was predominantly diagnosed in young to middle aged women. Female sex as a risk factor for PAH is emphasized by a higher penetrance of PAH in women (42%) than men (14%) in carriers of pathogenic mutations in the gene encoding the bone morphogenetic protein type II receptor (*BMPR2)* (Larkin et al., 2012) which is identified as the cause of most heritable cases of PAH (Lane et al., 2000, Machado et al., 2001). This sex predominance is not as clear in elderly PAH cohorts as observed in a collection of 28 registries across Europe reporting a female-male ratio of 1.2:1 in the elderly cohort (median age 71 years) in contrast to the ratio of 2.3:1 in the younger cohort with a median age of 54 (Hoeper et al., 2013). In the aging PAH population, comorbidities common to old age such as systemic hypertension, obesity and type II diabetes further complicate treatment plans and pharmacotherapy of PAH. An alternative to greater awareness as an explanation to an increased prevalence in the aging PAH population is the misdiagnosis of elderly idiopathic cases with undiagnosed left heart disease (Halpern and Taichman, 2009).

The proportions of PAH subtypes (Table 1.1) in registries differ between the developed and developing world. In Western countries, the largest subgroup is idiopathic (IPAH) and heritable PAH (HPAH) accounting for 30-50% of all PAH cases, whereas in developing countries PAH subgroups associated with other conditions are dominant (Lau et al., 2017). In China, 43% of all PAH cases come from the congenital heart disease associated subgroup reflecting uncorrected congenital heart defects (Lau et al., 2017). In Brazil, where schistosomiasis is an endemic, schistosomiasis infection associated PAH makes up 20% of all PAH cases (Alves et al., 2015).

1. Pulmonary arterial hypertension (PAH)
1.1. Idiopathic
1.2. Hereditary
1.2.1. *BMPR2* mutation
1.2.2. Other mutations (SMAD9, ACVRL1, CAV1, KCNK3, TBX4)
1.3. Induced by drugs and toxins
1.4. Associated with:
1.4.1. Connective tissue disease
1.4.2. HIV infection
1.4.3. Portal hypertension
1.4.4. Congenital heart disease
1.4.5. Schistosomiasis
1'. Pulmonary veno-occlusive disease and/or pulmonary capillary hemangiomatosis
1'.1. Idiopathic
1'.2. Hereditary
1'.2.1. *EIF2AK4* mutation
1'.2.2. Other mutations
1'.3. Induced by drugs, toxins, and radiation
1'.4. Associated with:
1'.4.1. Connective tissue disease
1'.4.2. HIV infection
1". Persistent pulmonary hypertension of the newborn
2. Pulmonary hypertension due to left heart disease
2.1. Left ventricular systolic dysfunction
2.2. Left ventricular diastolic dysfunction
2.3. Valvular disease
2.4. Congenital/acquired obstruction of the left ventricular inflow/outflow tract and congenital cardiomyopathy
2.5. Congenital or acquired pulmonary vein stenosis
3. Pulmonary hypertension due to respiratory disease and/or hypoxemia
3.1. Chronic obstructive pulmonary disease
3.2. Diffuse interstitial lung disease
3.3. Other lung diseases with mixed restrictive and obstructive pattern
3.4. Sleep disordered breathing
3.5. Alveolar hypoventilation
3.6. Chronic exposure to high altitudes
3.7. Lung development disorders
4. Chronic thromboembolic pulmonary hypertension and other pulmonary artery obstructions
4.1. Chronic thromboembolic pulmonary hypertension
4.2. Other pulmonary artery obstructions
4.2.1. Angiosarcoma
4.2.2. Other intravascular tumors
4.2.3. Arteritis
4.2.4. Congenital pulmonary artery stenosis
4.2.5. Parasitosis (hydatid disease)
5. Pulmonary hypertension with unclear and/or multifactorial mechanisms
5.1. Hematologic diseases: hemolytic anemia, myeloproliferative disorders, splenectomy
5.2. Systemic diseases: sarcoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis, neurofibromatosis
5.3. Metabolic disorders: glycogen storage disease, Gaucher's disease, thyroid disorders
5.4. Other: pulmonary tumor thrombotic microangiopathy, fibrosing mediastinitis, chronic renal failure

Table 1.1 Clinical Classification of Pulmonary Hypertension.
Adapted from Galie` et al 2016. The five main World Health Organization (WHO) subclasses are in bold. BMPR2 = bone morphogenetic protein receptor type 2; EIF2AK4 = Eukaryotic Translation Initiation Factor 2 Alpha Kinase 4.

1.3 Mechanisms of PAH

The remodeling of the pulmonary arteries in PAH illustrated in Figure 1.1 (Ramakrishnan et al., 2018, Stacher et al., 2012) affect all layers of the vessel, namely the intima (lining of the artery comprised of endothelial cells), media (middle layer comprised of smooth muscle cells – SMCs - of varying depth), and the adventitia (outer layer of connective tissue comprised of fibroblasts), resulting in a severe decrease in the size of the lumen and elasticity of the vessel, increasing pulmonary vascular resistance and the workload of the right heart (Schermuly et al., 2011). Reduced levels of the vasodilators nitric oxide (NO) (Ghofrani et al., 2004) and prostacyclin (also called prostaglandin I_2) (Christman et al., 1992) and increased levels of the vasoconstrictors endothelin (Giaid et al., 1993) and thromboxane (Christman et al., 1992) in PAH patients lead tovasoconstriction and proliferation of the arterial SMCs. In PAH, distal pre-capillary arteries and the proximal muscular arteries undergo severe muscularization via SMC proliferation induced by various possible mediators, such as serotonin (5-HT) (Eddahibi et al., 2006), or certain vasoactive mediators (Davie et al., 2002), and growth factors. Muscularized proximal arteries can subsequently form complex glomeruloid-like structures termed plexiform lesions (Figure 1.1). These plexiform lesions express vascular endothelial growth factor (VEGF) and VEGF receptors, however the role of VEGF has so far been uncertain with its upregulation in hypoxia-induced rats and downregulation in the monocrotaline induced rat model (Schermuly et al., 2011). Other growth factors implicated in PAH and vascular remodeling include basic fibroblast growth factor, epidermal growth factor, platelet derived growth factor and hepatocyte growth factor.

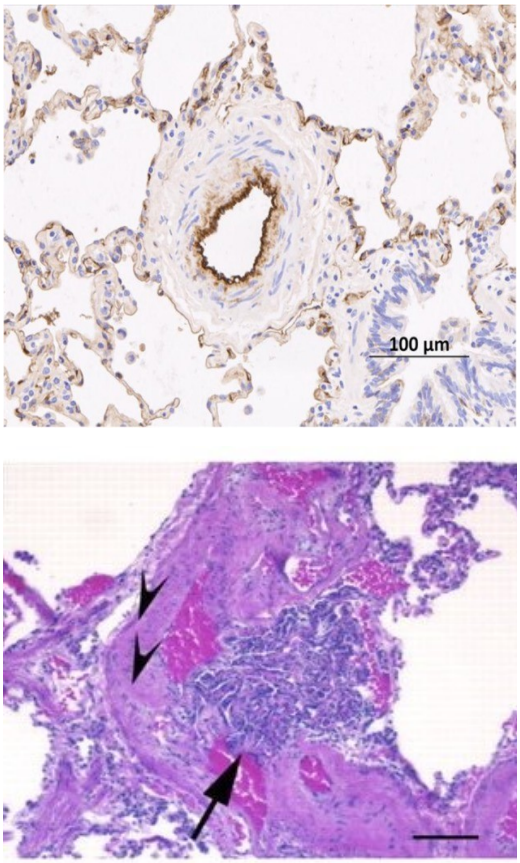


Figure 1.1 Remodeled pulmonary artery. ***Top:*** *Remodeled pulmonary arteriole from a patient with idiopathic pulmonary arterial hypertension taken after lung transplantation. Proliferation of the smooth muscle cells in the media led to thickening of the arteriole wall and narrowing of the lumen. Reprinted from Ramakrishnan et al. (2018). Pulmonary Arterial Hypertension: Iron Matters. Frontiers in Physiology.* ***Bottom:*** *The arrow points at a plexiform lesion fed by a muscular artery (arrowheads). Scale bar = 100 µm. Reprinted from Stacher et al. (2012). Modern age pathology of pulmonary arterial hypertension. AJRCCM.*

Inflammation is suspected to be a major contributor to PAH (Rabinovitch et al., 2014). Autoimmune conditions, such as systemic lupus erythematosus and scleroderma (Dib et al., 2012) and infectious diseases, such as human immunodeficiency virus (HIV) and schistosomiasis infections can lead to PAH, suggesting a link with immunity. Animal studies show that overexpression of the cytokine interleukin-6, (Steiner et al., 2009) or induction of the Th2 immune response (Daley et al., 2008) result in the development and progression of pulmonary vascular remodeling. Despite the convincing evidence in animal models, blocking inflammation has been unsuccessful for all but one subtype of PAH. In a south Indian cohort of 24 female patients with systemic lupus erythematosus-associated PAH, administration of the immunosuppressant cyclophosphamide in combination with a vasodilator (conventional therapy for IPAH) resulted in an improvement in NYHA functional class and an improvement in certain haemodynamic parameters (Kommireddy et al., 2015). A few explanations as to why immunosuppression would not work in a wider group of patients despite inflammation being universally present in PAH include a potentially dual role (both protective and deleterious) of inflammation in PAH (Kumar and Graham, 2018), the importance of inflammation in the development but not the maintenance of PAH and inflammation simply being an epiphenomenon in certain subtypes of PAH.

1.4 Genetic background of PAH

Heritable PAH (HPAH) – the subtype that makes up < 10% of all PAH cases – segregates in an autosomal dominant manner. Heterozygous mutations in the bone morphogenetic protein type II receptor gene (*BMPR2*) have been identified as the major cause of HPAH (Lane et al., 2000) accounting for 53% to 86% of PAH patients with a family history and 14% to 35% of those with idiopathic PAH (Machado et al., 2015). Over six hundred germline variants (corresponding to 384 independent variants) have been described in PAH in a 2015 study (Machado et al., 2015) More recently, a whole-genome sequencing study involving 1,048 PAH cases from the heritable, idiopathic (IPAH) and anorexigen-induced PAH classes reported 76%, 12% and 8% of patients harboring previously reported and novel mutations in *BMPR2,* respectively (Graf et al., 2018). Patients with *BMPR2* mutations are normally younger at diagnosis, have worse hemodynamic parameters and a worse prognosis and are less likely to respond to vasodilator therapy compared with other patients in the same PAH class without *BMPR2* mutations (Evans et al., 2016). Incomplete penetrance (42% in women and 12% in men) (Larkin et al., 2012) of *BMPR2* and a variable age of onset between cases carrying the same *BMPR2* mutations indicates that these heterozygous germline mutations are not sufficient for the development of PAH and that their effect has to be complemented by other genetic and/or environmental factors.

BMPR2 belongs to the group of bone morphogenetic proteins (BMP) of the transforming growth factor β (TGF-β) superfamily involved in fundamental cellular processes, such as apoptosis, proliferation, differentiation and migration. The list of genes underlying PAH is growing with new genes being identified in HPAH and IPAH cases, many of which are part of the BMP signaling pathway. Established genes harboring rare pathogenic mutations contributing to the development of PAH are summarized in Table 1.2. Genes other than *BMPR2* in the TGF-β pathway include *ACVRL1* and *ENG,* which together with BMPR2 form the BMPR2–ACVRL1–ENG receptor complex (David et al., 2009) and underlie an infrequent form of PAH associated with hereditary hemorrhagic telangiectasia. In some rare cases *ACVRL1* and *ENG* mutations have been identified in PAH patients without hereditary hemorrhagic telangiectasia, although typically in early-onset cases (Machado et al., 2015). Mutations in genes, in particular *SMAD5* and *SMAD9*, coding the BMP-specific series of cytoplasmic intermediary receptors called SMADs (mothers against decapentaplegic homolog) have been identified in PAH patients in multiple cohorts (Nasim et al., 2011, Graf et al., 2018). Two sequencing studies from Europe (Graf et al., 2018) and China (Wang et al., 2019) have reported deleterious mutations in the *GDF2* gene encoding bone morphogenetic protein 9 (BMP-9) that binds various TGF-β receptors and activates SMADs.

Rare and deleterious mutations in genes not involved in the TGF-β pathway including *KCNK3* (Ma et al., 2013, Graf et al., 2018) which encodes a potassium channel, and the scaffolding protein coding *CAV1* (Austin et al., 2012, Southgate et al., 2020) have been reported by multiple studies. The transcription factor coding *TBX4*, an essential regulator of embryonic development was first reported to harbor mutations in paediatric cases of PAH (Zhu et al., 2018, Kerstjens-Frederikse et al., 2013) and later on in adult onset cases too (Graf et al., 2018).

Gene symbol	Gene name	Function	References
BMPR2	Bone morphogenetic protein type 2 receptor	Type II receptor of the TGF-β family of signaling molecules. It forms a complex type I receptors such as ALK1/2/3/6 and BMPR1A or BMPR1B and activates intracellular SMAD proteins leading to target gene expression in the nucleus.	(Lane et al., 2000, Machado et al., 2015, Graf et al., 2018)
ACVRL1/ALK1	Activin A receptor like type 1/ Activin receptor-like kinase 1	Type I receptor of the TGF-β family of signaling molecules It forms a complex with BMPR2 and ENG.	(Harrison et al., 2003, Harrison et al., 2005, Graf et al., 2018)
ENG	Endoglin	Type III (accessory) receptor of the TGF-β family of signaling molecules It forms a complex with BMPR2 and ALK1.	(Harrison et al., 2003, Graf et al., 2018)
SMAD5, SMAD9	Mothers against decapentaplegic homolog	SMAD proteins transduce signals from TGF-β family members to the nuclear chaperone SMAD4.	(Nasim et al., 2011, Graf et al., 2018)
CAV1	Caveolin 1	Scaffold protein, main component of the caveolae plasma membranes found in most cell types and a regulator of BMP2-induced SMAD signaling.	(Austin et al., 2012, Zhu et al., 2018)
KCNK3	Potassium channel subfamily K member 3	pH-sensitive potassium channel, regulates the resting membrane potential in a variety of cell types.	(Ma et al., 2013, Graf et al., 2018)
GDF2	Growth differentiation factor 2	Encodes the bone morphogenetic protein-9 ligand which binds several TGF-β receptors.	(Wang et al., 2019, Graf et al., 2018)
TBX4	T-box transcription factor 4	Transcription factor involved in embryogenesis, especially in the development of hindlimb.	(Kerstjens-Frederikse et al., 2013, Zhu et al., 2018, Graf et al., 2018)

Table 1.2 Genes hosting rare pathogenic mutations in pulmonary arterial hypertension.

The common genetic variation of IPAH and HPAH was first investigated by a genome-wide association study (GWAS) (Germain et al., 2013) involving 625 PAH patients and 1,525 healthy controls. This reported one locus, namely *CBLN2*, above genome-wide significance ($P \leq 5x10^{-8}$). This locus was not replicated by the largest to-date PAH GWAS (Rhodes et al., 2018) using whole-genome sequence and genotyping array data from multiple cohorts totaling 11,744 individuals including 2,085 PAH cases. Rhodes et al. reported two loci above genome-wide significance near *SOX17* and in *HLA-DPA1/DPB1*, respectively. Experimental analyses showed that risk variants near *SOX17* exhibit enhancer activity and regulate *SOX17* expression (Rhodes et al., 2018). *SOX17* is involved in the development of vascular endothelium and its deletion leads to abnormal pulmonary vascular development in mice (Lange et al., 2014). Rare deleterious variants in *SOX17* have also been reported (Graf et al., 2018) , highlighting its importance in the pathogenesis of PAH.

1.5 Therapeutic options for PAH

Therapeutic options at present target vascular remodeling in PAH in an attempt to reduce right ventricular afterload and postpone right heart failure. In current PAH therapy, three pathways of vasoconstriction and cell proliferation are targeted, namely the prostacyclin, NO and endothelin pathways. These three main pathways are targeted by five classes of agents (Table 1.3) often used in the form of combination therapy, which has been shown to improve outcomes (Galie et al., 2016).

Pathway	Drug type	Generic names
Endothelin 1	Endothelin receptor antagonists	Bosentan Ambrisentan Macitentan
Prostacyclin	Prostanoids	Epoprostenol Treprostinil Iloprost
	Prostacyclin IP receptor agonist	Selexipag
Nitric oxide	PDE5 inhibitors	Sildenafil
	Soluble guanylate cyclase stimulator	Riociguat

Table 1.3 Pathways of disease mechanism in PAH targeted by current pharmacotherapy, drug classes and generic names.
Definition of abbreviations: PDE5 = phosphodiesterase 5.

1.5.1 Endothelin 1 (ET-1)

Two types of endothelin (ET) receptors, the ETA and ETB subtypes are targeted by ET 1. The A subtype is highly expressed in pulmonary artery SMCs while the B subtype is expressed both in the SMCs and endothelial cells of the pulmonary artery. ETB receptors facilitate ET-1 clearance and induce the release of vasodilator NO and prostacyclin (Hirata et al., 1993). Bosentan and macitentan target both ETA and ETB subtypes, whilst ambrisentan selectively targets the A subtype.

1.5.2 Phosphodiesterase 5

Phosphodiesterase enzymes regulate the breakdown of cyclic guanosine monophosphate (cGMP), an intracellular regulator that induces smooth muscle relaxation. Phosphodiesterase type 5 (PDE5) is highly expressed in the lung of PAH patients and can be inhibited by the drugs sildenafil and tadalafil to prevent cGMP hydrolysis (Galiè et al., 2005).

1.5.3 Soluble guanylate cyclase

Soluble guanylate cyclase (sGC) converts guanosine 5'-triphosphate into cGMP, leading to smooth muscle relaxation of the pulmonary arteries. Riocigulat, a stimulator of sGC can exert its effect both by stabilizing the NO-sGC complex (NO is an activator of sGC) and by stimulating sGC directly in the absence of NO (Ghofrani et al., 2013).

1.5.4 Prostanoids

Prostanoids are prostacyclin analogues which bind to prostacyclin receptors and lead to the production of cyclic adenosine monophosphate triggering vasodilation. The intravenous drug epoprostenol is regarded the therapy of choice for severe PAH (Barst et al., 1996). The chemically more stable Treprostinil (Jing et al., 2013) which can be administered via several routes is approved only in the United States at present. Iloprost (Olschewski et al., 2002) is an inhalable alternative of epoprostenol

1.5.5 Prostacyclin IP receptor agonist

Selexipag is the newest approved treatment for PAH (Sitbon et al., 2015). It selectively binds prostacyclin IP receptors (also called prostaglandin I2 receptors) unlike prostacyclin analogues mentioned above that have an affinity for all types of prostacyclin receptors. Selexipag can be administered orally and has a longer half-life than other prostanoids.

1.6 Iron homeostasis and iron biomarkers

The circulation of iron in blood plasma is facilitated by the glycoprotein, transferrin, which binds Fe(III) and delivers it to cells where it is taken up via the transferrin receptor. Transferrin in humans is normally 30% saturated with iron. Transferrin saturation levels much lower (< 16%) or much higher (> 45%) than that indicate iron deficiency or iron overload, respectively. Dietary iron absorption is capped at 1-2 mg per day to compensate for iron loss. Since there is no mechanism for the excretion of excess iron, absorption must be tightly controlled. Macrophages recycle and store about 20-25 mg of iron internally from aged erythrocytes. The stored iron in macrophages and other iron storing cells can be released into the circulation via ferroportin (Figure 1.2). The liver is the primary site of iron storage and it secretes hepcidin, the master regulator hormone of iron absorption and release. Hepcidin inhibits iron release from iron storing cells and enterocytes in the duodenum through binding to and internalizing ferroportin for lysosomal degradation (Figure 1.2). Iron is stored in the form of ferritin inside cells and serum ferritin levels are a useful indicator of cellular iron stores; however they are raised by inflammation to minimize iron available to pathogens. Ferritin levels are therefore disproportionately high in diseases with an inflammatory component, such as autoimmune disease, infections, chronic kidney disease, and cancer (Hentze et al., 2010).

Hepcidin plays a pivotal role in regulating iron homeostasis in the body. The production of hepcidin is increased when circulating and stored iron levels are high and in the presence of inflammation and infections (Figure 1.2). The production of hepcidin is inhibited by erythropoiesis, low iron levels and hypoxia (Figure 1.2), (Camaschella, 2015). The regulation of hepcidin production occurs at the level of transcription. Hepcidin transcription heavily depends on bone morphogenetic protein (BMP) signaling. Activation of the BMP receptor (BMPR) leads to the phosphorylation of the receptor-activated SMAD (RSMAD) protein, which then forms a transcriptional complex with SMAD4 and translocates into the nucleus to activate the expression of *HAMP,* the gene coding for hepcidin (Figure 1.2), (De Domenico et al., 2007). Erythropoiesis requires considerable amounts of iron with about 25 mg per day used for haemoglobin synthesis. The erythropoiesis stimulating hormone erythropoietin (EPO), reduces hepcidin and can overrule signals of hepcidin production. EPO likely suppresses hepcidin production by inducing erythropoiesis given that blocking erythropoietic activity prevents this effect (Pak et al., 2006). Growth differentiation factor-15 and twisted gastrulation-1 are released by erythroid precursor cells and inhibit BMP dependent activation of hepcidin transcription (Hentze et al., 2010).

Hypoxia refers to the deficiency in the amount of oxygen available for the living organism. Iron and oxygen metabolism are closely connected by erythropoiesis, as iron deficiency resulting in insufficient supply of iron for erythrocyte production leads to low levels of tissue oxygenation, aka tissue hypoxia, in the absence of low oxygen levels. Studies in mice and rats demonstrated that hypoxia leads to higher iron absorption and erythropoiesis (Mendel, Simpson). Hypoxia-inducible factors (HIF) play an important role in iron sensing and regulation. HIFs are transcription factors that initiate expression of a multitude of gene products with functional roles in erythropoiesis (↑EPO, ↑TfR, ↑transferrin, ↑ferroportin, ↓hepcidin), local oxygen delivery (inducible nitric oxide synthesis and vascular endothelial growth factor) and cellular metabolism under hypoxic conditions (Peyssonnaux et al., 2008). In the presence of oxygen, HIFs are hydroxylated by iron-dependent prolyl-hydroxylases (PHD) and degraded via its interaction with the von Hippel-Lindau tumor suppressor protein. PHDs enzymes require iron as a cofactor, the lack of which results in low PHD activity and stabilization of HIF (Ivan et al., 2001).

Iron deficiency can be defined as the decrease of the body`s total iron content, whereas iron deficiency anemia refers to iron deficiency severe enough to reduce erythropoiesis. Even though diagnostic criteria for anemia are straightforward with a haemoglobin threshold of 13 g/dL in men and 12 g/dL in women, the combination of iron biomarkers and levels of cut-off used for the definition of iron deficiency are more arbitrary. Serum ferritin levels lower than 30 ng/L indicate the presence of iron deficiency in healthy individuals, although in the presence of inflammation ferritin levels are disproportionately high which makes it unsuitable as a diagnostic biomarker for iron deficiency in individuals with chronic disease. A soluble form of the membrane receptor that binds transferrin has been identified to circulate in plasma as a complex of transferrin and its receptor (Shih et al., 1990). The vast majority (~80%) of transferrin receptors (TfR) are found in the erythroid marrow. Soluble TfR (sTfR) measured in the serum reflects the amount of cell-bound TfR originating almost exclusively from erythroblasts (R'Zik and Beguin, 2001). Raised sTfR levels are seen in increased erythropoiesis and functional tissue iron deficiency and are unaffected by inflammation (Cook et al., 1990) making it a better biomarker than ferritin, iron and transferrin in diseases with an inflammatory component, such as PH.

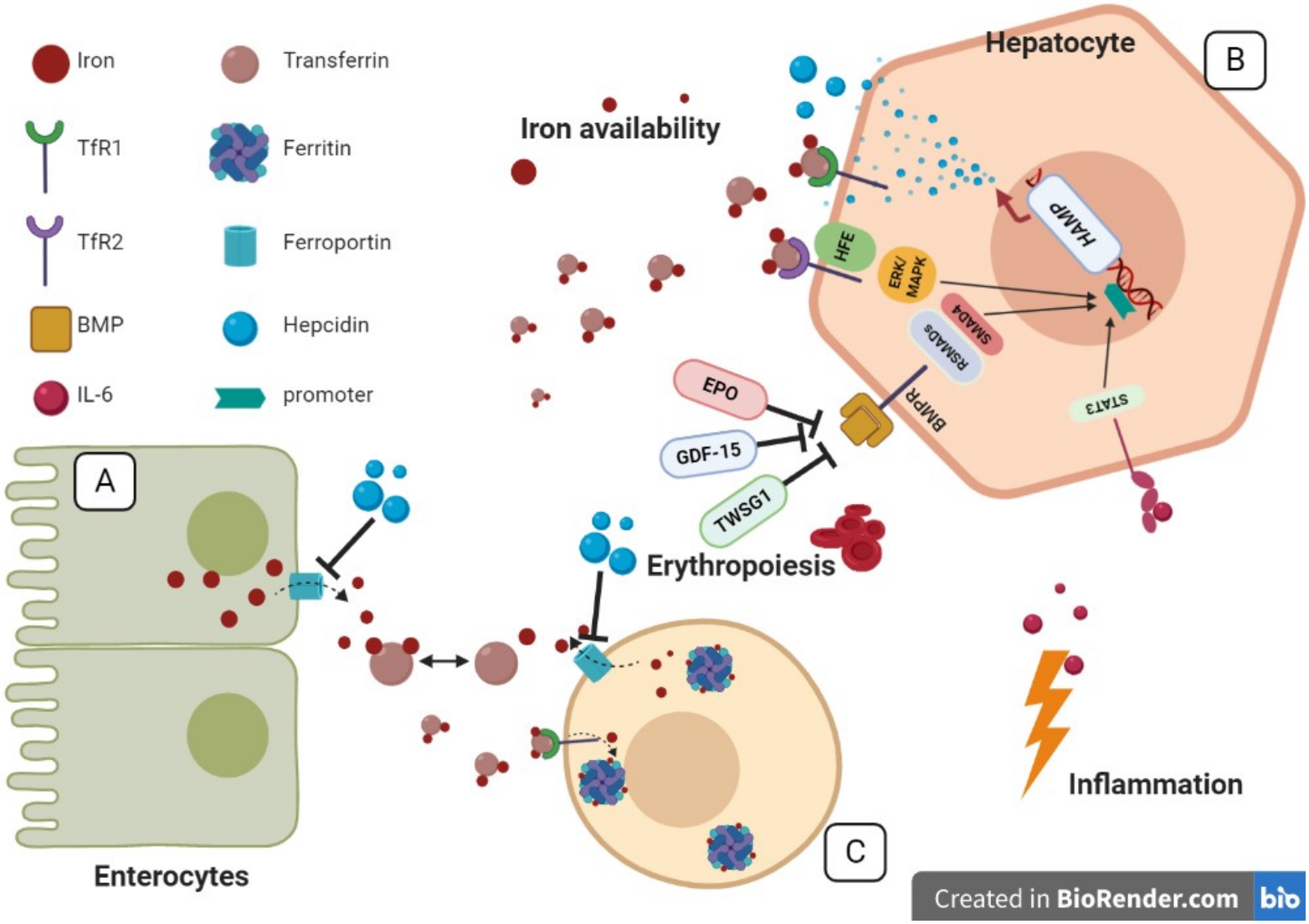


Figure 1.2 Regulation of hepcidin production.
A: Dietary iron is taken up by enterocytes in the duodenum and released into the circulation via ferroportin. Iron is loaded onto transferrin molecules that transports iron in the blood stream. B: Hepatocytes of the liver sense the saturation level of transferrin molecules and respond to systemic availability of iron accordingly. When transferrin saturation goes up, HFE interacts with TfR2 and signals via ERK/MAPK to activate HAMP expression, the gene that codes for hepcidin. Binding of BMP to BMPR leads to the generation of phosphorylated RSMADs which form dimers with SMAD4 and travels to the nucleus where it activates hepcidin transcription. Erythropoiesis blocks BMP-dependent activation of hepcidin transcription via the hormone EPO and erythroid factors GDF-15 and TWSG1. Cytokines such as IL-6 also trigger hepcidin production via STAT3 activation. C: Cells that store and utilize iron have TfR1 on their surface to bind the transferrin-iron complex and internalize iron. Iron is stored bound to ferritin inside the cell and released back into the blood stream via ferroportin. Definition of abbreviations: TfR1/TfR2 = transferrin receptor 1 and 2; BMP = bone morphogenic protein; IL-6 = interleukin-6; EPO = erythropoietin; GDF-15 = growth differentiation factor-15; TWSG1 = twisted gastrulation-1; BMPR = bone morphogenic protein receptor.

1.7 Iron deficiency in PAH

Little is known about the causal mechanisms and environmental exposures contributing to the pathogenesis of IPAH. Previous work has shown that in IPAH iron deficiency with or without anemia is both common (40% - 60%) and associated with poor prognosis, reduced exercise capacity and worsening hemodynamics (Ruiter et al., 2011, Soon et al., 2011, Rhodes et al., 2011a, Krasuski et al., 2011). Soon et al. found 50% of pre-menopausal women and 28% of men with IPAH to be iron deficient, whilst only 8% of pre-menopausal women and 2% of men in the control group consisting of chronic thromboembolic PH (CTEPH) patients had iron deficiency. Interestingly, the difference in prevalence was not significant between IPAH (14.3%) and CTEPH (6.5%) in post-menopausal women despite the larger sample size. Iron deficient status was defined by decreased ferritin levels (< 10 µg/L) with an elevated or normal transferrin (normal range 2.2 - 4.0 g/L) and normal C-reactive protein level. (Soon et al., 2011). Ruiter et al. found iron deficiency in 43% in a cohort of IPAH patients using transferrin saturation (< 15% in women and < 20% in men) and serum iron (< 10 µmol/L) as their cut-off values. Iron deficiency was equally common in pre- and post-menopausal women with a prevalence of 50% and 44%, respectively. Six-minute walk distance was significantly higher in iron replete patients; however, it did not differ between anemic and non-anemic individuals. Oral iron treatment of iron deficient patients increased ferritin levels in only 56% of patients, while in 44% it did not restore ferritin to normal levels (Ruiter et al., 2011). Rhodes and colleagues used soluble transferrin receptor (sTfR) levels above 28.1 nmol/L for the definition of iron deficiency and found 63% of their IPAH cohort to be iron deficient. They found hepcidin to be abnormally high in IPAH compared to healthy controls and to not correlate with interleukin-6 or C-reactive protein levels, suggesting that the phenomenon was not due to inflammation. A significant proportion of IPAH patients with iron deficiency had inappropriately high hepcidin levels, suggesting abnormal hepcidin production could contribute to iron deficiency observed in IPAH. High sTfR levels were associated with increased functional class (increased disease severity), lower 6-minute walk distance and higher mortality in IPAH. Rhodes et al. also showed that *BMPR2* knockout in liver cells leads to increased hepcidin levels as a response to BMP-6 stimulation (Rhodes et al., 2011a). BMPs bind activin type II receptors, in addition to BMPR, leading to SMAD-mediated hepcidin expression. Yu and colleagues demonstrated that ablation of BMPR2 in pulmonary artery smooth muscle cells augmented signaling by BMP-6 and BMP-7 via the activation of activin type II receptors (Yu et al., 2005). Atkinson et al. measured BMPR2 expression in the lungs of associated-, idiopathic and heritable PAH patients and found that BMPR2 expression was the lowest in the lungs of those harboring *BMPR2* mutations, but mutation-negative idiopathic and associated PAH patients also had

significantly lower BMPR2 levels than controls with associated PAH patients showing the smallest difference (Atkinson et al., 2002). Reduced levels of BMPR2 in PAH patients may result in abnormally high hepcidin levels even in the presence of iron deficiency, which might explain the ineffectiveness of oral iron therapy in IPAH patients observed by Ruiter et al. (Ruiter et al., 2011).

The biological rationale behind iron deficiency as a risk factor for PAH is served by multiple shared mechanisms between iron homeostasis and PAH. The iron deficient state is akin to being in a hypoxic environment. A physiological link between iron status and response to hypoxia exists in healthy volunteers whereby iron supplementation attenuates the pulmonary vasoconstriction and vascular resistance induced by hypoxia (Smith et al., 2008, Smith et al., 2009, Frise et al., 2016). Table 1.4 summarizes the evidence available to date on iron supplementation in attenuating pulmonary vasoconstriction in response to hypoxia in healthy individuals. This relationship between hypoxia and iron availability could be driven by iron's role in stabilizing HIFs, thereby mimicking the hypoxic state (see Iron homeostasis and iron biomarkers above) (Ivan et al., 2001).

Pathways of vascular remodeling – a hallmark of PAH – and iron homeostasis are interconnected by the role of the hepcidin/ferroportin axis in proliferative responses of human pulmonary artery SMCs. Hepcidin and interleukin-6 have been shown to induce pulmonary artery SMC proliferation by cellular iron retention achieved by hepcidin causing the degradation of ferroportin. This proliferative response was successfully reversed by stabilizing ferroportin expression thus inducing cellular iron excretion (Ramakrishnan et al., 2018). Ruiter et al. observed that oral iron therapy was effective in only half of IPAH patients with iron deficiency measured by serum ferritin levels, suggesting reduced iron absorption (Ruiter et al., 2011). Rhodes et al (Rhodes et al., 2011a) have argued that dietary iron absorption might be impaired by abnormally high levels of hepcidin in IPAH.

Iron homeostasis and endothelial dysfunction in PAH may also be connected through the imbalance between vasoconstrictors and vasodilators in PAH, namely endothelin 1 (ET1) and nitric oxide (NO). High iron levels were shown to reduce the expression of NO-synthase (Weiss et al., 1994) and iron loading in individuals with hereditary hemochromatosis has been linked to increased thickness of the carotid artery and reduced vasodilation through impaired endothelial function (Gaenzer et al., 2002).

The results of intervention studies investigating the effects of iron supplementation in PAH patients are not conclusive, with some improvement shown in certain outcomes (Table 1.4). However, evidence from a randomized controlled trial (RCT) of iron supplementation in PAH is lacking. An open label pilot study involving twenty IPAH patients with iron deficiency reported improved iron status, 6-minute walk distance (6MWD) and quality of life (QoL) two months after a single infusion of

1000 mg ferric carboxymaltose (Viethen et al., 2014). A subsequent, again open label, clinical study found no significant change in 6MWD (P = 0.07) three months after receiving 1000 mg of intravenous iron in fifteen IPAH patients with iron deficiency, although it reports restored iron levels of patients as well as improved exercise endurance capacity and QoL scores (Ruiter et al., 2015). A case report described substantial improvement of the symptoms of a 59-year-old woman with scleroderma-induced PAH and iron deficiency anemia after being treated with oral ferrous sulphate, indicating iron deficiency as a risk factor and/or an easily treatable comorbidity in PAH (Table 1.4) (Mehmood et al., 2016).

Author	Study type	Participants and study design		Endpoints	Results	Reference
PAH						
Viethen et al	Intervention study, not randomized, not placebo controlled	20 PAH patients (14 IPAH, 6 CTD PAH) with ID received single infusion of ≤ 1000mg ferric carboxymaltose. 20 matched PAH patients (13 IPAH, 4 CTD PAH, 3 CHD PAH) without ID did not receive infusion and served as controls. Assessment at baseline and at 8 weeks.		Iron status, 6MWD, CPET, QoL	All iron parameters improved (P < 0.001); 6MWD improved (P = 0.007); QoL improved (P = 0.01); Anaerobic threshold improved (P = 0.03).	(Viethen et al., 2014)
Ruiter et al	Intervention study, not randomized, not placebo controlled	30 IPAH patients with ID received single infusion of ≤ 1000mg ferric carboxymaltose. Assessment at baseline and at 12 weeks.		Iron status, 6MWD, NYHA class, CPET, RV function, QoL, Skeletal muscle oxygen handling	Time to anaerobic threshold increased (P < 0.001); Exercise endurance capacity improved (P < 0.001); QoL improved (P < 0.05); Myoglobin concentration improved (P < 0.05); Mitochondrial oxidative capacity (P < 0.05) improved; Myoglobin protein content improved (P < 0.05).	(Ruiter et al., 2015)
Mehmood et al	Case report	A 59-year-old woman with scleroderma-associated PAH and an implantable hemodynamic monitor had worsening NYHA class III symptoms and was found to have ID anemia. Oral ferrous sulfate (325 mg three times a day) treatment was initiated.		NA	After 6 months haemoglobin and iron levels normalized, cardiac output increased, RV function improved, NYHA functional class improved to class II.	(Mehmood et al., 2016)
Hypoxic pulmonary vasoconstriction						
Smith et al	Controlled crossover study	Iron group: 8 healthy volunteers not randomized, received saline first, 1 week later 200 mg iron(III)-hydroxide sucrose. Protocol: infusion, acute hypoxic challenge, 8 hrs in hypoxia chamber followed by another round of acute hypoxic challenge.	DFO group: 8 healthy volunteers randomized, received either saline or 4g/70kg desferrioxamine mesylate 1 week apart. Protocol: acute hypoxic challenge, infusion over 8 hrs followed by another round of acute hypoxic challenge.	PASP ΔPASP	Iron prevented the rise in PASP after 8 hrs of hypoxia (P < 0.001) and also in ΔPASP during second acute hypoxic challenge (P = 0.002). DFO increased basal PASP (P < 0.001) and ΔPASP (P = 0.01).	(Smith et al., 2008)

Smith et al	Randomized, double-blind, placebo-controlled trial	Healthy group: 22 sea level resident men studied for one week at Cerro de Pasco, Peru (altitude 4,340 m). On the 3rd day participants received 200 mg iron(III)-hydroxide sucrose or saline placebo.	CMS group: 11 high altitude resident men suffering from CMS at Cerro de Pasco studied for one month. CMS is often treated with venesection which induces ID. Venesection was performed during the first 4 days (2L blood in total) followed by either 400 mg iron(III)-hydroxide sucrose or saline placebo. On day 25 the saline and iron groups were crossed over.	Iron status PASP CO	In the healthy group, rise in PASP due to ascend to high altitude was attenuated by iron infusion (P = 0.01). Iron status was reduced by the 3rd day at high altitude (P < 0.01). In the CMS group, iron was depleted (P < 0.001) by venesection accompanied by an increase in PASP (P = 0.03). PASP was not significantly different (P = 0.64) between the group that received iron first and the one that received saline first.	(Smith et al., 2009)
Frise et al	Controlled observational study	13 ID and 13 iron replete healthy participants. Both groups were assessed before and after 6 hrs hypoxic exposure on the 1st day. On the 2nd day, both groups received iron infusions and the assessment was repeated.		Iron status PASP CO	On the 1st day, baseline PASP did not differ between groups. After 6 hrs hypoxia PASP was higher in the ID group (P < 0.001). On the 2nd day, rise in PASP was reduced in both groups by iron infusion and the absolute reduction was greater in the ID group (P = 0.035).	(Frise et al., 2016)

Table 1.4 Summary table of intervention studies to date investigating the effects of iron supplementation on outcomes in pulmonary arterial hypertension and studies investigating the role of iron availability in the pulmonary vasoconstrictive response to hypoxia in healthy individuals.
Definition of abbreviations: PAH = pulmonary arterial hypertension; CTD PAH = connective tissue disease-associated PAH; CHD PAH = congenital heart disease-associated PAH; ID = iron deficiency; 6MWD = 6-minute walk distance test; CPET = cardiopulmonary exercise testing; QoL = quality of life questionnaire; NYHA class = New York Heart Association functional classification; RV = right ventricular; DFO = desferrioxamine (iron chelator); PASP = pulmonary artery systolic pressure; ΔPASP = PASP response to acute hypoxia; CMS = chronic mountain sickness; CO = cardiac output.

1.8 Causal inference in epidemiology

The cause of an event can be defined as a factor or characteristic that was necessary for the event to occur at the time it occurred, holding all other factors constant. It is important to note the difference between a causal factor and a causal mechanism, a causal mechanism is sufficient by itself to cause the event and is often referred to as the sufficient cause, whilst a causal factor is a component of the causal mechanism insufficient to cause the event, often termed as the component cause. The true onset of a disease can be thought of as the completion of the sufficient cause and refers to the start of the disease process rather than the appearance of the symptoms (Rothman and Greenland, 2005).

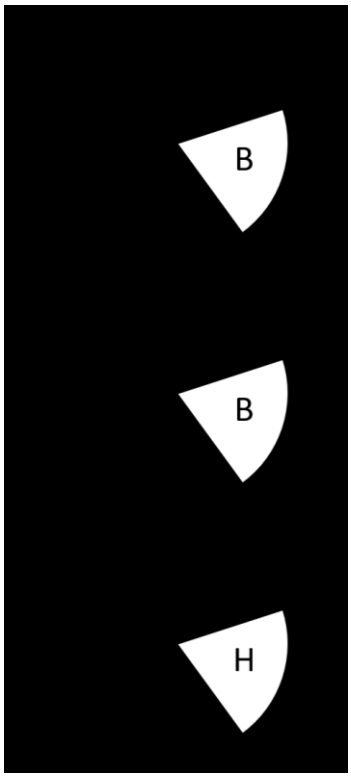


Figure 1.3 Causal pie model. *Three minimally sufficient causal mechanisms leading to the same disease in three hypothetical individuals. Each wedge in the pie charts represents a component cause.*

In biology and disease, most if not all components of the sufficient cause are unknown (Rothman, 1976). Various causal mechanisms can underlie a given disease comprising a number of component causes that jointly produce the disease. Figure 1.3 illustrates this multicausality model, where either one component cause is neither necessary nor sufficient for the disease to occur. For instance, component cause B in Figure 1.3 is not sufficient in itself to cause the disease in individual I. and II., hence the other five components in both sufficient causes. Component B may not be necessary either, as there are multiple causal mechanisms leading to the disease and not all of them include component B, as in individual III. in Figure 1.3. However, a cause need not be necessary or sufficient to be responsible for some cases of disease which can be prevented by blocking that single component.

Identifying possible environmental factors (hereon referred to as exposures) that could play a role in disease development is not an easy task. Unless the disease of interest is a monogenic disease with very high penetrance, the causes are likely to be part genetic and part environmental with the interaction of these two making identifying individual components even more challenging. In epidemiology, observational studies where the researcher has no control over the study population are the main source of putative causal factors of human disease for hypothesis generation. Observational studies are inherently (i.e. by design) prone to biases giving rise to spurious associations between exposures and outcomes that fail to replicate in RCTs. Beta-carotene

supplementation for the prevention of malignant tumors and cardiovascular disease (Hennekens et al., 1996) and antioxidant vitamin supplementation for the improvement of outcomes in individuals suffering from diabetes and/or coronary heart disease (Group, 2002) are examples of spurious associations from observational studies that led to negative RCTs. The reasons that lead to the estimate of an association and the true causal effect not being equal are collectively called biases. Reverse causation is one type of bias observational studies suffer from and it refers to the uncertain temporality of the relationship between exposure and outcome. For example, associations between biomarkers and disease do not provide information about the direction of the effect (if any) as perturbed biomarker levels are often caused by the disease process. The term confounding is used in epidemiology for when the exposure and the outcome share the same cause but are not causally linked. In Figure 1.4, a hypothetical observational study is divided into two groups based on whether or not they are exposed to a risk factor of interest (gray are exposed, green are unexposed). There are many more disease cases (red cross) observed in the exposed group which might suggest that the putative risk factor plays a role in the development of the disease. However, an unmeasured or unknown confounder (red outline) drives this association causing both the disease and the putative risk factor. The 2x2 contingency table in Figure 1.4 shows that 55% of individuals with the confounder develop the disease while this is only 9% in those without the confounder. Since the exposed (gray) group has a higher proportion of confounded individuals than the unexposed (green) group, the exposure shows an association with the disease when the confounder is not taken into account.

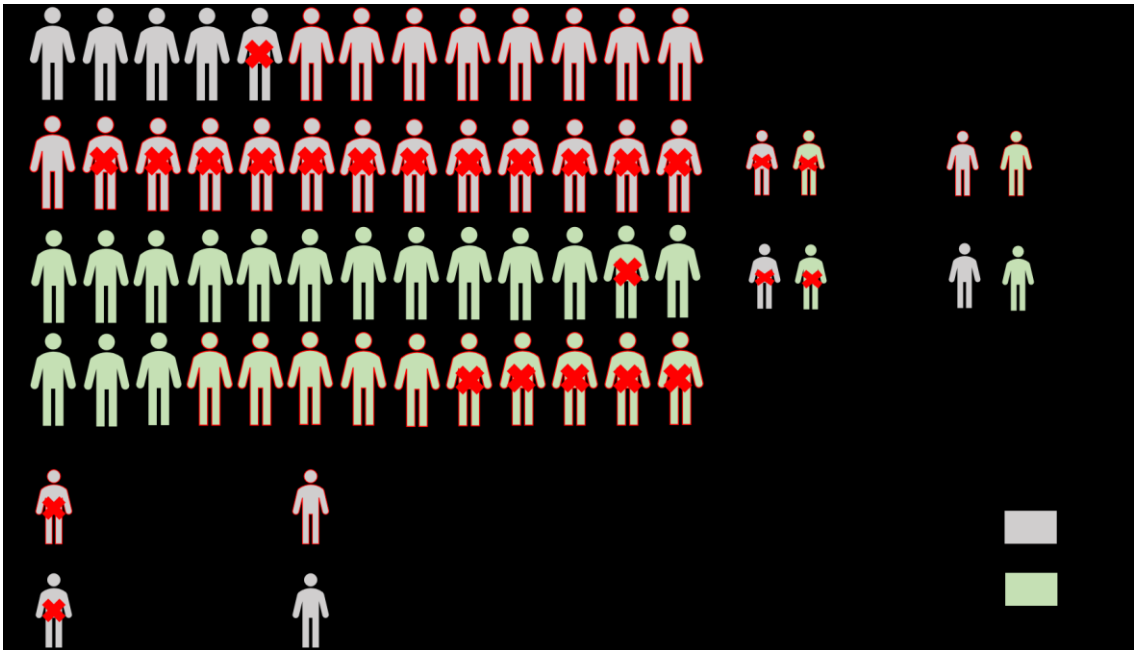


Figure 1.4 Schematic diagram of confounding in epidemiology.
Gray individuals form the exposed group in an observational study whilst green individuals are unexposed to the exposure of interest. The red cross signifies disease or outcome of interest. Individuals with a red outline are exposed to the confounder or common cause of both the exposure and outcome of interest. The 2x2 table on the right shows disease counts for individuals with and without the disease, stratified by the confounding factor.

Another source of bias in observational studies and major problem in today`s epidemiological research is called data-dredging. It refers to looking at a large number of associations in datasets with many exposure and outcome variables of which only a few associations passing the universally accepted significance threshold of $P<0.05$ will be real associations. The significance test of $P<0.05$ is often misinterpreted as the collection of tests which are false positives in only 5% of the cases, whereas the $P<0.05$ can be compatible with the majority of associations being false positives (Smith and Ebrahim, 2002).

1.9 Mendelian Randomisation

The use of genetic variants as instrumental variables as an alternative to clinical trials for testing causality have gained popularity in recent years. Instrumental variables outside of genetic epidemiology have been widely used in econometrics and rely on the central idea that when a variable - called the 'instrument' – unrelated to the biases of the exposure and outcome association is identified and used instead of the exposure itself, an unbiased causal inference can be made (Angrist and Krueger, 2001). The collection of methods for the application of genetic variants as unbiased instruments was termed Mendelian Randomisation (MR). The approach of MR is built on Mendel`s second law – the law of random assortment - which states that traits segregate independently from one another. These randomly segregating traits in Mendel`s second law are the different alleles present in a population which refer to the different forms of genetic variants detectable at a genetic locus. In brief, each individual inherits two alleles (one on each chromosome in the pair) at a specific genetic locus often referred to as one`s 'genotype', which could be homozygous in case the two alleles are identical or heterozygous when the two alleles are different. A genotype is said to be polymorphic when there is more than one allele present in the population. These polymorphisms can produce phenotypic differences which mirror the biological effects of environmental influences or exposures; therefore, they can be used as instruments in testing exposure-outcome relationships for causality. The above-mentioned features of genetic polymorphisms provide the opportunity for conducting natural experiments which can be thought of as randomized trials where individuals are randomly allocated at conception to a treatment or exposure group based on their genotypes. MR has been successfully used to help prioritize or deprioritize intervention and drug targets for several diseases (Pichler et al., 2013, Collaboration et al., 2011, Interleukin-6 Receptor Mendelian Randomisation Analysis et al., 2012, Lewis et al., 2011, Palmer et al., 2013). In general, candidate drugs with genetic evidence for effectiveness are more successful in drug trials compared to those without such genetic support (King et al., 2019).

The association between higher levels of the amino acid homocysteine and increased coronary heart disease (CHD) risk is a great example of how MR can be used to test an exposure-outcome relationship for causality. The homocysteine-CHD relationship is of great public health interest as it would be an easy intervention target for lowering CHD burden, since homocysteine can be easily lowered by a moderate increase in folate consumption (Collaboration, 1998). However, an observed association does not necessarily imply that intervening on the exposure, in this case lowering homocysteine levels with folate, would lead to a change in the outcome, in this case CHD. To robustly assess whether homocysteine is causally linked to CHD, lengthy and expensive randomized

trials of folate supplementation are needed. Alternatively, a functional genetic polymorphism in the *MTHFR* gene (at base 677), which is involved in homocysteine metabolism provides the opportunity of a natural experiment. Individuals with the *MTHFR* substitution C->T have reduced activity of the enzyme methylene tetrahydrofolate reductase involved in the conversion of folate to 5-methyl tetrahydrofolate. Since 5-methyl tetrahydrofolate is needed for the conversion of homocysteine to methionine, reduced levels lead to higher levels of homocysteine. Therefore, if higher homocysteine levels lead to increased CHD risk, individuals with the 677-TT genotype should be at higher risk of developing CHD than those with the 677-CC genotype. Given the random assortment of alleles as stated by Mendel`s Second Law, the two groups based on the *MTHFR* genotype should not differ with respect to any known or unknown confounders of the homocysteine-CHD relationship, such as socioeconomic and behavioral factors. Also, since the genotype of an individual is determined at conception, the association between genotype and disease should always go in the direction of genotype to disease.

The schematic of an MR study using the *MTHFR* genotype as an instrument for homocysteine levels is presented in Figure 1.5. Several MR studies investigating the relationship of homocysteine and CHD were meta-analyzed by Klerk et al. (Klerk et al., 2002) who found that the combined causal effect from 40 MR studies that used the *MTHFR* genotype as an instrument for impaired folate metabolism resulting in high homocysteine levels was very close to the effect estimate from the largest meta-analysis of observational studies (Collaboration, 2002), suggesting the protective effect of folate against CHD. The amount of homocysteine increase attributable to the 677-TT genotype (2.6 µmol/L) was estimated to increase CHD risk by 16% (95% CI = 1.05 – 1.28) in the MR meta-analysis and by 17% (95% CI = 1.10 – 1.25) in the meta-analysis of observational studies.

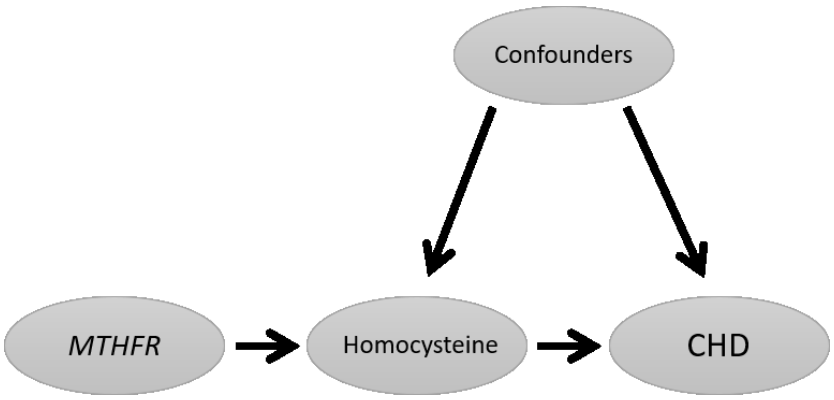


Figure 1.5 Mendelian Randomisation analysis of homocysteine and coronary heart disease.
Definition of abbreviations: MTHFR = methylene tetrahydrofolate reductase; CHD = coronary heart disease.

Equal distribution of confounding factors between genotype groups and no reverse causality in MR studies only hold true when genetic instruments are valid. The core assumptions of MR methods are discussed in detail in the *Methods* section.

1.10 Expression quantitative trait loci

Genome-wide association studies (GWAS), facilitated by the growing availability of genotyping and sequencing technologies, identify ever more common and low-frequency genetic variants associated with various disease definitions and quantitative phenotypes. As of March 5, 2020, the curated database of NHGRI-EBI GWAS Catalog (Buniello et al., 2019) contained 97,567 unique genome-wide significant genotype-phenotype associations reported by over three thousand published GWAS (Figure 1.6). The genome-wide significance threshold ($P \leq 5x10^{-8}$) reflects the number of independent common (minor allele frequency ≥ 5%) genetic variants throughout the genome- and therefore the number of independent tests, in European populations (Altshuler et al., 2005). Although GWAS discoveries have led to novel biological insight into disease mechanisms (Visscher et al., 2017), most genetic loci are located in the non-coding space of the genome (Maurano et al., 2012) and are not directly informative or pinpoint the causal genes, making the interpretation of GWAS results very challenging. Over three quarters of such non-coding genome-wide significant associations lie within DNase I hypersensitivity sites, characteristic of open chromatin structures in active regions of the genome and frequently overlap transcription factor recognition sequences within these sites, suggesting a regulatory instead of a gene product altering role for the majority of GWAS loci (Maurano et al., 2012).

Expression quantitative trait loci (eQTL) refer to genetic variants associated with the expression of a gene reflecting transcriptional regulatory relationships and are identified using methods similar to the GWAS strategies of other quantitative phenotypes (Michaelson et al., 2009). In recent years, eQTL mapping studies of transcriptome-wide gene expression in population-based studies have identified regulatory variants for a large proportion of tested genes and found that these eQTL often colocalize with GWAS loci of disease susceptibility and other quantitative traits (Zhernakova et al., 2017, Joehanes et al., 2017, Westra et al., 2013), offering a handle on linking GWAS loci to biology.

However, it should be noted that genetic control of disease susceptibility and all other phenotypes should be investigated in the context of environmental stimuli with which genetic effects are interlinked (Tam et al., 2019). For example, genetic susceptibility to obesity is highly dependent on and interacts with external environmental risk factors in the development of obesity, a condition

highly influenced by the environment (Reddon et al., 2016). Apart from the well-established tissue- or cell type specificity of eQTL effects (Aguet et al., 2019), environmental context, such as the activity of certain signalling pathways, can also modify transcriptional genetic control, especially genetic effects of genes that are differentially expressed upon the stimulus or stimuli under investigation (Grundberg et al., 2011, Zhernakova et al., 2017). This implies that eQTL studies of affected individuals characterized by signalling pathways important for the maintenance of disease pathobiology could be valuable sources of novel eQTL effects unobserved in healthy population-based studies. Previous work shows that various stimuli including inflammatory cytokines (Fairfax et al., 2014), viral (Çalışkan et al., 2015) and bacterial (Barreiro et al., 2012) infections uncover so-called `response eQTL` effects otherwise unobserved in the absence of the stimulus in question. Furthermore, such response eQTL were enriched in GWAS loci of related and relevant phenotypes, for instance response eQTL to Mycobacterium tuberculosis infection colocalized with GWAS loci for pulmonary tuberculosis susceptibility in the study of Barreiro et al.



Figure 1.6 NHGRI-EBI Catalog of published genome-wide association studies (GWAS).
Diagram showing genotype-phenotype associations discovered by GWAS on the twenty-two autosomal- and the X chromosomes with p-values less than $5x10^{-8}$, colored according to trait categories. The diagram was downloaded and contains all associations hosted on the GWAS Catalog website (https://www.ebi.ac.uk/gwas) on March 5, 2020.

1.11 Summary of background

Dysfunctional iron homeostasis is a hypothetical therapeutic target of interest in PAH; however, iron deficiency and its causes in other forms of PH have not been well characterized and it could be present in other pre- and also post-capillary forms of PH. The significance of iron deficiency in PAH development or prognosis is not yet clear as previous observational and open-label studies investigating this relationship were not designed to test causality. Current efforts characterizing common genetic variation underlying PAH susceptibility and the completion of the first randomized controlled trial of iron replacement therapy in PAH enable these questions of causality to be addressed. Genetic approaches, such as Mendelian Randomisation, to testing causality requires not only large sample sizes but also well-defined quantitative trait loci (QTL) for intermediate phenotypes (gene expression, protein levels, metabolites, etc.). These QTL often colocalize with susceptibility loci discovered in GWAS and they have been shown to heavily depend on tissue or cell type as well as the environmental context i.e. molecular milieu within the tissue/cell. This implies that QTL can be used to identify candidate genes and pathways of interest in disease development, but also that certain QTL effects cannot be detected in healthy individuals without the relevant stressors or stimuli and may be better characterised in the disease state.

1.12 Hypothesis

Iron deficiency is a common finding in PH but has a causal role in the pathogenesis of PAH and iron replacement will improve measures of disease severity. The genetic control of transcription differs between healthy individuals and individuals with PAH.

1.13 Aims

1. Assess the prevalence and severity of iron deficiency in a large cohort of pulmonary hypertension patients with varied etiologies referred to the National Pulmonary Hypertension Service at Hammersmith.
2. Test the observed relationship between red cell distribution width, a proxy for iron deficiency, and the development of PAH for causality in a Mendelian Randomisation framework.
3. Analyze the results of the first randomized controlled trial of iron supplementation in iron deficient idiopathic PAH patients.

4. Uncover potential PAH-specific expression quantitative trait loci using transcriptome-wide gene expression and whole genome sequencing data from the UK PAH Cohort and the UK National Institute for Health Research BioResource studies.

2 Materials and methods

2.1 Definition of patient groups

Pulmonary hypertension (PH) is defined as an increase in mean pulmonary arterial pressure (≥20 mmHg) (Simonneau et al., 2019) at rest and is divided into five subgroups based on similarities in pathology and clinical presentation devised by the World Health Organization (WHO) (*Introduction* Table 1.1).

Most analyses in this thesis were focused on WHO group 1 PH including idiopathic, hereditary or anorexigen-associated pulmonary arterial hypertension (PAH) patients. PAH was defined by internationally agreed criteria, specifically mean pulmonary artery pressure >25mmHg, pulmonary vascular resistance >3 Woods and pulmonary capillary wedge pressure <15 mmHg. PAH and other PH diagnostic subgroups were defined by haemodynamic criteria measured at right heart catheterization according to international guidelines (Galie et al., 2016).

PH groups 3 and 5, namely PH due to respiratory disease and/or hypoxemia and PH with unclear and/or multifactorial mechanisms (Table 1.1) were not included in these analyses, because most patients belonging to these two groups get referred to other PH services (e.g. Royal Free Hospital) and thus the number of samples available is not sufficient to do a well-powered survey of iron status in these patient groups.

2.2 Definition of iron deficiency in the National Pulmonary Hypertension Service

Blood biomarkers were measured at the clinical pathology laboratory at Hammersmith Hospital. Soluble transferrin receptor (sTfR) levels and interleukin-6 (IL-6) levels were determined by enzyme-linked immunosorbent assay and hepcidin was measured by competitive radioimmunoassay. Ferritin, iron and transferrin were measured by standard clinical pathology accredited hospital assays.

In the National Pulmonary Hypertension Service (NPHS) cohort, iron deficiency was defined by the combination of low ferritin (< 100 μg/L) and low transferrin saturation (TSAT) (< 20%) or by high sTfR levels (> 28.1 nmol/L, upper limit of normal range in healthy volunteers of clinical lab assay) alone. Thresholds for ferritin and TSAT were more liberal than those used for the definition of iron

deficiency in healthy individuals (ferritin < 30 µg/L and TSAT < 16%) to account for iron sequestration in chronic disease. The NPHS cohort was divided into four groups based on their iron status; those who are classified as iron deficient based on their sTfR levels only (referred to as 'sTfR-only'), those iron deficient based on their ferritin and TSAT levels only ('Ferr-TSAT-only'), those meeting both criteria ('Ferr-TSAT+sTfR') and finally those who meet neither and are therefore considered iron replete.

2.2.1 Iron biomarkers in chronic disease

Diagnosing iron deficiency based on clinical blood biomarkers in individuals with chronic disease is a challenging task as chronic disease itself and iron deficiency are the two most common causes of anaemia (Cash and Sears, 1989). Iron deficiency in individuals with chronic disease may be diagnosed by a bone marrow biopsy stained for iron, determining the absence of iron stores, although the procedure is invasive and expensive. Conventional blood biomarkers used for the assessment or iron status in healthy individuals including serum iron, transferrin and ferritin are confounded by the cytokines active in inflammatory or immune responses.

The circulating (soluble) form of transferrin receptors (sTfR) has been reported as a marker of iron deficiency unaffected by inflammation, providing an alternative for bone marrow biopsies in patients with acute or chronic inflammation (Punnonen et al., 1994, Cook et al., 1990). Transferrin receptors facilitate cellular iron uptake and are expressed in response to cellular iron requirement. Circulating sTfR is formed through the proteolysis of the membrane-bound receptor and originate primarily from erythroblasts in the bone marrow (Shih et al., 1990, R'Zik and Beguin, 2001). Levels of sTfR moderately increase in individuals with iron deficiency without anaemia and are more drastically elevated in iron deficiency anaemia when marrow functional iron deficiency develops (Punnonen et al., 1994), although high values are not specific to iron deficiency and are also observed in conditions associated with increased erythropoiesis (Mast et al., 1998). In a study using bone marrow biopsies or response to iron therapy for the detection of iron deficiency, sTfR showed high sensitivity (~100%) identifying most iron deficient individuals, but varied specificity in healthy individuals (specificity = 84%), anaemic individuals (specificity = 42%) and in individuals with microcytic anaemia (specificity = 65%) (Mast et al., 1998).

Studies in patients with inflammatory diseases indicated that sTfR levels might not distinguish accurately among patients with and without iron deficiency. In rheumatoid arthritis, sTfR levels show a positive correlation with inflammatory cytokines (Zoli et al., 1994) and levels are similar between iron replete and deficient patients whilst being generally higher in both patient groups than in healthy individuals (Nielsen et al., 1994). In a study of patients with chronic inflammatory diseases,

sTfR was elevated in 27% of individuals with normal or increased iron stores assessed by bone marrow staining (Juncà et al., 1998). Remacha et al. suggested the use of a higher sTfR cut-off (55 mmol/L) for the diagnosis of iron deficiency in individuals with chronic disease than that used in healthy individuals (Remacha et al., 1998). Combining ferritin and sTfR in a ratio (sTfR/ferritin) was proven useful in diagnosing iron deficiency in rheumatoid arthiritis patients (Bultink et al., 2001, Suominen et al., 2000). Bultink et al. reported improved specificity for sTfR/ferritin (94%) and the combined cut-off (97%) of sTfR and ferritin (sTfR > 29.4 nmol.L and ferritin < 50 µg/L) compared to the estimated specificities of sTfR (81%) and ferritin (84%) alone (Bultink et al., 2001). While iron stores are sufficient, sTfR is a marker of erythropoietic activity until the iron needs of erythropoiesis are not met and it becomes a marker of iron deficiency. In patients with chronic diseases, sTfR levels may remain in the normal range despite depleted iron stores due to inflammatory cytokines suppressing erythropoiesis (Nielsen et al., 1994) or may be elevated despite adequate iron stores because of increased erythropoietic activity (Mast et al., 1998). The combined use of ferritin and sTfR was also suggested based on bone marrow biopsies in a heterogenous patient cohort with various diagnoses including blood cancers, human immunodeficiency virus and other infections and chronic inflammatory diseases, where sTfR/ferritin showed both high specificity (93%) - comparable to that of ferritin alone - and high sensitivity (67%) - comparable to that of sTfR - in identifying iron deficiency (Means Jr et al., 1999). Similar studies using bone marrow staining to evaluate the sensitivity and specificity of iron indices in identifying iron deficiency have not been conducted in pulmonary hypertension to date. However, the results from the studies mentioned above (Nielsen et al., 1994, Juncà et al., 1998, Remacha et al., 1998, Mast et al., 1998, Means Jr et al., 1999, Suominen et al., 2000, Bultink et al., 2001) suggest that a higher sTfR cut-off than that used in healthy individuals might be needed in disease and that sTfR and ferritin in comination are the best blood biomarkers for distinguishing iron deficiency from anaemia of chronic disease in a wide range of chronic diseases.

2.3 Whole-genome DNA sequencing in the NIHRBR study

2.3.1 Next generation sequencing

Next generation sequencing (NGS) is an umbrella term referring to all novel technologies used for determining the sequence of nucleotides in a stretch of DNA. These massively parallel and high throughput technologies reduced the cost of sequencing - historically done by the Sanger dideoxy method - which meant it became widely available to research groups. The human genome is comprised of three billion nucleotide bases with millions of nucleotides being different between any two individuals` genomes. Large-scale human genome sequencing projects such as the 1000 Genomes Project (Auton et al., 2015) or the most recent Haplotype Reference Consortium (HRC) (the Haplotype Reference et al., 2016) aimed at characterizing and aggregating this variation in thousands (in the 1000 Genomes Project) or tens of thousands of people by the HRC. The 1000 Genomes Project provided the foundation for method development into analyzing sequencing data and has been an invaluable source of global human genetic variation for inferring untyped (i.e. not present on the array used for genotyping) genetic variants in genome-wide association studies (GWAS) (Birney and Soranzo, 2015). Inference of untyped genotypes, or as commonly called genotype imputation, has become widely available owing to the HRC reference panel (the Haplotype Reference et al., 2016) and the freely available services of the Sanger Imputation Service (https://imputation.sanger.ac.uk/) and the Michigan Imputation Server (https://imputationserver.sph.umich.edu/).

NGS technologies that started to emerge after 2005 can be divided into two main categories: luminescence-based technologies used by Illumina, Roche, Life Technologies (SOLiD) and BGI and the electrical-charge-based technologies used by Oxford Nanopore, Pacific Biosciences, Life Technologies (Ion Proton, Ion PGM), Helicos and ZS. Main steps prior to sequencing involve i) the preparation of the DNA library (DNA fragmentation and attachment of primer and adapter sequences), ii) the amplification of the DNA fragments either by using primers attached to a solid surface (Illumina), called solid-phase polymerase chain reaction (PCR) or by emulsion PCR using microbeads (Roche454/Ion Torrent/Life Technologies).

The luminescence-based technology called ‘reversible termination sequencing’ developed by Illumina currently dominates the market with superior output and lower reagent costs than that of their competitors. To ensure a detectable signal, DNA fragments are amplified on the array by PCR bridge amplification and form clusters of identical DNA sequences. Illumina requires short DNA fragments, typically 100 bp, and a typical run can last for multiple days with the industrial NextSeq

and HiSeq series. Illumina sequencing chemistry works on the basis of sequence by synthesis, whereby DNA fragments are extended by repeated cycles of single base incorporation directed by polymerase. Reversible terminators (nucleotides: A, C, G and T) are fluorescently labelled with different color fluorophores, which are removed in each cycle after imaging. The sequence of each cluster of DNA fragment is determined by quantifying the fluorescent signal from each cycle and applying a base-calling algorithm (Bentley et al., 2008).

2.3.2 The NIHRBR study

PAH patients with whole genome sequence (WGS) data contributing to the analyses in this thesis were recruited into the UK National Institute for Health Research BioResource (NIHRBR) study (https://bioresource.nihr.ac.uk/) which facilitated the whole-genome sequencing of 1,026 PAH patients from the idiopathic-, heritable- and drug-induced WHO class 1 pulmonary hypertension subclasses. Detailed description of the NIHRBR cohort as well as WGS procedures and variant calling have been described elsewhere (Graf et al., 2018).

In brief, DNA extracted from venous blood was prepared using the Illumina TruSeq DNA PCR-Free Sample Preparation kit (Illumina Inc., San Diego, CA, USA) and whole-genome sequenced with Illumina HiSeq 2500 or HiSeq X sequencer, generating 100 - 150 bp reads with a minimum coverage of 15X for ~95% of the genome (mean coverage of 35X). Read length chemistries changed from 100 bp to 125 bp and finally to 150bp which was taken into account in all analyses. Sequencing reads were pre-processed by Illumina with Isaac Aligner and Variant Caller (v2, Illumina Inc.) using human genome assembly GRCh37 as reference. Variants were normalised and merged into a multi-sample variant call format (VCF) file using the gvcf aggregation tool ‘agg’.

2.4 RNA sequencing in the PAH Cohort study

2.4.1 RNA sequencing

The information stored in genes in the form of DNA is transcribed into RNA which is then translated into protein as stated by the central dogma of molecular biology (Crick, 1970).The transcriptome of a cell or tissue is defined as the collection of RNA molecules that regulates its biological activity. The development of NGS technologies led to the improvement of gene expression quantification methods which historically relied on hybridization-based or Sanger sequencing techniques (Kukurba and Montgomery, 2015). RNA sequencing (RNASeq) starts with the isolation of RNA followed by its fragmentation to the desired length and its conversion by reverse transcriptase to complementary DNA (cDNA). The quality of the isolated RNA is usually assessed using an Agilent Bioanalyzer and

given an RNA Integrity Number - or RIN for short – that ranges from 1 to 10 (most to least RNA degradation). NGS library construction then involves attaching oligonucleotide adapters to the fragments and quantifying the resulting product. Since ribosomal RNA (rRNA) is the most abundant form of RNA in a sample, targeting polyadenylated messenger RNA (mRNA) molecules before library preparation is crucial for obtaining the desired depth of sequence coverage (Head et al., 2014, Kukurba and Montgomery, 2015).

2.4.2 PAH Cohort study

Patients with idiopathic, heritable or drug-induced PAH were recruited from expert centres across the UK as part of the PAH Cohort study (www.ipahcohort.com). 359 PAH patients - of which 276 patients also had their genomes sequenced as part of the WGS effort in the NIHRBR study - from the PAH Cohort study had their transcriptomes profiled through RNAseq.

Whole blood (3 ml) was collected in Tempus™ Blood RNA Tubes, which were stored at -80 oC until required. RNA was extracted using a Maxwell® 16 Tissue LEV Total RNA Purification Kit (Promega, Madison, WI, USA). All samples submitted for sequencing had a 260/230 ratio >1.5, a 260/280 ratio between 1.9-2.1 and an RNA integrity number (RIN)>7 as determined by Bioanalyser (Agilent, Santa Clara, CA, USA). Globin-Zero Gold rRNA Removal Kits (Illumina Inc, San Diego, CA) were used to remove ribosomal RNA contamination from whole blood RNA samples. 75 bp paired-end sequencing on Illumina Hiseq4000 was performed on pooled multiplexed libraries of ~80 samples on 9 sequencing lanes.

Fastq files (raw reads from RNAseq) were analysed using Salmon v0.9.1 (Patro et al., 2017) and GENCODE release 28 to produce transcript abundance estimates which were converted to gene expression data using tximport in R (Soneson et al., 2015). Salmon, the first transcriptome-wide quantifier to correct for fragment GC-content bias was used, which substantially improves the accuracy of abundance estimates and the sensitivity of subsequent differential expression analysis (Patro et al., 2017).

2.5 Expression quantitative trait loci mapping in PAH

'Omics' in gen-omics, transcript-omics, prote-omics, metabol-omics, epigen-omics and microbi-omics refers to the global assessment of a set of molecules present at a given time in an organism. Integrating genomic data (genome-wide genetic data) with global gene expression profiles (transcriptomic data) enabled the characterization of genetic control over gene expression which can help bridging the gap between disease susceptibility loci discovered in genome-wide association studies (GWAS) and biology (Zhernakova et al., 2017, Joehanes et al., 2017, Westra et al., 2013). Genetic regions that influence gene expression are commonly referred to as expression quantitative trait loci (eQTL). Apart from gene expression, quantitative trait loci have been described for several biological stages or intermediate phenotypes such as proteins (pQTL) (Battle et al., 2015), metabolites (mQTL) (Nath et al., 2017), gut microbes (mbQTL) (Bonder et al., 2016) ,methylation (meQTL) (Hannon et al., 2017), histone modification (hQTL) (McVicker et al., 2013) and more, creating layers of information in the cascade of events between disease-associated genetic loci and complex phenotypes.

For the mapping of eQTL in PAH patients, genetic variants with a minimum minor allele frequency of 5% were extracted from the NIHRBR variant call format (VCF) files and further filtered for variants called in at least 95% of the samples and had a p-value greater than 10^{-5} for Hardy Weinberg equilibrium. Various software tools were used in the eQTL mapping pipeline including PLINKv1.90 (Chang et al., 2015) for data extraction and computing multidimensional scaling components, QCTOOLv2.0 (Band and Marchini) for variant filtering, QUICKTESTv1.1 (Johnson et al., 2009) for association testing, and ANNOVAR (Wang et al., 2010) for variant annotation.

2.6 Mendelian Randomisation: one-sample design and assumptions.

An introduction to the concept of Mendelian Randomisation (MR) and an example for its application are given in the *Introduction* (0). Since the first application of a genetic instrument to assess the effectiveness of allogenic sibling bone marrow transplantation in acute myeloid leukaemia published in 1991 (Gray and Wheatley, 1991), ongoing development of MR methodology has led to more sophisticated techniques allowing the use of summary-level data both boosting statistical power and circumventing the need for sharing individual-level information (Bowden et al., 2015, Bowden et al., 2016, Burgess et al., 2017a, Burgess et al., 2020). Initial MR studies used a single genetic instrument in one study sample with data available for the genetic polymorphism, the exposure and the outcome of interest. The causal effect in the one-sample setting can be estimated with two-stage least squares (2SLS) regression (Burgess et al., 2017b) as shown in Figure 2.1 below, using the causal relationship between homocysteine and coronary heart disease (CHD) as an example.

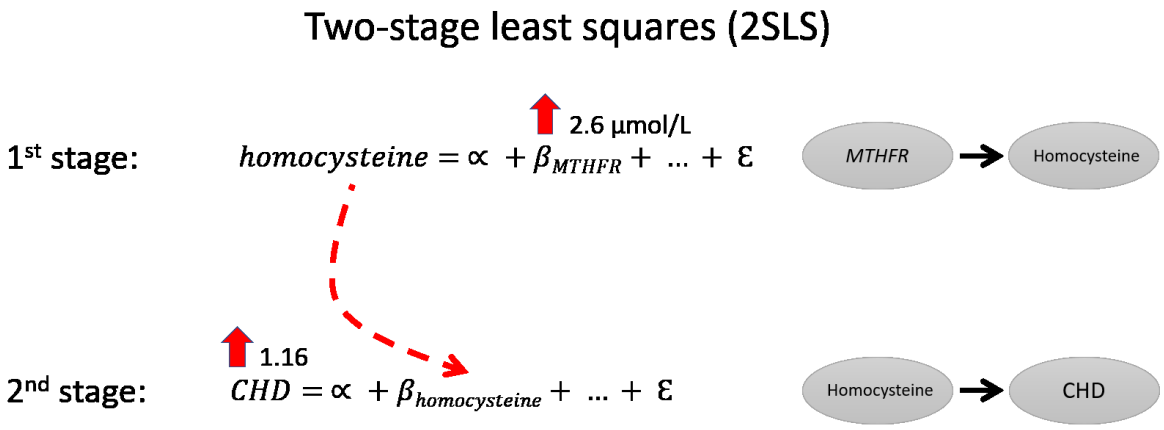


Figure 2.1 Two-stage least squares regression for estimating causal effect in the one-sample setting.
Definition of abbreviations: MTHFR = methylene tetrahydrofolate reductase; CHD = coronary heart disease.

In the first stage of the 2SLS regression (Figure 2.1), homocysteine is regressed on the *MTHFR* 677 genotype and – given that the effect of the genetic instrument is significant - fitted homocysteine values from the model are taken to the next stage of the regression. In the second stage, CHD is regressed on the fitted homocysteine values which estimates the effect of the genetically determined amount of homocysteine increase (in the above example 2.6 µmol/L) on CHD.

However, the MR causal effect is an unbiased estimate of the real causal effect only if some assumptions hold true. These assumptions state that the genetic instrument is associated with the exposure (Figure 2.2/ 1.), the genetic instrument is independent of the outcome given the exposure (Figure 2.2/2.), and the genetic instrument is independent of all confounders, known and unknown,

that confound the exposure-outcome association (Figure 2.2/3.). In practice, only the first assumption can be tested, although known and measured confounders can be checked for their independence of the genetic instrument as an exploration of assumption three.

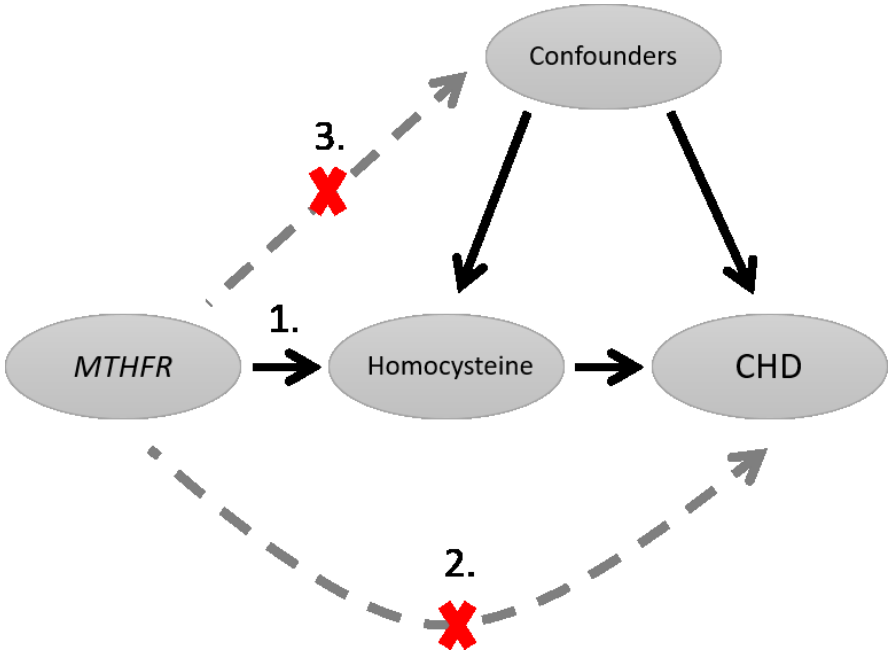


Figure 2.2 Assumptions of Mendelian Randomization.
Definition of abbreviations: MTHFR = methylene tetrahydrofolate reductase; CHD = coronary heart disease.

An example for the violation of MR assumptions is the presence of a genetic variant in linkage disequilibrium (LD) – non-random association of alleles at two or more loci - with the genetic instrument which is associated with any of the confounders of the exposure-outcome relationship or affects the outcome via any other pathway (VanderWeele et al., 2014). Examples of genetic loci with multiple correlated alleles that affect disease risk - after conditioning on all other variants in the locus - are emerging as genetic association studies are becoming more powered to detect ever smaller effects (Evangelou et al., 2018, Mahajan et al., 2018, Eyre et al., 2012). Controlling for the potentially confounding genetic marker in LD with the instrument can solve the problem where there is knowledge of such an association. However, given the relatively short distances over which LD is seen in the human genome, the chances of a genetic variant influencing disease risk being associated with another variant influencing the disease are much less concerning than confounding present in observational studies (Davey Smith and Ebrahim, 2003).

Genetic instruments prior to the emergence of large-scale genome-wide association studies (GWAS) and easily available summary-level data (Buniello et al., 2019) were mostly based on single well-defined genetic loci often with functional evidence linking it to the exposure of interest, similar to the example of *MTHFR* and homocysteine. As the genetic susceptibilities of ever more phenotypes

are characterized through genotyping or sequencing efforts of disease-focused consortia (Harst and Verweij, 2018, Scott et al., 2012) and population-based studies (Bycroft et al., 2018), it is now feasible to investigate causal relationships between a large number of traits which presents new opportunities but also new pitfalls to consider.

2.7 Mendelian Randomisation: two-sample design.

Summary-level data of genetic effects from large consortia and population-based studies refer to the association estimates (beta coefficients and their standard errors) of genetic variants and phenotypes, often on the genome-wide scale. These summary estimates are often made publicly available by the consortium itself and on curated websites of genetic associations such as the NHGRI-EBI GWAS Catalog of published genome-wide association studies (Buniello et al., 2019) or the NCBI molecular QTL browser.

Two-sample Mendelian Randomisation (MR) requires summary-level data for the genetic variants used as instruments for the exposure (usually variants that reached genome-wide or a study-specific significance threshold) on both the exposure and the outcome. The population in which the exposure (genetic) effects are estimated and the population in which the outcome (genetic) effects are estimated should be non-overlapping to avoid capturing the confounded association between the exposure and the outcome (Burgess et al., 2015).

By using summary-level data, much larger sample sizes can be achieved through the meta-analysis of genetic effect estimates which, together with utilizing multiple instrumental variants, increases the power of MR studies. MR and instrumental variable analyses in general are low-power analyses due to the instrument explaining only a proportion of the exposure`s variability. The power of an MR analysis depends on the strength of the instrument i.e. the proportion of variation the genetic instrument explains in the exposure, the size of the underlying causal effect and the sample size available for the outcome (Burgess, 2014, Pierce et al., 2011). If the genetic instrument is weak, meaning it explains a small proportion of the variation in the exposure for the available sample size (Pierce et al., 2011), the MR causal estimate is biased towards the confounded observational exposure-outcome association. By combining multiple genetic variants in the instrument, the proportion of variation explained (R^2) increases.

Apart from boosting statistical power, using multiple genetic variants as instruments can shed light on possible violations of the assumptions of MR, as the chances of multiple independent variants all

being confounded (for example by nearby variants in LD) is increasingly smaller as the number of instrumental variants increases.

Causal effect estimation in the two-sample MR setting with summary-level data is illustrated through the example of low-density lipoprotein (LDL) and coronary heart disease (CHD) (Figure 2.3). Since LDL is a more complex phenotype than levels of a specific protein or the transcript abundance of a gene, several genetic regions are associated with altered levels of LDL acting via distinct pathways and processes (Liu et al., 2017).

All variants used as instruments provide a causal estimate for the relationship of LDL and CHD, also called the ratio estimate or Wald ratio given by simply dividing the variant`s effect on the outcome (CHD) by the variant`s effect on the exposure (LDL) as shown by the forest plot in Figure 2.3/b. Ratio estimates represent the change in the outcome per unit change in the exposure. These individual causal estimates are meta-analyzed to get an overall causal estimate whilst giving more weight to instrumental variants with precisely estimated effects on the outcome, by using the reciprocal/inverse of the variance [S_{Yk}^{-2}] (Burgess et al., 2015).

The above described meta-analysis of ratio estimates is equivalent to regressing the variant-exposure associations on the variant-outcome associations (Figure 2.3/a) with the same inverse variance weighting, whilst forcing the regression line to pass through the origin. The method is often referred to as the 'inverse-variance weighted' (IVW) method (Burgess et al., 2015) conventionally used in two-sample MR analyses. The causal estimate from the IVW method (β_{IVW}) is:

$$\beta_{IVW} = \frac{\sum_{k=1}^{K} X_k Y_k s_{Yk}^{-2}}{\sum_{k=1}^{K} X_k^2 s_{Yk}^{-2}}$$

where:
k is an index for each instrumental variant
X is the effect estimate on the exposure
Y is the effect estimate on the outcome
S is the standard error.

The standard error of the causal estimate is:

$$se(\beta_{IVW}) = \sqrt{\frac{1}{\sum_{k=1}^{K} X_k^2 s_{Yk}^{-2}}}$$

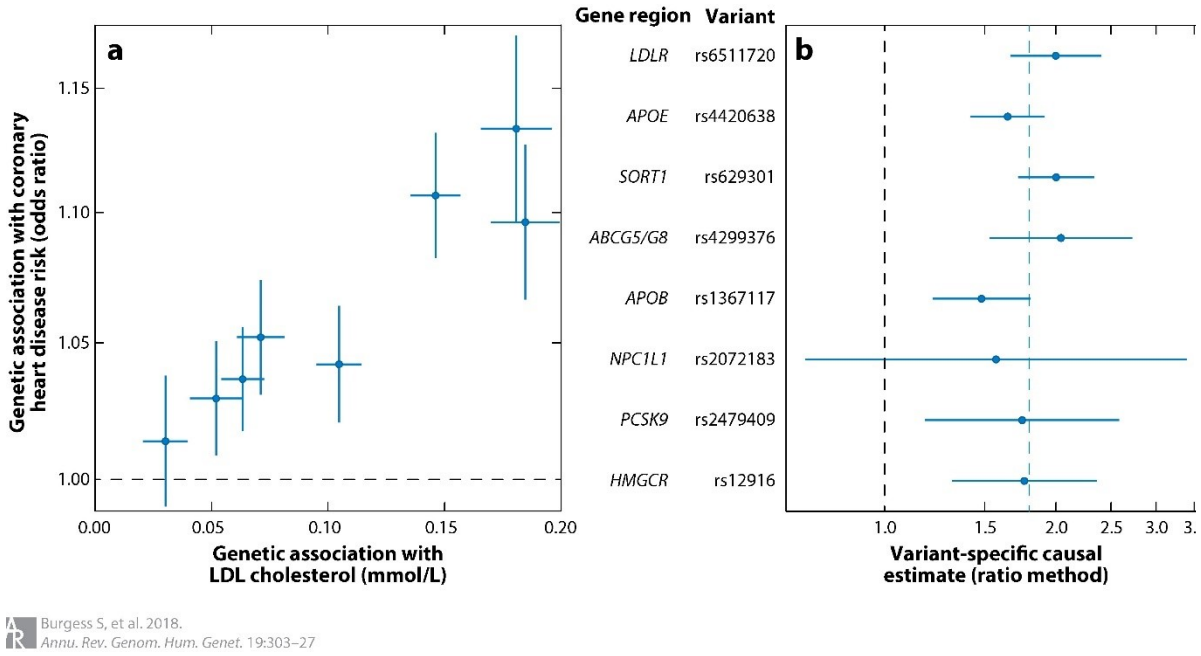


Figure 2.3 Mendelian Randomisation using summary-level data: low-density lipoprotein and coronary heart disease.
a: *Instrumental variant associations with low-density lipoprotein (LDL) levels and coronary heart disease (CHD) risk for eight genetic variants. Blue lines are 95% confidence intervals for the genetic associations.* ***b:*** *Individual causal estimates (odds ratio for CHD per 1-mmol/L increase in LDL) from the ratio method. The solid blue lines are 95% confidence intervals for the causal estimates; the dashed blue line is the (meta-analyzed) inverse-variance weighted estimate. Reprinted from 'Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data' by Stephen Burgess et al., 2018, Annual Review of Genomics and Human Genetics, Vol. 19:303-327.*

LDL-variant and CHD-variant associations plotted against each other in Figure 2.3/a show a dose-response relationship with variants that cause larger increase in LDL associated with higher CHD risk. When individual causal estimates from several genetic variants are mostly similar but some causal estimates differ sharply, this could be indicative of pleiotropic effects at those variants. In pleiotropy, a single genetic locus influences multiple phenotypes (Stearns, 2010). There are two types of pleiotropy: the desirable 'vertical' or 'mediated' pleiotropy which refers to a cascade of events (as in the example of LDL and CHD where variants affecting blood lipid levels also associate with CHD) and the problematic 'horizontal' or 'biological' pleiotropy where the genetic locus affects multiple phenotypes which are not all part of the same cascade of events. However, judgement on the validity of any genetic instrument should not be made based on its causal estimate being different from those of the other instrumental variants, unless there is biological knowledge justifying the exclusion of such variants.

Alternative methods for causal effect estimation using summary-level data and multiple – often hundreds of – variants are constantly being developed. Methods such as the weighted-median or MR-Egger regression methods relax certain assumptions and are more robust to the inclusion of

invalid instruments (Bowden et al., 2015, Bowden et al., 2016), although they make their own assumptions. The use of likelihood-based methods is also possible and these have been described elsewhere (Burgess et al., 2013, Burgess et al., 2017b).

In the MR analyses in this thesis, IVW and weighted median (WM) causal effects were calculated. The WM estimator will provide a consistent causal estimate when up to (but not including 50%) of the variants in the genetic instrument are invalid instruments, if no single variant contributes more than 50% of the weight (Bowden et al., 2016). Analogously to the IVW method (described above), a Wald ratio is calculated for each variant. These Wald ratios are then ordered and weighted by the same weights used in the IVW method (see above). Let wj be the weight of the jth ordered Wald ratio estimate.

$$s_j = \sum_{k=1}^{j} w_k$$

where:
k is an index for each of the variants used in the two-sample MR analysis
w is the weight of the variant
s_j is the sum of weights up to and including the jth Wald ratio estimate

The weights are standardized, so that the sum of weights is 1. The WM estimator is the median of the empirical distribution of weighted Wald ratios. Each Wald ratio is the $100(s_j - \frac{w_j}{2})$th percentile of this distribution.

Assessing the causal relationship using more than one method and applying sensitivity analyses for detecting horizontal pleiotropy (Bowden et al., 2015) and outliers (Verbanck et al., 2018) can increase confidence in the results of the MR analysis.

2.8 Genetic effects for red cell distribution width and pulmonary arterial hypertension

Effect estimates for genetic variants used as instruments in the Mendelian Randomisation analyses were derived from the summary statistics of the largest genome-wide association studies (GWAS) to date. Summary statistics comprise beta coefficients and standard errors taken from regression of the phenotype of interest on each genetic variant usually adjusted for sex and fine-scale population structure (Burgess et al., 2017a).

Genetic instruments serving as a proxy for red cell distribution width (RDW) were obtained from the largest-to-date GWAS on hematological traits in 173,480 individuals from a population-based study involving the UK Biobank and INTERVAL cohorts (Astle et al., 2016). RDW-variant associations were estimated in a linear mixed model - which accounts for the genetic component of phenotypic correlation generated by relatedness - assuming an additive genetic model. Additional adjustments were made for recruitment center and the first ten principal components for large-scale population structure.

The effects of RDW genetic instruments on PAH were derived from a recent GWAS of PAH susceptibility published by our group and others investigating the effects of common genetic variation on PAH risk involving a total of 11,744 individuals of which 2,085 were PAH cases (Rhodes et al., 2018). Effect estimates of the PAH GWAS were combined through the meta-analysis of four independent studies, one of which is the NIHRBR study with 847 PAH cases and 5,048 rare disease controls. The other major contributing study, the US PAH Biobank (PAHB) used a control population with mixed common diseases recruited from outpatient clinics (694 PAH cases and 1,560 controls), whilst the two smaller studies used population-based controls, with 269/275 PAH cases and 1,068/1,983 controls in the PHAAR and BHFPAH studies, respectively. The association of PAH diagnosis with each available variant was tested in a logistic regression framework assuming a log-additive genetic model adjusted for sex, read length chemistry (NIHRBR only), and population structure using the first four (NIHRBR and PHAAR), three (PAHB), or ten (BHFPAH) principal components.

2.9 Genetic risk score calculation in the VUMC study

2.9.1 The VUMC study

The hospital population-based Vanderbilt University Medical Centre (VUMC) study provided longitudinal clinical laboratory red cell distribution width (RDW) measurements, detailed clinical diagnoses and genome-wide genotype array data (genotyping platform: Illumina MEGAex) for 181 idiopathic PAH patients and 27,995 common disease controls for the RDW and PAH observational and Mendelian Randomisation analyses. Individuals with conditions including polycystic kidney disease, chronic kidney disease, liver disease and transfusion therapy, which likely influence RDW levels were excluded. Children and adolescents under the age of 18 as well as individuals with extreme RDW values (below 10%, N=20 or above 30%, N=15) were also excluded. In total, 118 PAH cases and 15,889 common disease controls from the VUMC were included in the analyses.

Individuals in the VUMC were genotyped in 6 batches (30,886 in total) using the Infinium Expanded Multi-Ethnic Global Array-8 (MEGA-ex) array (Illumina, San Diego, California, US). Unfiltered genotype data was received in binary plink files. Genotyped genetic variants were quality-checked prior to being submitted for imputation to the Haplotype Reference Consortium panel using the free Sanger Imputation Service (the Haplotype Reference et al., 2016). Variants were excluded if they had a low call rate (< 95%), deviated from the Hardy-Weinberg equilibrium (P ≤ 10-5), were rare (minor allele frequency ≤ 1%) or had more than two alleles. Sample quality control excluded individuals with high proportions (> 5%) of missing genotype data, unresolved sex discrepancies (discordant phenotype-genotype sex information), heterozygosity outliers, self-reported and/or principal component-based ethnic outliers, intentional duplicates and related individuals (PI-HAT > 0.2). Imputed variants were excluded if they showed a deviation from the Hardy-Weinberg equilibrium (P ≤ 10-5), were rare (minor allele frequency ≤ 1%, which are often poorly imputed) or were otherwise of low-quality with an INFO score lower than 0.9.

2.9.2 Genetic risk score calculation

Genetic risk scores (GRS) reflect the propensity to a trait or disease for an individual. GRS-based risk prediction has become more powerful in recent years by the improved accuracy of genetic variant effect estimates achieved through genotyping and deep phenotyping of large cohorts such as the UK Biobank (Bycroft et al., 2018). A study by Khera and colleagues (Khera et al., 2018) showed that GRSs comprising millions of common genetic variants from large genome-wide association studies for coronary artery disease (CAD), type 2 diabetes and other conditions are able to identify groups at high risk of developing disease with an accuracy comparable to that of well-established rare genetic

mutations which often confer several-fold disease risk to carriers. For example, the genome-wide GRS for CAD identified 20 times as many individuals as rare familial hypercholesterolemia mutations in the population of the UK Biobank who were at a CAD risk at least as high as rare mutation carriers (Khera et al., 2018), demonstrating an enormous gain in the fraction of the high-risk population identified who may be targeted early for disease prevention. Apart from individual-level risk prediction, GRSs can also be used for assessing shared disease aetiology and as proxy measurements for unmeasured phenotypes or for identifying target populations for clinical studies (Choi et al., 2018).

To estimate the predictive performance of the genetic variants used as instruments for red cell distribution width (RDW) in the Mendelian Randomisation analyses, weighted genetic risk scores consisting of the 179 RDW quantitative trait loci from the genome-wide association study of Astle et al. (Astle et al., 2016) for RDW (RDW GRS) were constructed for common disease controls in VUMC. GRS for an individual is the summation of the effect (RDW-increasing) alleles (0, 1 or 2) weighted by the effect size of the variant taken from the summary statistics of Astle et al. An additive model was used, meaning that homozygotes for the effect allele had twice the increase in RDW levels than heterozygotes, which was in line with the association model used by Astle et al. when estimating per-allele effects. The GRSs were calculated using the software PRSice-2 (Euesden et al., 2015).

The formula for the GRS is as below:

$$GRS_{i=} \sum_{j=1}^{m} wj\ xij$$

where:
m is the total number of variants
xij is the number of effect alleles at the jth variant of the ith individual
wj is the effect size of the variant.

2.10 Clinical studies of intravenous iron in PAH

Two randomised, double-blind studies of intravenous iron versus placebo were conducted in idiopathic and heritable PAH to assess the safety and clinical benefit of iron supplementation in iron deficient patients. The trials took place across four specialist centers in the United Kingdom and one center in Germany(Hammersmith Hospital, London; Papworth Hospital, Cambridge; Royal Hallamshire Hospital, Sheffield; and Giessen in Germany) where 42 patients were recruited in total and one center in Beijing, China where 17 patients were enrolled.

Patients were randomized in a 1:1 ratio in this 24-week crossover study. Iron was administered intravenously as a single 1000 mg (15 mg/kg if patient weight < 66.7 kg) infusion of Ferinject (ferric carboxymaltose, Europe) or 20mg/kg CosmoFer (iron dextran, China). Participants were iron deficient at recruitment and met either one of four stringent biomarker cut-offs: soluble transferrin receptors (sTfR) > 28.1 nmol/L where available, or ferritin < 37 µg/L or transferrin saturation < 16.4% or serum iron < 10.3 µmol/L. These cut-offs were based on previous work (Rhodes et al., 2011a) demonstrating that 83% of IPAH patients with an sTfR > 28.1 nmol/L could be identified correctly using the above cut-offs for the other three, more widely available biomarkers. Blood biomarker levels (serum iron, ferritin, transferrin saturation, sTfR, hepcidin, interleukin 6) were measured by a central laboratory and used to define iron deficiency.

3 Iron deficiency and anaemia in pulmonary hypertension diagnostic groups: a cross-sectional study in the National Pulmonary Hypertension Service, Hammersmith Hospital

3.1 Introduction

Pulmonary hypertension (PH) is a haemodynamic and pathophysiological condition that can complicate cardiovascular, respiratory, infectious and immune-related diseases or emerge in a sporadic or heritable form. Table 1.1 in the main *Introduction* shows the classification of PH as devised by the World Health Organisation, dividing the condition into five main groups based on similarities in haemodynamics, clinical presentation and disease aetiology. Groups 1, 3 and 4 constitute pre-capillary forms of PH while group 2 is defined by post-capillary hemodynamics assessed at right heart catheterization using the pulmonary arterial wedge pressure (PAWP) cutoff of 15 mmHg (pre-capillary ≤ 15 mmHg, post-capillary > 15 mmHg). PAWP estimates pressure in the left atrium with high values suggesting left ventricular failure. Comparing the prevalence of different PH subgroups is not an easy task given the lack of comprehensive epidemiological data on certain subgroups, however PH due to left heart disease is without doubt the most common form of PH (Galie et al., 2016, Hoeper et al., 2016).

Group 1 PH, namely pulmonary arterial hypertension (PAH), has been widely studied with patient registry databases set up both in the Western world and in developing countries (Hoeper et al., 2013, Lau et al., 2017, Alves et al., 2015). These registries provide researchers with valuable information on the demographics, disease course, treatment effectiveness and safety and much more over a course of time for observational analyses. The prevalence of PAH is estimated to be between 15 and 60 subjects per million population in Europe (Galie et al., 2016). PAH can develop as a complication of connective tissue diseases (CTD), portal hypertension, congenital heart disease (CHD), infection with the human immunodeficiency virus or schistosomiasis and as a result of exposure to certain toxins. The epidemiology, mechanisms and treatment options of PAH are described in detail in the main Introduction.

Idiopathic PAH (IPAH) is the most common form of PAH in registries from the Western world where it constitutes 30% to 50% (Lau et al., 2017). CTD associated PAH (PAH CTD) is the second most prevalent form of PAH constituting 15% to 30% of registered patients in PAH registries from

developed countries (McGoon et al., 2013). Systemic sclerosis (also called scleroderma) – an autoimmune condition characterised by the buildup of scar tissue and abnormally high collagen production – leads to PAH in 7% to 12% of patients (Simonneau et al., 2013) accounting for 61% to 76% of PAH CTD cases in the developed world (McGoon et al., 2013). PAH CTD presents with the worst prognosis of all PAH subgroups with a 1-year mortality rate of 30% (Tyndall et al., 2010) compared to 9%-17% in IPAH (McGoon et al., 2013). PAH can develop as a complication of uncorrected congenital heart disease (CHD) especially in patients with left-to-right shunts (hole between the chambers of the heart). In PAH CHD, increased blood flow from the left side of the heart (systemic circulation) to the right side of the heart (pulmonary circulation) leads to increased pulmonary vascular resistance through vascular remodeling and dysfunction. In Eisenmenger`s syndrome, which is the most advanced form of PAH CHD, the direction of the shunt reverses or becomes bidirectional (Alto and Mahadevan, 2012). PAH CHD is much less prevalent in the developed world where it makes up 10% to 23% of patients in PAH registries (McGoon et al., 2013) than in the New Chinese Registry, for example, where PAH CHD constitutes 43% of all PAH subgroups (Lau et al., 2017). This phenomenon is a result of the differences in availability of health care and paediatric surgery to close heart defects.

Group 4 PH, namely chronic thromboembolic PH, is a form of pre-capillary PH that emerges as a result of pulmonary embolism, the obstruction of the pulmonary arteries by thrombus (blood clot) formation. The obstructed pulmonary arteries cause pulmonary vascular resistance to rise and PH to develop. The first line of treatment for CTEPH is pulmonary endarterectomy, the surgical removal of thrombi from the pulmonary arteries, however not all patients suffering from CTEPH qualify for surgery. The mechanisms behind pulmonary embolism leading to chronic PH is not fully understood, as chronic PH develops in some patients with acute pulmonary embolism even after resolving the emboli (Jamieson et al., 2003). The obstruction of the pulmonary arteries is not likely to be the sole driving factor behind CTEPH, instead a combination of factors initiated by pulmonary embolism lead to pulmonary vascular remodeling, a hallmark of PAH, and chronic PH (Hoeper Marius et al., 2006, Jamieson et al., 2003). The prevalence of CTEPH is estimated to be lower than that of PAH with 3.2 cases per million population reported by a Spanish cohort (Escribano-Subias et al., 2012).

Group 2 PH or PH due to left heart disease (PH LHD) is a post-capillary form of PH characterised by elevated pressure in the left atrium of the heart which can be estimated at right heart catheterization using PAWP as a surrogate measure. PAWP above 15 mmHg is used as a threshold for distinguishing pre-capillary PH from post-capillary PH. The pressure from the left atrium gets transmitted backwards via the pulmonary veins, pulmonary capillaries and eventually the pulmonary arteries. The backward transmission of pressure to the pulmonary circulation commonly leads to the

passive elevation of the pulmonary arterial pressure and in some cases to reactive pulmonary vasoconstriction leading to vascular remodeling as seen in PAH and irreversible pulmonary vascular disease (Oudiz, 2007). A multitude of left heart conditions can lead to increased left atrial pressure and therefore PH, including left ventricular dysfunction, aortic and mitral valve stenosis and valve regurgitation (Charalampopoulos et al., 2018). Although estimates of the prevalence of PH LHD taking all forms of heart disease associated with PH LHD together is lacking, PH prevalence in cohorts of single underlying causes have been estimated. Over 50% of patients with heart failure with preserved ejection fraction, a frequent condition (prevalence: 1% - 5%) in western countries (Lam et al., 2011), develop PH which is associated with worse outcomes and higher mortality in these patients (Leung et al., 2010). Most patients with severe mitral valve stenosis and up to 65% of patients with aortic stenosis develop PH (Galie et al., 2016). These estimates suggest that PH LHD is the most common form of PH (Hoeper et al., 2016). PH LHD patients are generally older than PAH patients and present with morbidities common in old age, such as obesity, diabetes, systemic hypertension, coronary artery disease and dyslipidemia (Charalampopoulos et al., 2018).

Iron deficiency with or without anaemia have been established to be common and associated with worse survival in idiopathic and heritable forms of PAH (Ruiter et al., 2011, Soon et al., 2011, Rhodes et al., 2011a, Krasuski et al., 2011). Intervention studies investigating the effects of iron supplementation in PAH (IPAH, PAH CHD, PAH CTD) patients with iron deficiency reported promising results (Table 1.3), including improvement in 6-minute walk distance (a prognostic marker of PH) and exercise capacity measures (Ruiter et al., 2015, Viethen et al., 2014). High prevalence (38%) of iron deficiency and its association with adverse outcomes have also been described in a cohort of Eisenmenger`s syndrome patients (severe PAH CHD) (Van De Bruaene et al., 2011). Lower iron status – assessed by ferritin levels only – in patients with chronic obstructive pulmonary disease, a condition that can lead to group 3 PH, was associated with higher pulmonary artery pressure and higher odds of developing PH (OR = 3.3, 95% CI = 1.3-8.6) (Plesner et al., 2017). Also, iron trials in healthy individuals exposed to hypoxia - at high altitudes or acute hypoxia in hypoxic chambers – report the attenuation of the pulmonary vasoconstrictor response in those receiving iron supplementation (Table 1.3) (Smith et al., 2008, Smith et al., 2009, Frise et al., 2016), and vice versa, exacerbation with iron chelation.

There has been growing interest and enthusiasm for the role of iron in PAH following the success of iron replacement therapy in chronic heart failure patients with iron deficiency with or without anaemia. These clinical trials unanimously report improved symptoms including functional class, exercise capacity measures, left ventricular function and more (Okonko et al., 2008, Toblli et al., 2007, Anker et al., 2009). To date, the prevalence and/or prognostic value of iron deficiency have

not been assessed in a diverse cohort of PH patients from a single centre including subgroups other than group 1 (PAH). The aim of this study was to assess the prevalence of iron deficiency and its possible underlying mechanisms, such as inflammation and hepcidin dysregulation, in PH patients referred to the National Pulmonary Hypertension Service at Hammersmith Hospital with heterogeneous etiologies in, to my knowledge, the largest and most diverse PH population studied to date.

3.2 Methods

3.2.1 Data

Patient data were retrieved from the electronic health records of the National Pulmonary Hypertension Service at Hammersmith Hospital, London. Plasma samples were obtained from the Imperial College Pulmonary Hypertension biorepository. This biorepository hosts blood samples from consenting patients referred to the National Pulmonary Hypertension Service (NPHS) at Hammersmith Hospital between 2005 and 2018. This cohort (total N=812) comprised of disease controls (No PH, N=173) who were referred based on a suspicion of PH to the NPHS but did not receive a diagnosis of PH, idiopathic and heritable pulmonary arterial hypertension (H/IPAH) patients (N=193), chronic thromboembolic pulmonary hypertension (CTEPH) patients (N=275), patients with pulmonary arterial hypertension associated with congenital heart disease (PAH CHD) (N=30), patients with pulmonary arterial hypertension associated with connective tissue disease (PAH CTD) (N=54) and patients (N=87) with PH due to left heart disease (PH LHD). IPAH patients included in a previous study published by Rhodes et al. were excluded from this analysis (Rhodes et al., 2011a). Soluble transferrin receptor (sTfR) and interleukin (IL)-6 levels were determined by enzyme-linked immunoadsorbent assay. Hepcidin levels were determined by a competitive radioimmunoassay as previously described (Busbridge et al., 2009). All other iron parameters including serum iron, ferritin, transferrin and hemoglobin (Hgb) were measured by standard clinical pathology accredited hospital assays.

World Health Organization (WHO) functional class (I-IV : increasing severity) and six-minute walk distance (6MWD) assessments within 3 months of iron parameter measurements were retrieved. Baseline haemodynamic measurements of pulmonary vascular resistance (PVR) and mean pulmonary arterial pressure (PAPm) were taken at the time of diagnosis at right heart catheterization. Self-declared ethnicity information was available for a subset of patients with three major groups identified; white (consisting of 478 British, 8 Irish, 81 any other white background),

South Asian (consisting of 35 Indian and 5 Pakistani) and black (consisting of 25 African, 24 Caribbean and 5 any other black background). Survival status was censored on March 28, 2019. Those who underwent surgery including lung or heart transplantation (H/IPAH: 3; PAH CHD: 1), pulmonary endarterectomy (CTEPH: 148) or atrial septostomy (H/IPAH: 2; PAH CHD: 1) were censored on the day of the surgery. Demographic characteristics and iron indices of patients are shown in Table 3.1. Clinical characteristics are shown in Supplementary Table 8.1.1

3.2.2 Statistical analyses

Iron deficiency was defined by the combination of low ferritin (< 100 µg/L) and low transferrin saturation (TSAT) (< 20%) or by high sTfR levels (> 28.1 nmol/L, upper limit of normal range in healthy volunteers of clinical lab assay) alone. Thresholds for ferritin and TSAT were more liberal than those used for the definition of iron deficiency in healthy individuals (ferritin < 30 µg/L and TSAT < 16%) to account for iron sequestration in chronic disease. The cohort was divided into four groups based on their iron status; those who are classified as iron deficient based on their sTfR levels only (referred to as 'sTfR-only'), those iron deficient based on their ferritin and TSAT levels only ('Ferr-TSAT-only'), those meeting both criteria ('Ferr-TSAT+sTfR') and finally those who meet neither and are therefore considered iron replete. To ensure accurate iron status assessment for all individuals, patients whose ferritin, TSAT, sTfR and Hgb levels were not measured within the timeframe of one month were not assigned to an iron status group. The sample size per diagnostic group with complete iron biomarker data was 752 (No PH = 161, H/IPAH = 179, PAH CTD = 51, PAH CHD = 32, PH LHD = 81, CTEPH = 248). Anaemia was assessed by haemoglobin levels with a threshold of 11.5 g/dL or lower for females and 13 g/dL or lower for males. Blood biomarkers are all presented in the regression models as z-scores, i.e. in units of standard deviation (σ) calculated as follows: $(x_i - \bar{x})/\sigma_{\bar{x}}$. Skewed biomarker levels including IL-6, sTfR, hepcidin and ferritin were log-transformed before being converted to standard units. Ethnicity was coded as white or non-white given the relatively small number of non-white patients. All models were adjusted for ethnicity and/or sensitivity analyses were carried out in each group separately. Cox regression analyses were performed to test the prognostic value of iron status.

	No PH		H/IPAH		PAH CTD		PAH CHD		PH LHD		CTEPH	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
N (%)	116 (72)	45 (28)	131 (73)	48 (27)	40 (78)	11 (22)	19 (59)	13 (41)	49 (60)	32 (40)	120 (48)	128 (52)
Age at diagnosis, mean (SD)	57.95 (16.72)	57.22 (14.55)	49.25 (18.19)	57.61 (18.02)	56.9 (14.92)	64.62 (13.23)	46.3 (15.88)	52.95 (20.25)	70.4 (11.95)	73.67 (8.35)	61.85 (16.51)	62.43 (13.83)
Age at sample, mean (SD)	58.04 (16.87)	57.36 (14.41)	52.12 (17.43)	59.65 (16.26)	58.96 (13.97)	65.47 (12.88)	49.99 (16.08)	55.28 (19.36)	70.91 (11.14)	73.64 (8.35)	64.27 (16.17)	64.63 (13.73)
WHO class, N (%)												
available	88 (76)	36 (80)	107 (82)	39 (81)	34 (85)	7 (64)	18 (95)	11 (85)	41 (84)	28 (88)	108 (90)	119 (93)
I	9 (10)	8 (22)	3 (3)	3 (8)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2)	1 (4)	5 (5)	10 (8)
II	24 (27)	8 (22)	19 (18)	7 (18)	11 (32)	0 (0)	3 (17)	4 (36)	5 (12)	3 (11)	17 (16)	37 (31)
III	54 (61)	20 (56)	75 (70)	26 (67)	19 (56)	4 (57)	12 (67)	6 (55)	32 (78)	24 (86)	73 (68)	66 (55)
IV	1 (1)	0 (0)	10 (9)	3 (8)	4 (12)	3 (43)	3 (17)	1 (9)	3 (7)	0 (0)	13 (12)	6 (5)
Ethnicity, N (%)												
available	69 (59)	28 (62)	110 (84)	34 (71)	32 (80)	10 (91)	14 (74)	10 (77)	32 (65)	21 (66)	105 (88)	118 (92)
white	50 (72)	23 (82)	95 (86)	31 (91)	25 (78)	9 (90)	12 (86)	9 (90)	29 (91)	18 (86)	93 (89)	107 (91)
South Asian	7 (10)	0 (0)	12 (11)	2 (6)	3 (9)	1 (10)	2 (14)	1 (10)	2 (6)	2 (10)	0 (0)	2 (2)
black	12 (17)	5 (18)	3 (3)	1 (3)	4 (12)	0 (0)	0 (0)	0 (0)	1 (3)	1 (5)	12 (11)	9 (8)
Iron µmol/L	13 (10 - 17)	16 (12 - 20)	12 (9 - 15)	13.5 (9.75 - 17)	10 (6.75 - 14)	11 (7 - 12)	13 (9.5 - 20)	12 (11 - 14)	11 (8 - 15)	11.5 (9 - 16.25)	12 (8.75 - 17)	14 (11 - 19)
sTfR nmol/L	28 (23 - 39.6)	28.6 (24.4 - 47.9)	27 (21.2 - 35.17)	28.8 (25.6 - 37.51)	33.85 (26.83 - 43.17)	41.6 (29.85 - 56.85)	28.4 (22.7 - 41.33)	25.9 (21.5 - 42.1)	29.3 (26.6 - 41.4)	27.75 (22.85 - 40.95)	30.65 (24.53 - 44.45)	28.2 (23.43 - 35.52)
Log sTfR/ferritin ratio	-0.72 (-1.31 - 0.09)	-1.42 (-2.03 - -0.52)	-0.72 (-1.38 - 0.11)	-1.16 (-1.83 - -0.48)	-0.59 (-1.39 - 0.28)	-0.54 (-1.97 - 0.49)	-0.39 (-1.41 - 0.26)	-1.1 (-1.93 - 0.06)	-1.03 (-1.53 – -0.02)	-1.35 (-1.88 – -0.37)	-0.57 (-1.16 - 0.26)	-1.17 (-1.79 - -0.44)
Transferrin *g/L*	2.8 (2.5 - 3.08)	2.6 (2.23 - 2.9)	2.7 (2.4 - 3.3)	2.6 (2.4 - 3.1)	2.7 (2.4 - 3.12)	2.9 (2.55 - 3.2)	2.8 (2.6 - 2.98)	2.5 (2.3 - 3.1)	2.9 (2.6 - 3.2)	2.8 (2.4 - 3.02)	2.8 (2.7 - 3.12)	2.55 (2.3 - 2.9)
TSAT *%*	18.5 (13.8 - 26)	26 (18 - 33)	19 (14 - 24)	22 (16 - 28)	17 (11.8 - 22.3)	15 (10.5 - 18.5)	21 (13.5 - 32)	18 (15 - 25)	18 (12 - 21)	17 (13 - 26)	18 (12.8 - 24)	25 (18 - 32)
Ferritin *µg/L*	54.5 (29.5 - 104.3)	128 (59 - 261)	58 (27 - 98.5)	103.5 (44.5 - 205.3)	57 (34.5 - 115.8)	58 (37.5 - 186)	34 (25.5 - 70.5)	89 (39 - 157)	79 (49 - 132)	114.5 (62 - 159.8)	53 (27 - 97.8)	92.5 (50 - 164.5)
Hepcidin *ng/mL*	16.2	17.7	13.97	21	6.55	1.8	6.5	9.56	14.05	18.95	5.15	7.85

	(6.2 - 31.5)	(6.8 - 36.5)	(5.7 - 30.8)	(9.9 - 39.1)	(1.4 - 13.9)	(1 - 9.2)	(1.6 - 11.7)	(4 - 23.7)	(7.4 - 25.9)	(4.6 - 29.5)	(1.1 - 10.9)	(2.8 - 16)
Haemoglobin *g/dL*	12.75 (11.8 - 13.5)	14.1 (12.2 - 14.7)	13.8 (12.6 - 14.8)	15.45 (13.4 - 16.9)	12.5 (11.6 - 14.4)	13.8 (12.5 - 14.5)	13.8 (13.3 - 15.35)	16 (13.4 - 17.7)	12 (11.4 - 13.2)	11.95 (10.4 - 13.9)	13.7 (12.4 - 14.5)	14.4 (13.6 - 16.2)
Haematocrit *%*	39 (36 - 41)	43 (36 - 44)	42 (39 - 45)	45 (41 - 49)	40 (36 - 43)	42 (40 - 46)	42 (40 - 47)	46 (42 - 52)	37 (36 - 40)	37 (34 - 41)	42 (38 - 44)	44 (41 - 48)
RDW *%*	14.25 (13.4 - 15.6)	14.5 (13.3 - 15.8)	14.4 (13.55 - 15.5)	14.6 (13.9 - 15.17)	15.35 (14.5 - 17.05)	17.6 (14.9 - 18.4)	14.7 (13.4 - 16.35)	15 (13.5 - 15.7)	14.7 (14.1 - 16.1)	15.35 (14.6 - 16.8)	14.9 (14.2 - 16)	14.4 (13.8 - 15.4)
IL-6 *pg/mL*	2.9 (1.6 - 5.3)	3.8 (1.4 - 5.9)	3.2 (1.7 - 5.9)	3.72 (2.2 - 6.6)	4.35 (2.1 - 7.6)	7.9 (6.3 - 15.8)	2.1 (1 - 4.1)	5.1 (2.7 - 10.4)	5 (3.8 - 7.7)	7.75 (4.5 - 13.9)	4 (2.6 - 6.7)	4.1 (2.7 - 7)

Table 3.1 Demographic characteristics and iron indices of patients referred to the National Pulmonary Hypertension Service at Hammersmith Hospital.
Values reported are median and interquartile range unless otherwise specified. Definition of abbreviations: No PH = disease controls; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; PAH CTD = pulmonary arterial hypertension associated with connective tissue disease; PAH CHD = pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension; sTfR = soluble transferrin receptor; TSAT = transferrin saturation; RDW = red cell distribution width; IL-6 = interleukin-6.

3.3 Results

All individuals were assigned to one of four groups based on iron status; i) iron deficient based on sTfR (> 28.1 nmol/L) levels (sTfR-only), ii) iron deficient based on ferritin (< 100 µg/L) and TSAT (< 20%) levels (Ferr-TSAT-only), iii) iron deficient based on sTfR, ferritin and TSAT (Ferr-TSAT+sTfR) levels and iv) iron replete. Demographic and clinical characteristics of the four iron status groups are summarized in Supplementary Table 8.1.2 for the whole cohort. Supplementary Table 8.1.3 and 8.1.4 describe the four iron status groups stratified by ethnicity.

3.3.1 Iron deficiency is common in all patient groups attending the PH clinic.

Iron deficiency, by either definition, was common (34-56% meeting Ferr-TSAT and 39-69% meeting sTfR threshols) in all patient groups including disease controls without PH. The proportion of iron replete individuals was between 20% and 33% across all diagnoses (Figure 3.1).

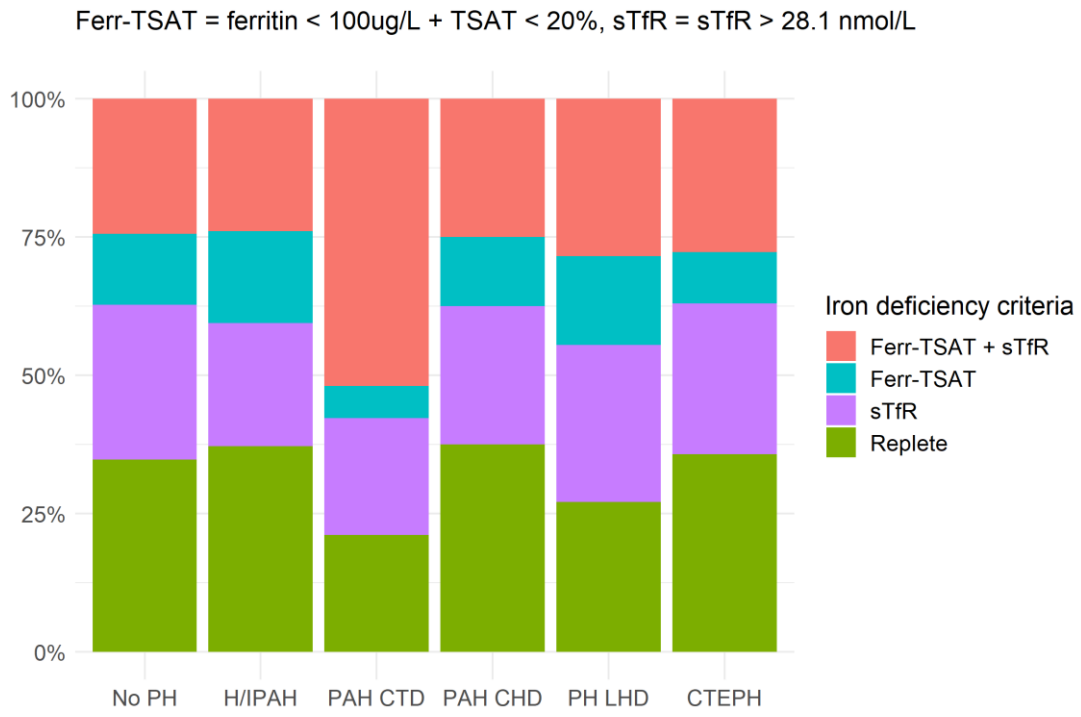


Figure 3.1 Proportions of iron status groups in disease controls and five PH diagnostic groups.
Definition of abbreviations: Ferr = ferritin; TSAT = transferrin saturation; sTfR = soluble transferrin receptors; No PH = disease controls; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; PAH CTD = pulmonary arterial hypertension associated with connective tissue disease; PAH CHD = pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension.

Since patients were not excluded based on whether or not they underwent surgery, a sensitivity check was conducted excluding those patients from the CTEPH diagnostic group who underwent pulmonary endarterectomy before their iron biomarkers were measured. Eigthy-five of the 248 CTEPH patients had undergone surgery before the date of their blood test used in these analyses. In the whole CTEPH cohort, 69 (28%) were iron deficient based on the Ferr-TSAT+sTfR criteria, whilst the proportion of Ferr-TSAT+sTfR patients (N = 43, 26%) in the smaller subcohort was not significantly different from that of the whole CTEPH cohort (Z = -0.41; p-value = 0.68).

The prevalence of Ferr-TSAT+sTfR iron deficiency in PAH CTD patients is much higher (50%) than that of the other diagnostic groups (23% - 29%, Figure 3.1). This association remained significant (OR_{CTD} = 2.61; 95% CI = 1.32 - 5.16; p-value = 0.006) after adjusting for age, sex, ethnicity, 6MWD as a measure of disease severity and IL-6 levels as a measure of inflammation by logistic regression using the whole of the cohort without PAH CTD as reference (Figure 3.2).

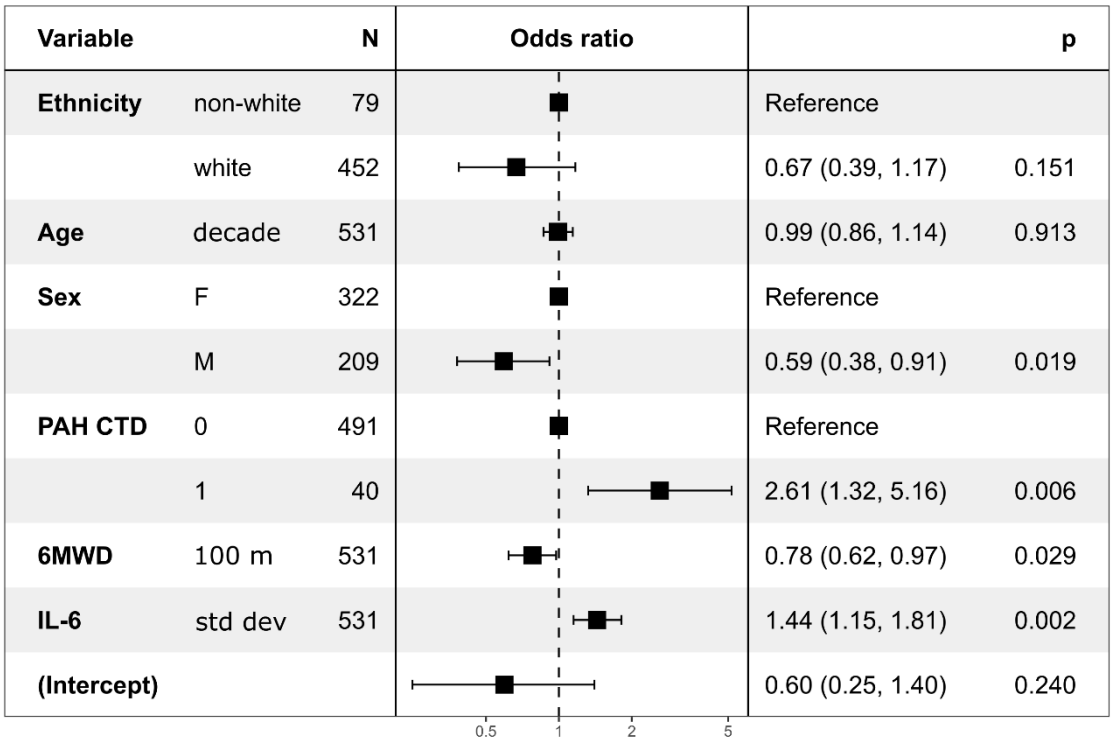


Figure 3.2 Predictors of `Ferr-TSAT+sTfR` iron deficiency.
Ferr-TSAT+sTfR iron deficiency is defined by the combination of three biomarkers: sTfR (> 28.1 nmol/L), ferritin (< 100 µg/L) and TSAT (< 20%). Definition of abbreviations: F = female; M = male; PAH CTD = pulmonary arterial hypertension associated with connective tissue disease; 6MWD = 6-minute walk distance; IL-6 = interleukin-6; std dev = standard deviation.

3.3.2 Haemoglobin levels are higher in H/IPAH, PAH CHD and CTEPH than in controls.

Anaemia was found in 23% of disease controls, 14% of H/IPAH, 28% of PAH CTD, 10% of PAH CHD, 44% of PH LHD and 14% of CTEPH patients (Figure 3.3). Ethnicity is a strong predictor of Hgb with non-white patients presenting with lower levels in most diagnostic groups ($Effect_{ethnicity}$ = -0.64; 95% CI = -0.85 – -0.43; p-value = $3.0x10^{-9}$, reference ethnicity: white) (Figure 3.4), so this was corrected for in subsequent analyses.

Haemoglobin (Hgb) levels of H/IPAH patients ($Effect_{H/IPAH}$ = 0.55; 95% CI = 0.32 – 0.77; p-value = $2.5x10^{-6}$, reference: disease controls), PAH CHD patients ($Effect_{CHD}$ = 1.35; 95% CI = 0.95 – 1.76; p-value = $1.24x10^{-10}$, reference: disease controls) and CTEPH patients ($Effect_{CTEPH}$ = 0.60; 95% CI = 0.39 – 0.81; p-value = $2.5x10^{-8}$, reference: disease controls) were higher than those of disease controls

after adjustments for age, sex and ethnicity (Figure 3.3). Even though the proportion of anaemic individuals in the PH LHD group is higher than that in the control group ($\chi^2_{(df=1)}$ = 10.6; 95% CI = -0.034 - -0.08; p-value = 0.001), their Hgb levels are not significantly different in the model adjusted for age, sex and ethnicity ($Effect_{LHD}$ = -0.23; 95% CI = -0.52 – 0.07; p-value = 0.13, reference: disease controls).

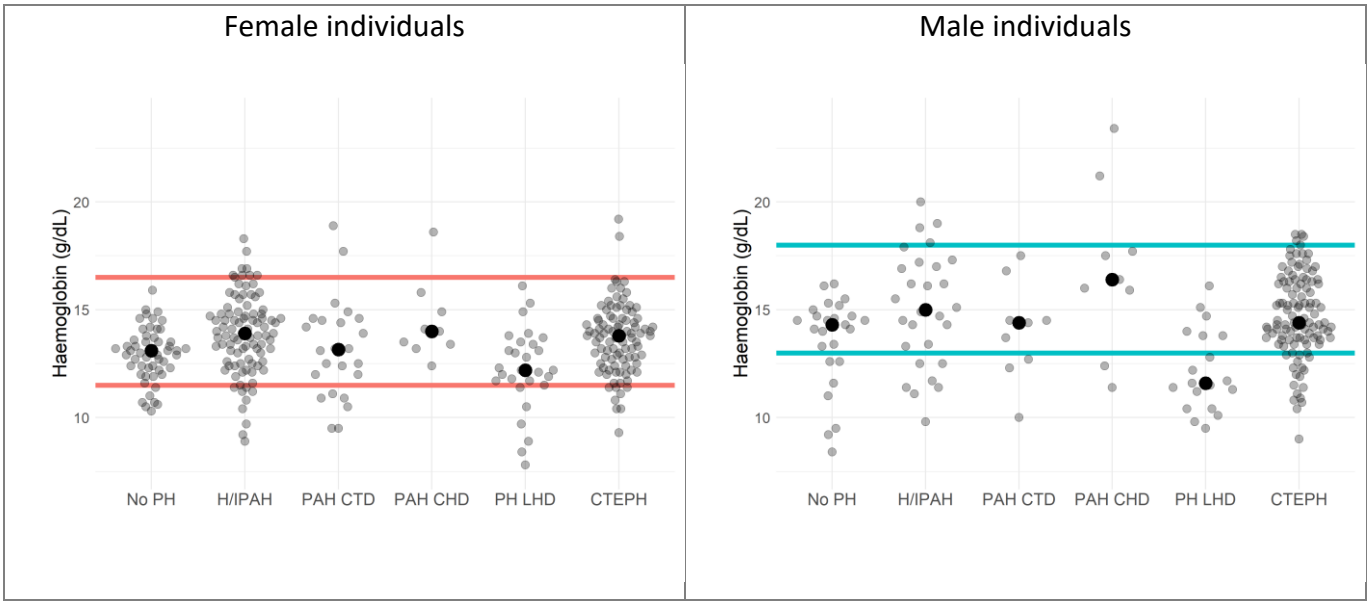


Figure 3.3 Haemoglobin levels in disease controls and five PH diagnostic groups in white patients. ***Left panel:*** *Black dots represent the median value of observations, red lines delineate the normal haemoglobin range for females (11.5 – 16.5 g/dL).* ***Right panel:*** *Black dots represent the median value of observations, blue lines delineate the normal haemoglobin range for males (13 – 18 g/dL). Definition of abbreviations: No PH = disease controls; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; PAH CTD = pulmonary arterial hypertension associated with connective tissue disease; PAH CHD = pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension.*

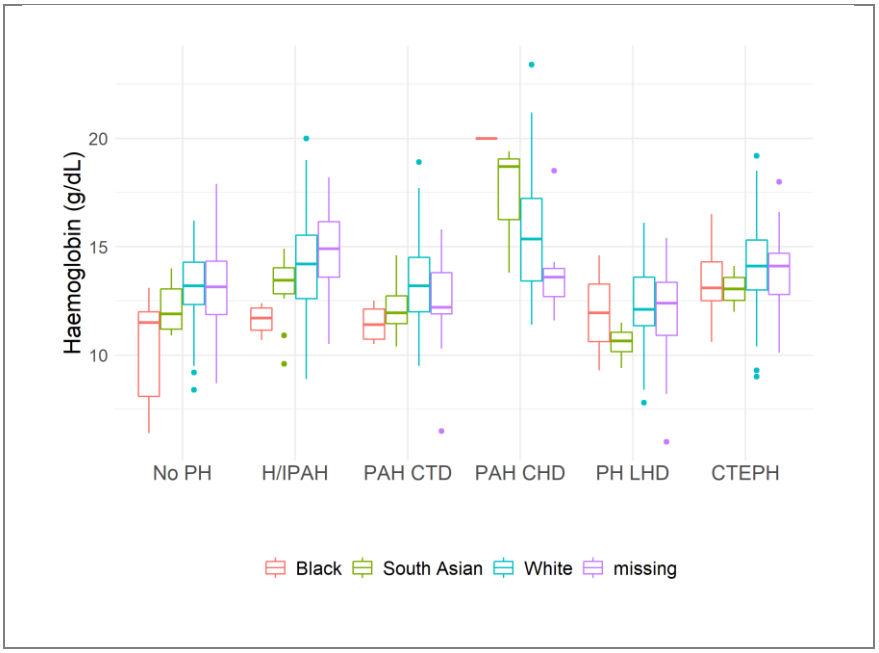


Figure 3.4 Haemoglobin levels by ethnicity in different diagnostic groups.
The bottom and the top lines of the box indicate the 25th and 75th percentiles, while the centre line indicates the median value. The whiskers extend to 1.5 times the interquartile range from both ends of the box with individual points being more extreme observations. Those in the `missing` category had no ethnicity information available.

3.3.3 The ratio of soluble transferrin receptors and ferritin is not associated with inflammation, while soluble transferrin receptor levels are in white individuals.

Soluble transferrin receptor (sTfR) has been the choice of iron status biomarker in PAH for its property of not being affected by inflammation (Rhodes et al., 2011a). In this study, the sTfR-only group was defined as iron deficient based on sTfR levels only (> 28.1 nmol/L) while the Ferr-TSAT+sTfR group had low ferritin (< 100 µg/L) and transferrin saturation (TSAT < 20%) levels as well high sTfR levels (> 28.1 nmol/L). To investigate what factors are associated with sTfR and how it varies with diagnosis, I modeled sTfR as a function of ferritin, Hgb, the inflammatory cytokine IL-6, ethnicity, diagnosis, age and sex.

Ethnicity was the strongest predictor of sTfR levels with Black or South Asian ethnicity associated with higher levels ($Effect_{ethnicity}$ = 0.71; 95% CI = 0.50 – 0.93; p-value = $1.66x10^{-10}$, reference ethnicity: white). Other significant predictors in the model included ferritin, Hgb and IL-6 ($Effect_{Ferr}$ = -0.23; 95% CI = -0.30 – -0.15; p-value = $4.27x10^{-9}$. $Effect_{Hgb}$ = -0.19; 95% CI = -0.28 – -0.11; p-value = $4.24x10^{-6}$. $Effect_{IL-6}$ = 0.23; 95% CI = 0.15 – 0.30; p-value = $3.1x10^{-9}$).

Given the large effect of ethnicity on sTfR levels, other predictors in the sTfR model were investigated in the white and non-white cohorts separately (Figure 3.5). These stratified models showed that whilst ferritin and IL-6 are important predictors of sTfR levels in white patients ($Effect_{Ferr}$ = -0.29; 95% CI = -0.37 – -0.22; p-value = $4.41x10^{-13}$. $Effect_{IL-6}$ = 0.26; 95% CI = 0.18 – 0.33; p-value = $3.91x10^{-11}$.), they do not predict sTfR in the non-white cohort ($Effect_{Ferr}$ = -0.09; 95% CI = -0.30 – 0.12; p-value = 0.38. $Effect_{IL-6}$ = 0.004; 95% CI = -0.23 – 0.24; p-value =0.97.). Furthermore, Hgb shows no effect on sTfR in the white cohort ($Effect_{Hgb}$ = -0.06; 95% CI = -0.14 – 0.02; p-value = 0.16) while being a strong predictor of sTfR in non-white patients ($Effect_{Hgb}$ = -0.61; 95% CI = -0.88 – -0.34; p-value = $2.04x10^{-5}$).

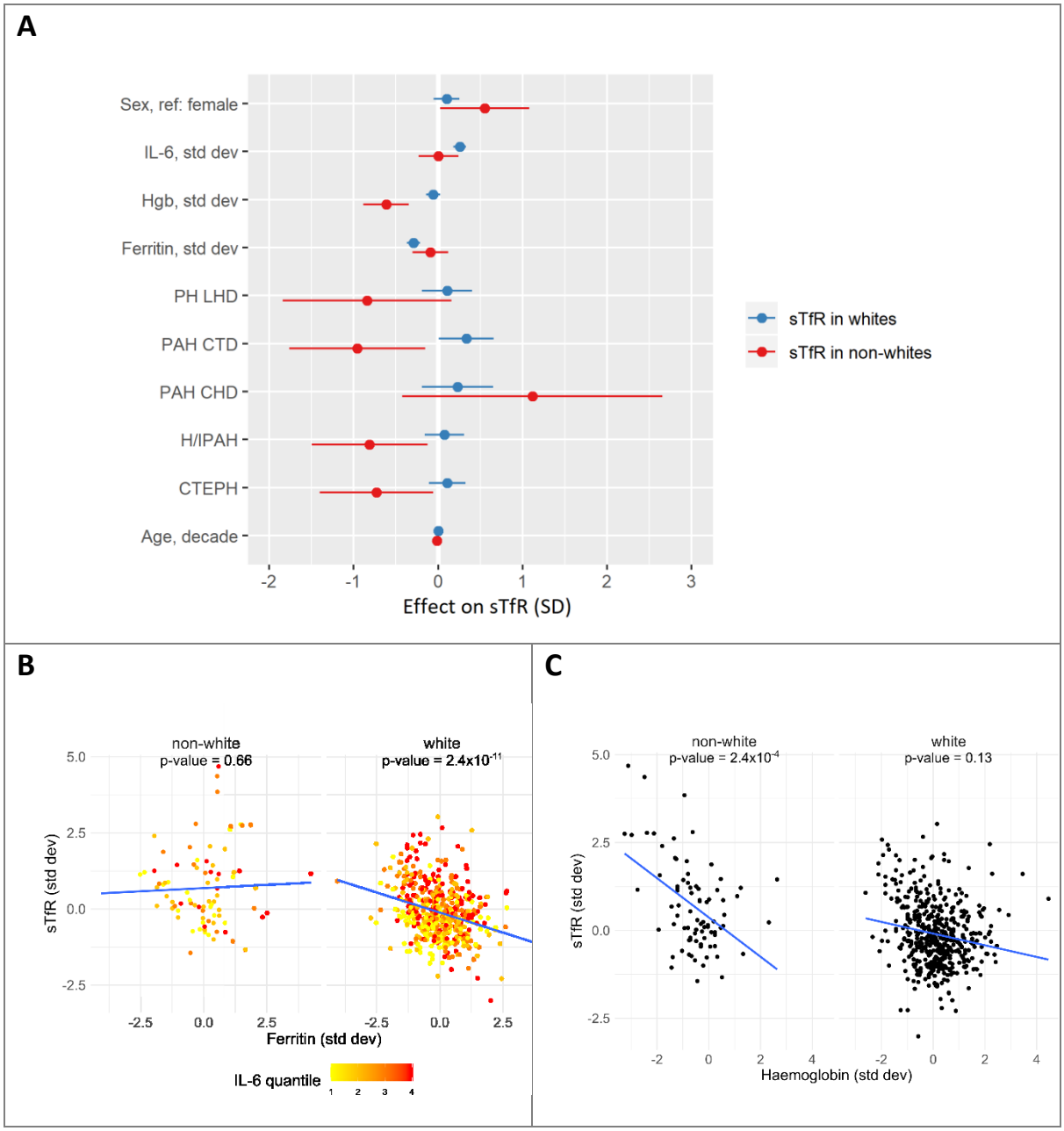


Figure 3.5 Predictors of soluble transferrin receptor levels in patients by ethnicity.
A: *Effect size is in sTfR standard units. Horizontal error bars represent 95% confidence intervals.* ***B:*** *Association of sTfR and ferritin, colored by IL-6 quantiles in white and non-white patients. P-values are given for the regression coefficient of ferritin from the full sTfR prediction models.* ***C:*** *Association of sTfR and Hgb in white and non-white patients. P-values are given for the regression coefficient of Hgb from the full sTfR prediction models. Definition of abbreviations: IL-6 = interleukin-6; Hgb = haemoglobin; sTfR = soluble transferrin receptors; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; PAH CTD = pulmonary arterial hypertension associated with connective tissue disease; PAH CHD = pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension.*

In white patients, diagnosis has little effect on sTfR levels with PAH CTD patients showing a trend for higher sTfR levels than disease controls ($Effect_{CTD}$ = 0.33; 95% CI = 0.01 – 0.66; p-value = 0.04) in the full model. sTfR levels in the non-white cohort were lower in three diagnostic groups (PAH CTD, H/IPAH, CTEPH) than those of the disease controls in the full model (Figure 3.5).

IL-6 is higher in both the sTfR-only and Ferr-TSAT+sTfR groups than in the iron replete and Ferr-TSAT groups in white patients ($Effect_{sTfR\text{-}only\ and\ Ferr\text{-}TSAT+sTfR}$ = 0.52; 95% CI = 0.35 – 0.68; p-value = $3.0x10^{-9}$, reference groups: replete and Ferr-TSAT) after adjustments for age, sex and diagnosis. IL-6 levels in iron replete and Ferr-TSAT-only patients are similar between white and non-white individuals, and IL-6 does not differ by iron status in non-white patients (Supplementary Table 8.1.3 and Supplementary Table 8.1.4). The prediction model of sTfR using age, sex, diagnosis, ferritin, Hgb and IL-6 levels as predictors explained 36% ($P < 1.8x10^{-6}$) of the total variability in sTfR levels of non-white patients and 20% ($P < 2.2x10^{-16}$) of the total variability in sTfR in white patients.

The ratio of sTfR and ferritin has been proposed as a measure of tissue iron availability with higher specificity than sTfR and higher sensitivity than ferritin alone, especially in individuals with chronic disease (Bultink et al., 2001, Suominen et al., 2000, Remacha et al., 1998, Means Jr et al., 1999). Log-transformed ratio of sTfR and ferritin (sTfR/ferritin) did not differ by diagnosis after adjustments for age, sex and IL-6 levels in either the white or non-white cohort. In white patients, sTfR/ferritin increases a small amount in the sTfR-only group ($Effect_{sTfR\text{-}only}$= 0.39; 95% CI = 0.22 – 0.57; p-value = $8.11x10^{-6}$, reference group: replete), almost twice as much in the Ferr-TSAT group ($Effect_{Ferr\text{-}TSAT}$= 0.71; 95% CI = 0.50 – 0.91; p-value = $8.98x10^{-11}$, reference group: replete) and increases considerably in the Ferr-TSAT+sTfR group ($Effect_{Ferr\text{-}TSAT+sTfR}$= 1.64; 95% CI = 1.47 – 1.81; p-value < $2x10^{-16}$, reference group: replete) compared to iron replete patients, independent of the effects of age, sex and IL-6 (Figure 3.6).

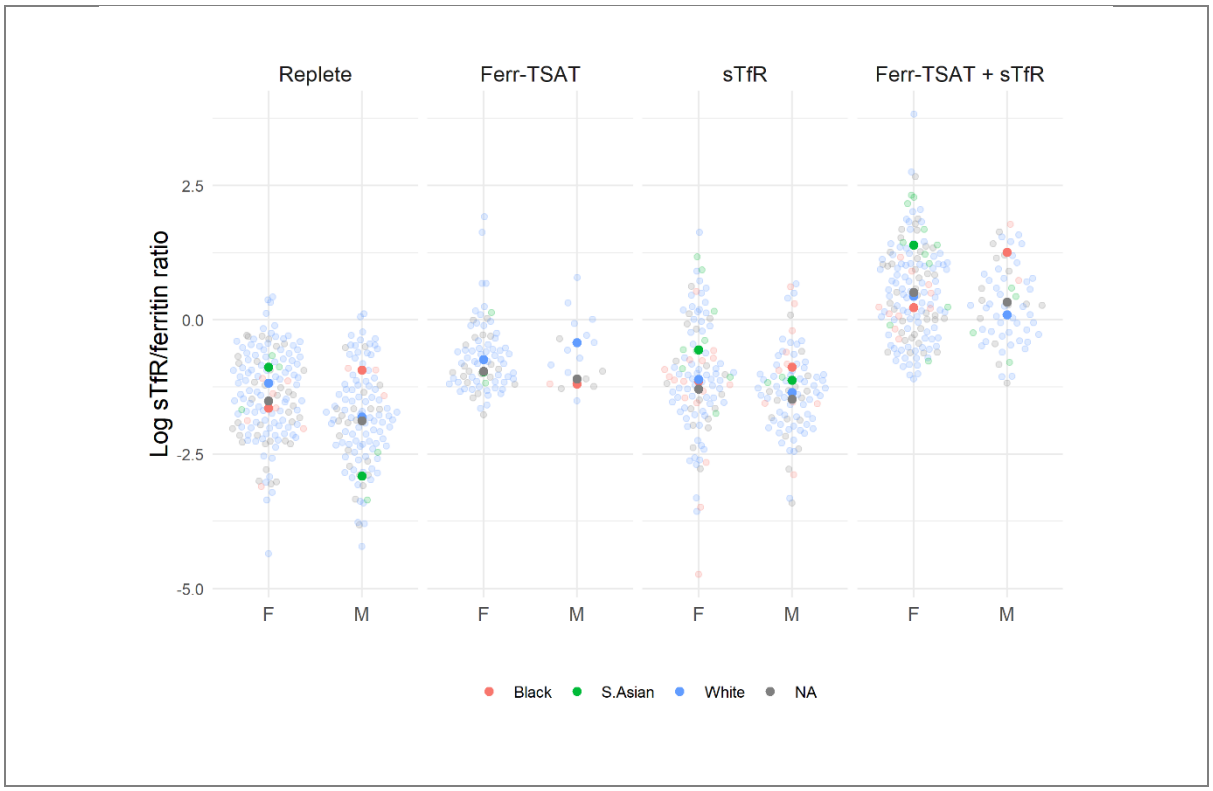


Figure 3.6 Ratio of soluble transferrin receptor and ferritin levels by iron deficiency group.
Each observation (dot) is colored by ethnicity. Dark red, green, and blue dots represent the median for black, South Asian and white groups, respectively. Observations in grey are of patients with missing ethnicity information. The Ferr-TSAT group meets both ferritin (< 100 µg/L) and TSAT (< 20%) cut-offs for iron deficiency, the sTfR group meets the sTfR threshold (> 28.1 nmol/L) only, while the Ferr-TSAT+sTfR group meets the cut-off for all three biomarkers of iron deficiency. Individuals in the replete group do not meet any of the iron deficiency criteria. Definition of abbreviations: F = female; M = male; S.Asian = South Asian; log = natural logarithm; sTfR = soluble transferrin receptors; ferr = ferritin; TSAT = transferrin saturation.

There was no association between sTfR/ferritin ratio and IL-6 in a regression model adjusted for age, sex and diagnosis in either the white ($Effect_{IL\text{-}6}$ = 0.06; 95% CI = -0.03 – 0.14; p-value = 0.22) or the

non-white cohort ($Effect_{IL\text{-}6}$ = -0.1; 95% CI = -0.34 – 0.14; p-value = 0.4). sTfR/ferritin ratio is considerably higher in the Ferr-TSAT+sTfR group of the non-white cohort ($Effect_{Ferr\text{-}TSAT+sTfR}$= 1.73; 95% CI = 1.18 – 2.27; p-value = $1.97x10^{-8}$, reference group: replete), but the smaller differences between the replete and sTfR-only and replete and Ferr-TSAT groups were not significant.

3.3.4 Hepcidin levels are similar between H/IPAH, PAH associated with LHD and disease controls.

Hepcidin was observed to be abnormally high in iron deficient (iron deficiency defined as sTfR > 28.1 nmol/L) IPAH patients and proposed as a possible mechanism behind the high prevalence of iron deficiency in IPAH patients (Rhodes et al., 2011a). Abnormally high hepcidin could result from decreased *BMPR2* expression in PAH (Atkinson et al., 2002) which leads to BMP-mediated upregulation of hepcidin via the activin type II receptor (Rhodes et al., 2011a, Yu et al., 2005).

In this study, hepcidin levels were modeled as a function of sTfR/ferritin ratio, IL-6, diagnosis, age and sex, stratified by ethnicity. In the white cohort, levels of hepcidin are similar in H/IPAH, PH LHD and disease controls whilst PAH CTD, PAH CHD and CTEPH patients present with lower hepcidin levels ($Effect_{CTD}$ = -0.74; 95% CI = -1.00 – -0.48; p-value = $4.12x10^{-8}$. $Effect_{CHD}$ = -0.47; 95% CI = -0.8 – -0.13; p-value = 0.006. $Effect_{CTEPH}$ = -0.71; 95% CI = -0.88 – -0.53; p-value = $4.08x10^{-15}$, reference group: disease control) independent of iron availability as measured by the sTfR/ferritin ratio (Figure 3.7). In the non-white cohort, H/IPAH patients have higher ($Effect_{H/IPAH}$ = 0.63; 95% CI = 0.25 – 1.01; p-value = 0.001, reference group: disease control) hepcidin levels than disease controls (Figure 3.7).

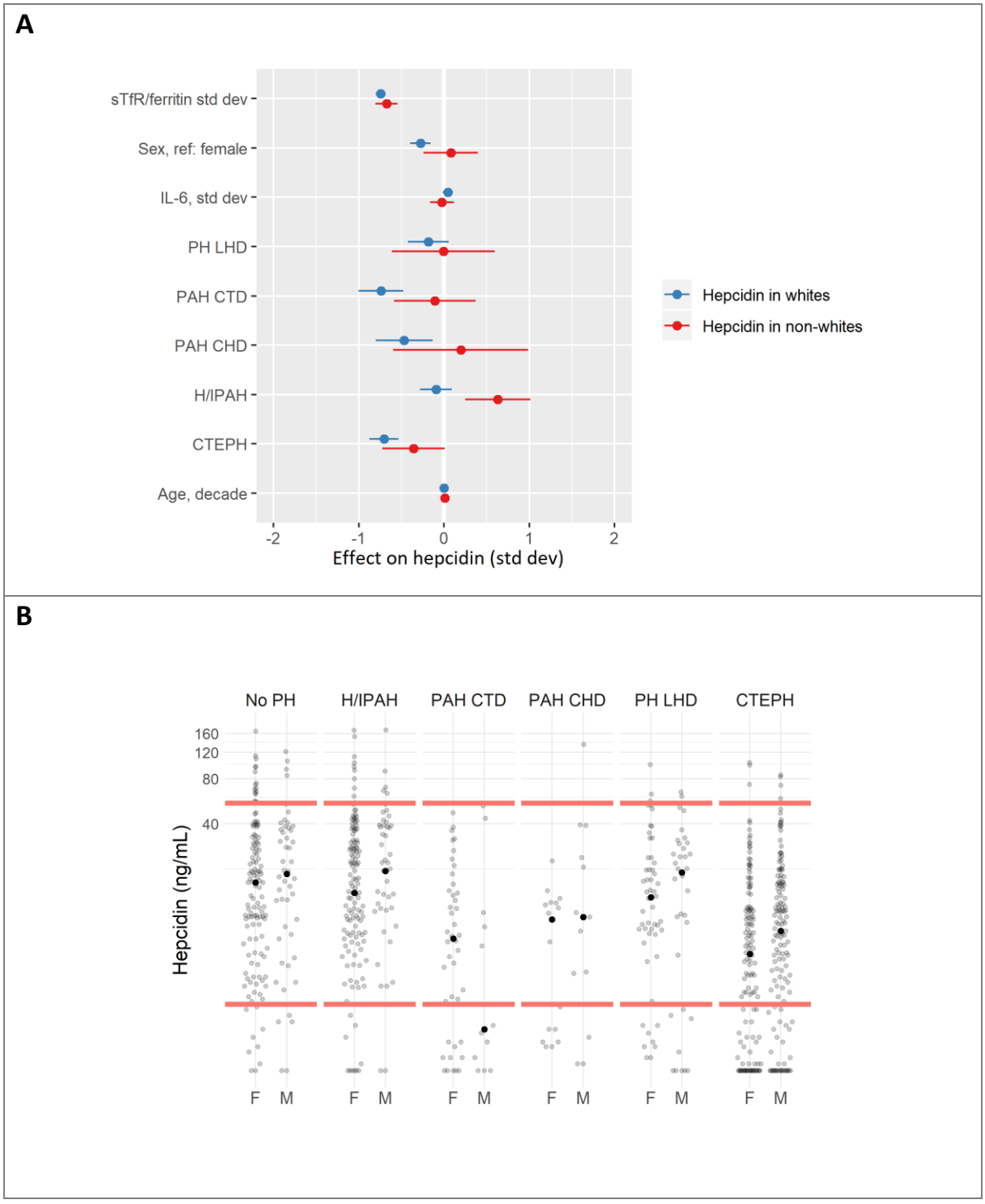


Figure 3.7 Predictors of hepcidin and its levels in disease controls and five PH diagnostic groups.
A: *Effect size is in hepcidin standard units. Horizontal error bars represent 95% confidence intervals.* ***B:*** *Black dots represent the median value of observations, horizontal red lines delineate normal hepcidin range (2.5 – 55 ng/mL). Definition of abbreviations: IL-6 = interleukin-6; Ferr = ferritin; sTfR = soluble transferrin receptors; F = female; M = male; No PH = disease controls; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; PAH CTD = pulmonary arterial hypertension associated with connective tissue disease; PAH CHD = pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension.*

The strongest predictor of hepcidin levels is sTfR/ferritin ratio in both white and non-white patients (white: $Effects_{TfR/Ferr}$= -0.74; 95% CI = -0.80 – -0.68; p-value < $2x10^{-16}$. non-white: $Effects_{TfR/Ferr}$= -0.67; 95% CI = -0.81 – -0.54; p-value = $4.29x10^{-16}$). Compared to the iron replete group, hepcidin is lower in the Ferr-TSAT ($Effect_{Ferr-TSAT}$= -0.54; 95% CI = -0.77 – -0.31; p-value = $5.43x10^{-6}$, reference group: replete) and Ferr-TSAT+sTfR groups ($Effect_{Ferr-TSAT+sTfR}$= -1.23; 95% CI = -1.41 – -1.06; p-value < $2x10^{-16}$, reference group: replete) after adjustments for age, sex, diagnosis and ethnicity (Figure 3.8). No difference in hepcidin was observed between iron replete patients and those belonging to the sTfR-only group ($Effect_{sTfR-only}$= -0.13; 95% CI = -0.31 – 0.05; p-value = 0.15, reference group: replete) (Figure 3.8).

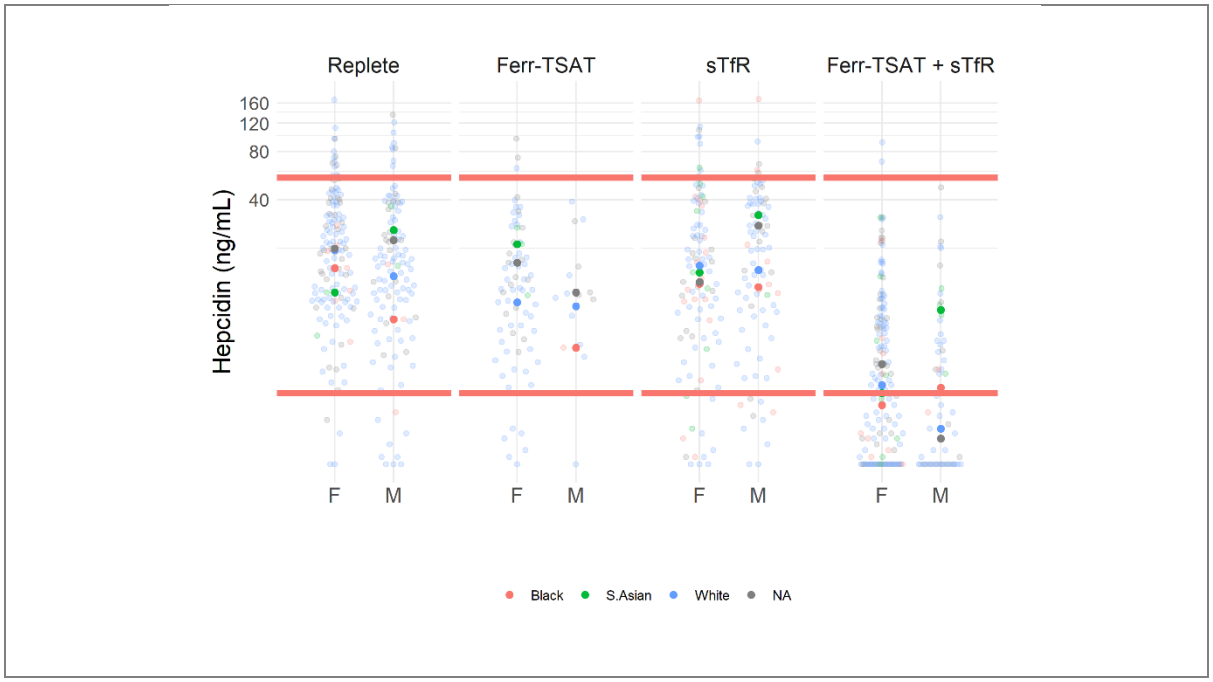


Figure 3.8 Hepcidin levels by iron deficiency status.
Each observation is colored by ethnicity. Dark red, green, and blue dots represent the median for black, South Asian and white groups, respectively. Observations in grey are of patients with missing ethnicity information. Horizontal red lines delineate normal hepcidin range (2.5 – 55 ng/mL). The Ferr-TSAT group meets both ferritin (< 100 µg/L) and TSAT (< 20%) cut-offs for iron deficiency, the sTfR group meets the sTfR threshold (> 28.1 nmol/L) only, while the Ferr-TSAT+sTfR group meets the cut-off for all three biomarkers of iron deficiency. Individuals in the replete group do not meet any of the iron deficiency criteria. Definition of abbreviations: F = female; M = male; S.Asian = South Asian; sTfR = soluble transferrin receptors; ferr = ferritin; TSAT = transferrin saturation.

3.3.5 sTfR levels do not prognosticate, but belonging to the sTfR-only group predicts mortality.

Iron deficiency was reported (Rhodes et al., 2011a) to predict mortality independent of 6MWD and functional class in PAH, using sTfR levels as the iron status biomarker in their prognostic model. The prognostic value of iron status in this study was investigated using sTfR levels, the ratio of sTfR and ferritin (sTfR/ferritin) and the four definitions of iron deficiency (Replete, Ferr-TSAT, sTfR-only, Ferr-TSAT+sTfR). All Cox proportional hazards models to predict all-cause mortality from time of patient blood sample were adjusted for age, sex, IL-6 levels, 6-minute walk distance (6MWD), ethnicity and diagnostic group. PAH patients with associated conditions (PAH CTD and PAH CHD) were combined in a group called 'other PAH'.

Median follow-up time was 4.5 years during which 188 patients died. As expected, disease controls show the best, and PAH CTD patients the worst prognosis among diagnostic groups (p-value = 0.0023) (Supplementary Figure 8.1.1). Soluble transferrin receptor (sTfR) levels did not predict mortality (HR_{sTfR} = 1.08; 95%CI = 0.90 – 1.29; p-value = 0.40) after adjusting for age, sex, IL-6, 6MWD, ethnicity and diagnosis (Figure 3.9).

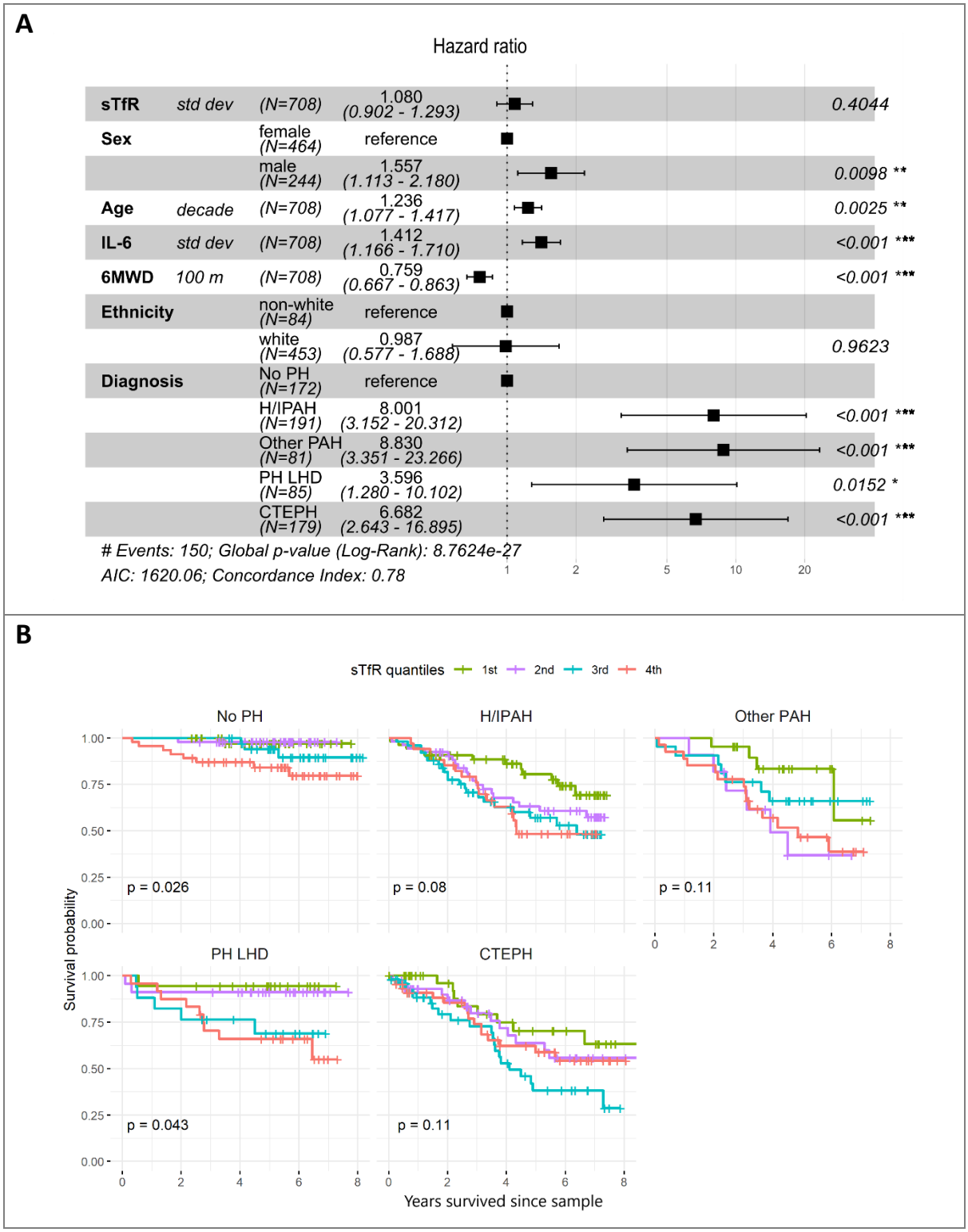


Figure 3.9 Prognostic value of soluble transferrin receptor (sTfR) levels.
A: *Results of the Cox proportional hazards model.* ***B:*** *Kaplan-Meier survival estimates stratified by quantiles of soluble transferrin receptor (sTfR) levels. Green, purple, blue and red curves correspond to the first, second, third and fourth sTfR quantile, respectively. Definition of abbreviations: No PH = disease controls; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; Other PAH = pulmonary arterial hypertension associated with connective tissue disease and pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension; IL-6 = interleukin 6; 6MWD = 6-minute walk distance.*

The ratio of sTfR and ferritin showed some prognostic value (HR = 0.81; 95%CI = 0.68 – 0.96; p-value = 0.01) independent of the effects of age, sex, IL-6, 6MWD, ethnicity and diagnosis (Figure 3.10). One standard unit increase in sTfR/ferritin was associated with a 19% decrease in risk of dying which translates to worse iron status (as defined by sTfR/ferritin) being protective.

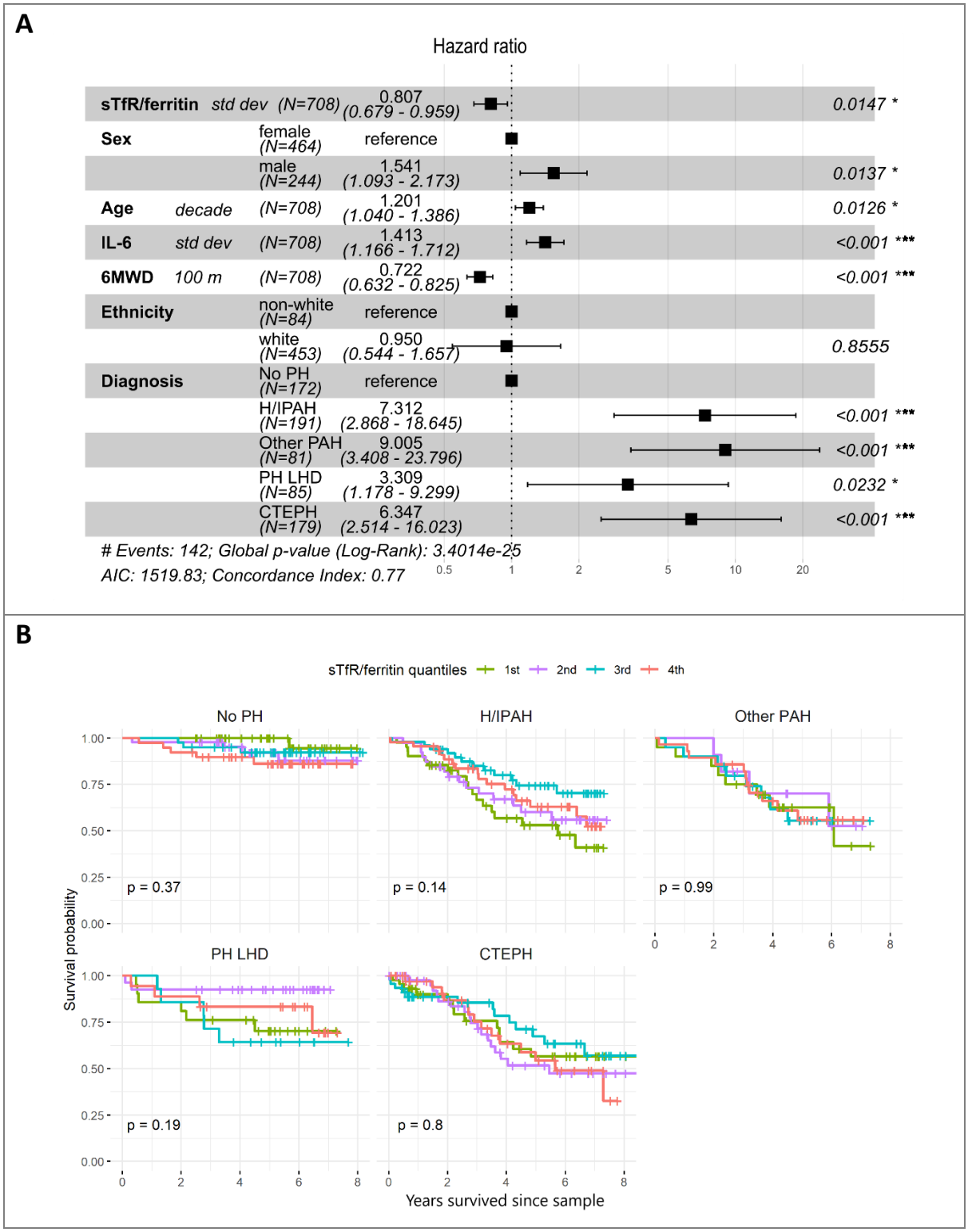


Figure 3.10 Prognostic value of soluble transferrin receptor and ferritin ratio (sTfR/ferritin).
A: *Results of the Cox proportional hazards model.* ***B:*** *Kaplan-Meier survival estimates stratified by quantiles of sTfR/ferritin. Green, purple, blue and red curves correspond to the first, second, third and fourth sTfR quantile, respectively. Definition of abbreviations: No PH = disease controls; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; Other PAH = pulmonary arterial hypertension associated with connective tissue disease and pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension; IL-6 = interleukin 6; 6MWD = 6-minute walk distance.*

Iron deficiency status as divided into four groups based on ferritin, transferrin saturation and sTfR levels predicted mortality ($HR_{sTfR\text{-}only}$ = 2.1; 95%CI = 1.31 – 3.30; p-value = 0.002, reference: replete) independent of age, sex, IL-6, 6MWD, ethnicity and diagnosis. Those belonging to the sTfR-only group had more than twice the risk of dying compared to iron replete individuals, which is the second largest effect on survival after diagnosis in the model. The other two iron deficient groups (Ferr-TSAT and Ferr-TSAT+sTfR) were not prognostic (Figure 3.11). Poor prognosis of patients in the sTfR-only group is most evident in the H/IPAH diagnostic group (Figure 3.11/B), but the survival effect is robust to excluding the H/IPAH cohort ($HR_{sTfR\text{-}only}$ = 2.0; 95%CI = 1.14 – 3.50; p-value = 0.02, reference: replete). Sensitivity analyses were run excluding the sTfR-only group to test a potentially masked effect of iron deficiency. There was no evidence for the prognostic value of sTfR levels or the two iron deficiency groups (HR_{sTfR} = 0.95; 95%CI = 0.74 – 1.22; p-value = 0.68. $HR_{Ferr\text{-}TSAT}$ = 0.81; 95%CI = 0.38 – 1.72; p-value = 0.59. $HR_{Ferr\text{-}TSAT+sTfR}$ = 1.01; 95%CI = 0.60 – 1.71; p-value = 0.96, reference: replete) in the cohort excluding the sTfR-only group.

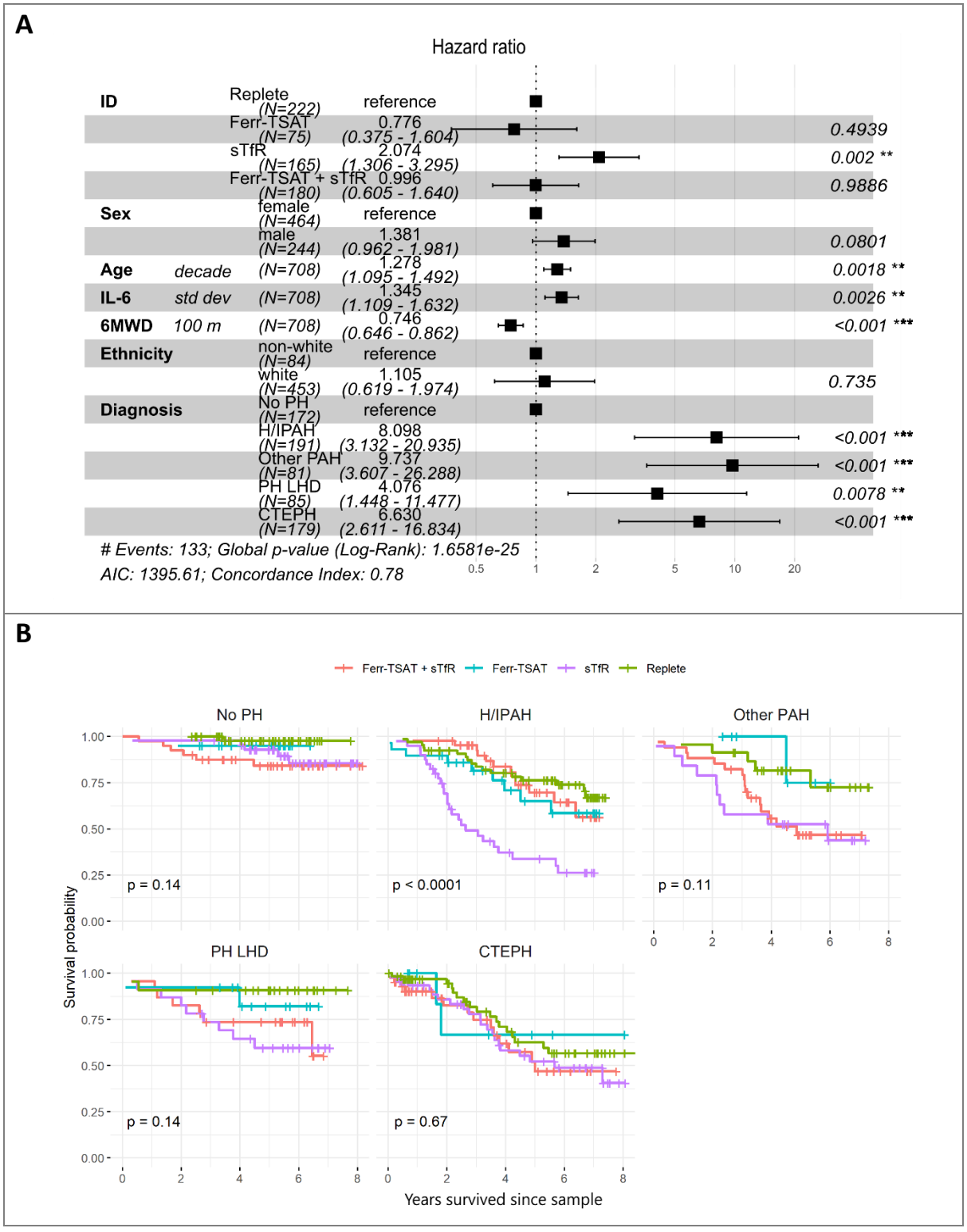


Figure 3.11 Prognostic value of iron deficiency groups.

A: *Results of the Cox proportional hazards model.* ***B:*** *Kaplan-Meier survival estimates stratified by iron deficiency group. The Ferr-TSAT group meets both ferritin (< 100 µg/L) and TSAT (< 20%) cut-offs for iron deficiency, the sTfR group meets the sTfR threshold (> 28.1 nmol/L) only, while the Ferr-TSAT+sTfR group meets the cut-off for all three biomarkers of iron deficiency. Individuals in the replete group do not meet any of the iron deficiency criteria. Definition of abbreviations: No PH = disease controls; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; Other PAH = pulmonary arterial hypertension associated with connective tissue disease and pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension; IL-6 = interleukin 6; 6MWD = 6-minute walk distance.*

3.4 Discussion

Previous studies have shown that iron deficiency with or without anemia is both common (40% - 60%) and associated with poor prognosis, reduced exercise capacity and worsening hemodynamics in idiopathic PAH (IPAH) (Rhodes et al., 2011a, Soon et al., 2011, Ruiter et al., 2011, Krasuski et al., 2011, Sonnweber et al., 2018). High prevalence (38%) of iron deficiency and its association with adverse outcomes have also been described in patients with Eisenmenger`s syndrome, a severe form of PAH CHD (Van De Bruaene et al., 2011). This study aimed to assess the prevalence and prognostic value of iron deficiency in a large single-center cohort of pulmonary hypertension (PH) patients with diverse etiologies, referred to the National Pulmonary Hypertension Service at Hammersmith Hospital. Three of the five PH diagnostic groups were represented: WHO group 1 PH including IPAH, PAH associated with congenital heart disease (PAH CHD) and connective tissue disease (PAH CTD), WHO group 4 chronic thromboembolic PH and WHO group 2 PH due to left heart disease.

In this study, iron deficiency defined by either elevated soluble transferrin receptors (sTfR > 28.1 nmol/L) or low ferritin (Ferr < 100 µg/L) and transferrin saturation (TSAT < 20%) levels, was common (34-56% meeting Ferr-TSAT and 39-69% meeting sTfR thresholds) in all patient groups including disease controls without PH. PAH CTD patients had a higher frequency (50%) of iron deficiency than controls and other diagnostic groups considering both ferritin/TSAT-based and sTfR-based criteria. PAH CTD is the most severe form of PAH and shows the worst prognosis among PH diagnostic groups. The frequent cooccurence of low ferritin, (TSAT) and high sTfR in PAH CTD could be a sign of advanced disease where high demand for iron by increased red blood cell production (causing elevated sTfR) to cope with worsening tissue hypoxia depletes iron stores. It is unclear at present whether iron replacement therapy has any clinical benefit through either cardiovascular pathology or peripheral effects in any of the PH diagnostic groups. However, iron deficiency being a ubiquitous comorbidity across several PH diagnostic groups and disease controls from the PH clinic might suggest that a causal effect is less likely or very modest.

Diagnosing iron deficiency based on clinical blood biomarkers in individuals with chronic disease is a challenging task as chronic disease itself and iron deficiency are the two most common causes of anaemia (Cash and Sears, 1989). The circulating (soluble) form of transferrin receptors (sTfR) has been reported as a marker of iron deficiency unaffected by inflammation, providing an alternative for bone marrow biopsies in patients with acute or chronic inflammation (Punnonen et al., 1994, Cook et al., 1990). Levels of sTfR moderately increase in individuals with iron deficiency without anaemia and are more drastically elevated in iron deficiency anaemia when marrow functional iron

deficiency develops (Punnonen et al., 1994), although high values are not specific to iron deficiency and are also observed in conditions associated with increased erythropoiesis (Mast et al., 1998). Studies evaluating the predictive ability of iron status biomarkers in individuals with chronic inflammatory diseases using bone marrow biopsies found that whilst sTfR showed high sensitivity, it had generally low specificity (Mast et al., 1998, Zoli et al., 1994, Juncà et al., 1998, Nielsen et al., 1994), and suggested the use of a higher sTfR cut-off and the use of sTfR and ferritin combined in a ratio index (Remacha et al., 1998, Bultink et al., 2001, Means Jr et al., 1999, Suominen et al., 2000). Similar studies using bone marrow staining to evaluate the sensitivity and specificity of iron indices in identifying iron deficiency have not been conducted in pulmonary hypertension to date. However, the results in a wide range of other chronic diseases mentioned above possibly apply to PH as well.

Individuals in this study were assigned to one of three iron deficient- or the iron replete category of which the one defined by the combination of ferritin, TSAT and sTfR (Ferr-TSAT+sTfR) is comparable to the combined use of ferritin and sTfR suggested by the experimental studies mentioned above. The log ratio of sTfR and ferritin, shown to outperform any other available blood biomarker for correctly identifying iron deficiency in chronic disease, served as an index of iron availability in the prediction and prognostic models of this study. Levels of sTfR/ferritin were lowest in the iron replete group, increased twice as much from iron replete to the Ferr-TSAT group than from iron replete to the sTfR-only group and were drastically higher (1.6 standard units) in the Ferr-TSAT+sTfR group.

Anaemia prevalence ranged between 10% and 44% in the five PH groups and disease controls. Levels of haemoglobin (Hgb) were higher in H/IPAH, PAH CHD and CTEPH patients than disease controls in white and non-white individuals alike. Hgb was generally lower in individuals from black or South Asian backgrounds than in white individuals and its levels increased with increasing tissue iron availability measured by the ratio of sTfR/ferritin in white patients only. Low Hgb, high sTfR and normal ferritin levels in the non-white cohort could suggest anaemia due to factors different from iron deficiency, such as vitamin B12 or folate deficiencies.

Soluble transferrin receptor (sTfR) levels were independently predicted by the inflammatory cytokine IL-6, the iron storage protein ferritin and the diagnosis of PAH CTD (in whom they were higher) in white patients. In the patient cohort from black or South Asian backgrounds, decreasing Hgb levels predicted elevated sTfR, suggesting stimulated erythropoiesis, although this could not be tested as erythropoietin measurements were not available. These differences in iron status and anaemia between white and non-white individuals should be considered in future investigations into iron deficiency in PH. H/IPAH, PAH CTD and CTEPH diagnostic groups had significantly lower sTfR than disease controls after adjusting for the effects of Hgb, IL-6 and ferritin in the non-white cohort.

sTfR was higher in the non-white cohort than in white individuals and was not associated with IL-6 and ferritin levels in non-white individuals. The correlation of sTfR and Hgb and the lack of correlation between sTfR and iron markers such as ferritin have been described with erythropoietic activity in the iron replete state (Beguin, 2003). Comparisons between white and non-white individuals were all adjusted for diagnosis, although analyses of ethnicity effects within diagnostic groups were not possible due to the small sample size of non-white individuals in the PH diagnostic groups.

Unlike sTfR, the ratio of sTfR and ferritin was not associated with IL-6 and it predicted hepcidin levels. Hepcidin plays a pivotal role in iron homeostasis by suppressing iron absorption from the gut. Hepcidin expression is upregulated by circulating iron and inflammation and inhibited by erythropoiesis (*See* 1.6 Iron homeostasis and iron biomarkers). A previous study reported abnormally high hepcidin levels in a significant proportion of IPAH patients defined as iron deficient based on sTfR levels (> 28.1 nmol/L) only (Rhodes et al., 2011a). Rhodes et al. argued that inappropriate hepcidin production may contribute to iron deficiency in IPAH by inhibiting dietary iron uptake and the release of iron from iron storing cells. They also demonstrated that knocking down *BMPR2* in hepatoma cells leads to increased hepcidin levels as a response to BMP-6 stimulation (*See* 1.8 Iron deficiency in PAH). *BMPR2* is underexpressed in the lungs of PAH patients harboring *BMPR2* mutations and also in the lungs of mutation-negative idiopathic and associated PAH patients (Atkinson et al., 2002). Hepcidin dysregulation in IPAH is supported by the observation that only a small proportion of IPAH patients respond to oral iron therapy (Ruiter et al., 2011). In this study, whilst hepcidin levels were similar in disease controls, H/IPAH and PH LHD patients in the white cohort, in non-white patients H/IPAH diagnosis predicted higher hepcidin levels independent of sTfR/ferritin, IL-6, age and sex. However, higher sTfR and lower Hgb levels as well as the lack of correlation between sTfR and ferritin in the non-white cohort could be signs of erythropoiesis which was not accounted for in the hepcidin model and could result in lower hepcidin in non-white controls. Rhodes et al. found that the difference in hepcidin levels between IPAH patients and healthy controls was robust to adjusting for erythropoietin levels (Rhodes et al., 2011a).

Iron deficiency in previous studies was associated with worse prognostic markers, such as 6-minute walk distance (6MWD) and predicted survival in IPAH patients (Rhodes et al., 2011a, Ruiter et al., 2011). In the study of Rhodes et al. sTfR associated with worse survival in IPAH patients independent of 6MWD. In this cohort, sTfR did not predict survival after adjustments for 6MWD, IL-6, ethnicity, diagnosis, sex and age. On the other hand, belonging to the sTfR-only group had a strong effect on survival which was especially pronounced in the H/IPAH diagnostic group. Those in the sTfR-only group had more than twice the risk of dying than those in the iron replete group. The other two iron

deficiency groups were not at increased risk compared to iron replete individuals. This suggests that the survival effect of being in the sTfR-only group is likely not caused solely by iron deficiency, as if that was the case, the more severely iron deficient Ferr-TSAT+sTfR group would show at least some degree of risk increase against the iron replete group. A possible explanation for the poor prognosis of patients in the sTfR-only group is the cooccurrence of iron deficiency and inflammation, i.e. they might have elevated sTfR due to iron status but fail to meet ferritin and TSAT criteria due to inflammation. Even though inflammation was accounted for in the form of interleukin-6 (IL-6) levels in the survival models, residual confounding may remain as a result of measurement imprecision and the inability of IL-6 levels to fully capture systemic inflammation. Residual inflammation could explain the modest protective effect of sTfR/ferritin, which would suggest, somewhat counterintuitively, that worse tissue iron availability decreases risk of dying. Patients in the sTfR-only group have similar iron indices to those in the iron replete group and therefore are not likely to represent a milder form of iron deficiency. Sensitivity analyses excluding the sTfR-only group did not provide evidence for the prognostic value of iron deficiency proxied by either sTfR levels or iron deficiency status. These observations suggest that other processes underlying more advanced disease instead of iron deficiency could be the cause of elevated sTfR in the sTfR-only group. The survival effect of non-iron related sTfR increase could be explained by a systemic response to severe chronic local hypoxia, i.e. the expansion of immature red blood cell precursors in the bone marrow, leading to more membrane bound TfR as well as higher sTfR. A gene expression signature specific to erythroid differentiation discriminated between a group of PAH patients including idiopathic and scleroderma-associated cases (PAH SSC) and a group of controls including systemic sclerosis patients without PH and healthy controls (Cheadle et al., 2012). This gene expression cluster contained genes expressed almost exclusively in erythroid progenitors and in the bone marrow. The genes *ALAS2* (aminolevulinate, delta-, synthase2) and *ERAF* (erythroid differentiation associated factor) within the cluster showed markedly increased levels in IPAH and PAH SSC patients, both of which are essential for the terminal differentiation of erythroid cells. This gene signature was also associated with increasingly severe disease measured by hemodynamic parameters in IPAH but not PAH SSC, which is in line with the prominent risk increase of the sTfR-only group in IPAH patients compared to other diagnostic groups in this study. The presence of this gene signature characterized by Cheadle et al. and the elevated sTfR levels observed in PAH are indicative of increased red blood cell recruitment and may constitute a systemic response to severe chronic local hypoxia.

The congenital heart disease- and connective tissue disease associated PAH groups (PAH CHD and PAH CTD, respectively) were pooled into a single group of associated PAH for the survival analyses given the relatively small sample size and therefore small number of events in these two groups.

Therefore it remains possible that the prognostic value of iron deficiency differs between PAH CHD and PAH CTD groups that this study was not powered to detect.

This study differs in certain aspects from previously published observational studies of iron deficiency in PAH which, at least in part, could underlie the disparities between them. Iron indices and/or cut-offs used in the criteria for iron deficiency differ between this study and previous studies of iron deficiency in PAH (Rhodes et al., 2011a, Ruiter et al., 2011, Soon et al., 2011). Rhodes et al. (Rhodes et al., 2011a) defined iron deficiency on the basis of sTfR levels using the upper limit of the normal range (> 28.1 nmol/L) as a cut-off (same cut-off as in the sTfR-only and Ferr-TSAT+sTfR groups in this study) and found 62 out of 98 IPAH patients (63%) iron deficienct. Ruiter et al. used the combination of serum iron (< 10 µmol/L) and TSAT (< 15% in females and < 20% in males) for the definition of iron deficiency which classified 30 out of 70 IPAH (43%) iron deficient. Soon et al. (Soon et al., 2011) used ferritin (< 10 µg/L), transferrin (> 2.2 g/L) and normal C-reactive protein levels for defining iron deficiency in a cohort of 85 IPAH patients of which 26 (31%) were found iron deficient. The cut-off values in the studies of Ruiter et al. and Soon et al. seem conservative for defining iron deficiency in chronic disease patients which could have lead to underestimation of iron deficiency prevalence in their study populations. On the other hand, the iron deficiency criterium applied by Rhodes et al. might have been too liberal considering the evidence that the upper limit of normal sTfR (28.1 nmol/L) is likely too low in individuals with chronic disease (Nielsen et al., 1994, Juncà et al., 1998, Remacha et al., 1998, Mast et al., 1998, Means Jr et al., 1999, Suominen et al., 2000, Bultink et al., 2001). For instance, the ferritin thereshold of 10 and 22 µg/L for females and males has an estimated specificity of 99%, but sensitivity as low as 25% whilst the sTfR cut-off of 28.1 nmol/L is less specific (74%) but more sensitive (71%) when predicting bone marrow iron content in a heterogenous population of patients (Means Jr et al., 1999).

Prevalence of iron deficiency was similar between CTEPH and H/IPAH patients in this cohort, which is not in agreement with the findings of Soon et al. who reports 30% of IPAH patients and only 5% of CTEPH patients who served as controls in their cohort to be iron deficient (Soon et al., 2011). Twenty-five percent of H/IPAH patients and 28% of CTEPH patients were iron deficient in this study, considering the strictest criteria including all three iron indices (Ferr-TSAT+sTfR). Prevalence of iron deficiency in CTEPH patients was still 25% after excluding those who underwent pulmonary endarterectomy before blood biomarkers were measured. Age at diagnosis and the age at which patients were sampled are similar between this study and the study of Soon et al. for both diagnostic groups, however the sample size is over twice as large in this cohort which could result in more accurate prevalence estimates. Even though no exclusions based on clinical history or co-existing morbidities such as gastrointestinal bleeding, transfusion-dependent conditions or renal failure were

made in this study, this should not impact considerably the relative proportions of iron deficiency since exclusions were not applied to either CTEPH or H/IPAH. This would hold true assuming that the two diagnostic groups had similar proportions of patients with co-morbidities likely affecting iron status, as observed in the study of Soon and colleagues (Soon et al., 2011).

3.5 Conclusions

Iron deficiency has not been well characterized in PH, other than IPAH. This study suggests that iron deficiency is prevalent in all patients admitted to the PH clinic, but it is not associated with worse survival. Elevated levels of soluble transferrin receptor associated with worse survival, which was especially prominent in IPAH patients, although this seemed to be more complicated than simply being due to iron deficiency. The diagnosis and assessment of iron deficiency severity in a chronic disease like PH is a challenging task, but soluble transferrin receptor levels used in combination with ferritin could provide a measure with higher specificity. The fact that iron deficiency was equally common in patients admitted to the PH clinic but found not to have PH begins to question the likelihood of iron deficiency being a causal pathway in the development of PH.

4 Mendelian randomisation analysis of red cell distribution width in pulmonary arterial hypertension

4.1 Introduction

In chapter three I showed that iron deficiency is common in all patients attending the PH clinic. In this chapter I focus on group 1 PH, specifically IPAH and heritable PAH, to assess the association of iron status and PAH susceptibility for causality using genetic variants as instruments. Iron deficiency is currently under investigation as a therapeutic target in PAH (Howard et al., 2013).

PAH is a rare and complex disease with both genetic and environmental factors contributing to the pathogenesis of PAH (Southgate et al., 2020). The effects of rare (Graf et al., 2018) and common (Rhodes et al., 2018) genetic variants have recently been described in the largest to date cohort of PAH patients utilizing genome-wide array and whole genome sequencing data from multiple centers (*Materials and methods* 2.9). Approximately 70–80% of heritable PAH and 10–20% of IPAH patients are known to harbour mutations in the bone morphogenetic protein type II receptor (*BMPR2*) gene (Evans et al., 2016). The study of Gräf et al. involving over one thousand PAH patients confirmed the prevalence of causal mutations in *BMPR2*, as well as in five other established genes (*TBX4, ACVRL1, ENG, SMAD9,* and *KCNK3*), and identified PAH-associated mutations in four new genes (*ATP13A3, SOX17, AQP1,* and *GDF2*), all together accounting for 23.5% of the cases studied (Graf et al., 2018). The rare mutations in all these genes are inherited in an autosomal dominant manner and exhibit reduced penetrance, indicating that other genetic, epigenetic and/or environmental factors influence the development of PAH.

Red cell distribution width (RDW) has been demonstrated to strongly correlate with survival in PAH (Rhodes et al., 2011b, Hampole et al., 2009). A recent hypothesis-free phenome-wide analysis indicated PAH, among several disease descriptors, as the most strongly associated trait with RDW (odds ratio, OR = 2.0, 95% CI = 1.75-2.4 per % increase in RDW) in a hospital population (Thayer et al., 2019). RDW - a measure of red blood cell (RBC) size variability in an individual – is part of the full blood count in standard hospital practice and readily available as a biomarker. RDW correlates well with iron status biomarkers and high values can signal early stage iron deficiency or iron deficiency anemia (Emans et al., 2011). RDW increases with decreasing iron as available body iron stores fail to meet the iron demand of RBC synthesis resulting in RBCs of varied size.

An observed correlation between two traits does not necessarily imply that interventions on one trait will change the other and there are numerous examples where false positive associations have led to unsuccessful randomized controlled trials (Hennekens et al., 1996, 2002). Clinical trials are expensive and time-consuming, and recruitment can be challenging, especially in rare diseases. With the growing availability of genetic data in large disease-focused and population-based studies, testing causal relationships between traits of interest has become possible by harnessing the naturally occurring genetic variation in the population. The collection of methods used to test causal relationships using genetic variants is called Mendelian randomisation (MR)(described in detail in *Introduction* 1.9 and *Materials and methods* 2.6, 2.7). MR has been successfully used to help prioritize intervention and drug targets and to identify causal factors for several diseases (Pichler et al., 2013, Collaboration et al., 2011, Interleukin-6 Receptor Mendelian Randomisation Analysis et al., 2012, Lewis et al., 2011, Palmer et al., 2013). In general, candidate drugs with genetic evidence for effectiveness are more successful in drug trials compared to those without such genetic support (King et al., 2019).

It remains to be investigated whether elevated RDW is largely a consequence of PAH or plays a causal role in the condition. The common genetic variation determining RDW levels has been defined in very large population studies described in *Materials and methods* 2.9 (Astle et al., 2016), with more power than equivalent studies of iron status which makes RDW a strong candidate for MR analysis. The primary aim of this chapter is to investigate the epidemiological association between RDW and PAH by assessing causality using MR. I refined the estimate of association between RDW and PAH using biomarker data in 642 PAH cases and over 15,000 controls with common diseases from the Imperial College PH Biobank, the UK PAH Cohort and the Vanderbilt University Medical Centre (VUMC) and applied MR to investigate whether being genetically predisposed to higher levels of RDW increases the odds of developing PAH.

4.2 Methods

4.2.1 Definition of pulmonary arterial hypertension (PAH)

PAH was defined by internationally agreed criteria (Galie et al., 2016), specifically mean pulmonary artery pressure >25mmHg, pulmonary vascular resistance >3 Woods and pulmonary capillary wedge pressure <15 mmHg. Patients with concurrent diseases known to cause pulmonary hypertension were excluded such that all were considered to have idiopathic, heritable or anorexigen-induced PAH.

4.2.2 Genetic and phenotype data in contributing studies

4.2.2.1 Individual-level data

Individual-level data including clinical laboratory RDW was available for a total of 524 PAH patients from the Imperial College PH Biobank and the multi-centre UK PAH Cohort, a study that facilitates collaboration and the sharing of data and samples between the specialist pulmonary hypertension centres in the United Kingdom.

In addition, the hospital population-based VUMC study provided longitudinal clinical laboratory RDW measurements, detailed clinical diagnoses and genome-wide genotype array data (genotyping platform: Illumina MEGAex) for an additional 118 PAH patients and 15,889 common disease controls (Table 4.1). VUMC hosts a collection of electronic medical records linked to genetic data derived from blood collected during routine clinical assessment in outpatient clinics where all patients are shown the consent form during check-in (Roden et al., 2008). Details on the quality control and genotype imputation procedures of VUMC genotype array data are described in *Methods* 0.

	PAH (118 VUMC; 524 NIHRBR)		Controls (15,889 VUMC)	
	Female (%)	Male (%)	Female (%)	Male (%)
N	445 (69)	197 (31)	8,539 (54)	7,350 (46)
RDW (mean/SD)	15.1/2.20	15.5/2.33	13.6/1.41	13.6/1.39
Age (mean/SD)	52.4/17.6	58.2/16.2	54.3/16.1	58.1/14.7

Table 4.1 Characteristics of the study population used for estimating the association between RDW and PAH, stratified by sex.
Definition of abbreviations: PAH – Pulmonary Arterial Hypertension, VUMC - Vanderbilt University Medical Centre, NIHRBR - UK National Institute for Health Research BioResource, RDW – red cell distribution width.

Individuals from non-white ethnic backgrounds were excluded to avoid potential bias from ethnicity effects. Since RDW is known to be elevated in a number of diseases, all individuals with either of the

following in their medical history were excluded: polycystic kidney disease, chronic kidney disease, liver disease and transfusion therapy received. Furthermore, children and adolescents under the age of 18 as well as individuals with extreme RDW values (below 10%, N=20 or above 30%, N=15) were not included in the analysis (Figure 4.1). Out of the 35 excluded for extreme RDW, 14 were cases from VUMC in the range of 43-55 and 20 were all below 10 from one of the NIHRBR centers. Both centers confirmed reporting errors for these samples.

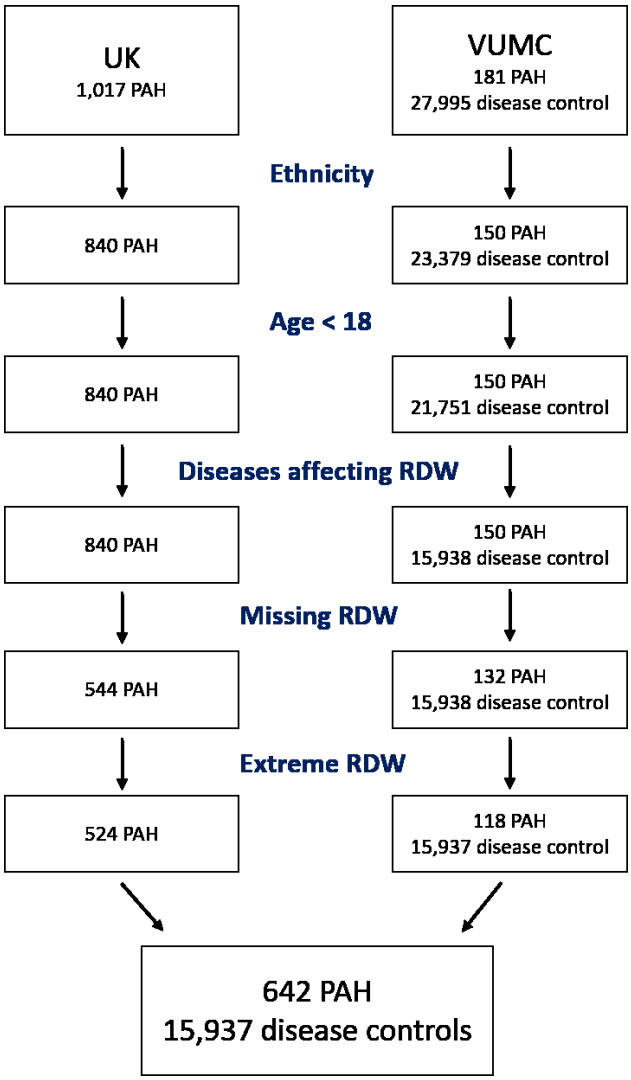


Figure 4.1 Flow diagram of the exclusion steps in the UK centers and the Vanderbilt University Medical Centre (VUMC) participating in the RDW and PAH association analysis.
Definition of abbreviations: PAH – Pulmonary Arterial Hypertension, VUMC - Vanderbilt University Medical Centre; RDW – red cell distribution width.

4.2.2.2 Summary-level data

Summary-level data for RDW and PAH susceptibility were used in the MR analyses to maximize the sample size and therefore the statistical power. Genetic instruments serving as a proxy for RDW were obtained from the largest-to-date (over 170,000 individuals) GWAS on hematological traits

(hereon referred to as the 'RDW GWAS') in a population-based study by Astle et al. (Astle et al., 2016).

A large proportion of the idiopathic and familial PAH cases from the UK PAH Cohort were whole genome sequenced as part of the UK National Institute for Health Research BioResource (NIHRBR) (Rhodes et al., 2018, Graf et al., 2018) study. For genetic effects on PAH, I used the largest GWAS of PAH susceptibility published by our group and others investigating the effects of common genetic variation on PAH risk (hereon referred to as the 'PAH GWAS') involving a total of 11,744 individuals of which 2,085 were PAH cases. Effect estimates of the PAH GWAS were combined through the meta-analysis of four independent studies, one of which is the NIHRBR with 847 PAH cases and 5,048 rare disease controls. The other major contributing study, the US PAH Biobank (PAHB) used a control population with mixed common diseases recruited from outpatient clinics (694 PAH cases and 1560 controls), whilst the two smaller studies used population-based controls, with 269/275 PAH cases and 1068/1983 controls in the PHAAR and BHFPAH studies, respectively (Rhodes et al., 2018).

4.2.3 Statistical analyses

4.2.3.1 RDW and PAH association

To confirm and refine the estimate for the association between RDW and PAH, PAH patients from the Imperial College PH Biobank, the UK PAH Cohort and the VUMC were combined and compared to common disease controls from the VUMC recruited in outpatient clinics. RDW values were natural log-transformed and for ease of interpretation Z-score normalized to have a mean of zero and a standard deviation of one. The association between RDW levels and PAH was tested in a logistic regression framework adjusted for age and sex. Closest to diagnosis RDW measurements were used for PAH cases and the first available RDW measurement was retrieved for controls.

4.2.3.2 Causal effect estimation

To test for causality between RDW and PAH, I applied a two-sample MR strategy that requires effect estimates for the genetic instrument on the risk factor (here RDW) and the outcome (here PAH) from two non-overlapping datasets (Figure 4.2). The genetic instrument for RDW comprised genetic variants associated with RDW levels in the RDW GWAS at a study specific significance level ($P \leq 8.31\times10^{-9}$) devised for the genome-wide testing of common, low-frequency and rare variants from whole genome sequencing data (Xu et al., 2014). In the PAH GWAS, 179 variants (inclusive of 12 proxy variants with a minimum r^2 of 0.8) out of the 212 independent ($r^2 < 0.01$) RDW quantitative trait loci (QTL) were available after excluding 13 palindromic variants (A/T or C/G) with intermediate allele frequencies (minor allele frequency > 45%) to ensure that the effects of the variants for the

two traits can be harmonized to the same allele. To obtain the causal estimate, I applied the inverse variance weighted (IVW) (Lawlor et al., 2008) - and weighted median (Bowden et al., 2016) methods. Heterogeneity between the causal estimates from each QTL was estimated using the Cochran's Q test.

In brief, the alleles of RDW QTL were aligned to correspond to an increase in RDW followed by harmonizing PAH effects to the same allele. The causal effect was estimated with IVW and weighted median methods as implemented in the MR-Base software (Hemani et al., 2018). In the conventional IVW method, each RDW QTL provides a causal estimate calculated by simply dividing the variant effect on PAH by the variant effect on RDW (ratio estimate or Wald ratio). These individual causal estimates are then meta-analyzed in a random-effects model weighted by the reciprocal of the standard error of the variant association with PAH, giving more weight to the more precise estimates (detailed description of the causal effect estimation is given in *Methods* 2.7).

In the primary MR analysis, all available genome-wide significant RDW QTL (n = 179) were included. The secondary analysis explicitly tested the role of iron deficiency in the RDW-PAH association using five RDW QTL mapped to genes involved in iron homeostasis (*HFE, TMPRSS6, TFRC* and *TFR2*) from the full set of genome-wide significant RDW QTL (Figure 4.2). All of these five RDW QTL concomitantly increase serum iron, ferritin, and transferrin saturation and decrease transferrin, reflecting systemic iron status (Supplementary Table 8.2.1) and were first reported by an independent GWAS - the Genetics of Iron Status GWAS - as genome-wide significant signals for at least two of the above-mentioned iron status biomarkers (Benyamin et al., 2014). These five RDW QTL are among the signals which explain the highest proportion of variance in the RDW GWAS, highlighting the importance of iron availability in RDW levels.

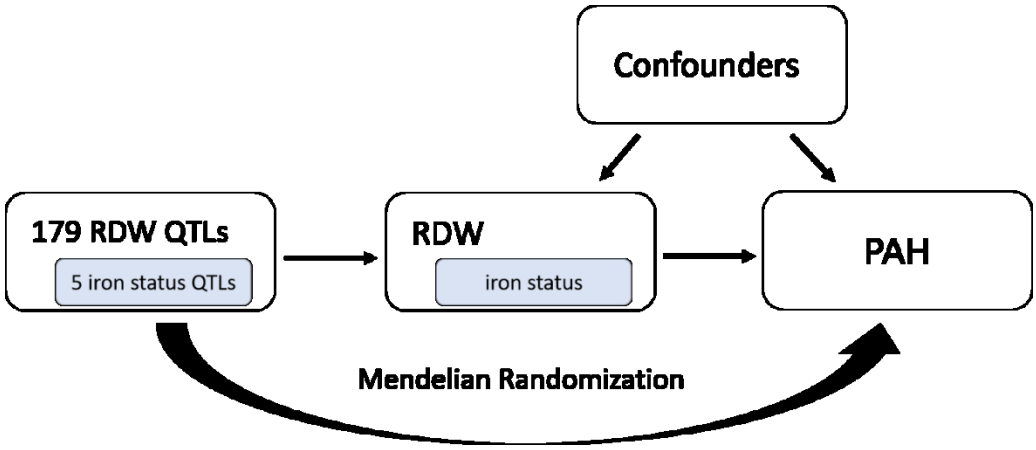


Figure 4.2 Mendelian Randomisation (MR) analyses of red cell distribution width (RDW) and pulmonary arterial hypertension (PAH).
RDW quantitative trait loci (QTL) and their effect estimates were taken from the largest-to-date population based RDW genetic association study (GWAS). Effect estimates for the RDW QTL on pulmonary arterial hypertension (PAH) susceptibility were obtained from the largest-to-date PAH GWAS. The primary MR analysis included all 179 RDW QTL while the secondary MR analysis was restricted to five from the 179 RDW QTL acting via iron status.

4.2.3.3 Estimating the variance explained by the RDW genetic instrument

The validity of the RDW genetic instrument – consisting of 179 RDW QTL - as a proxy for RDW levels was tested in 14,964 common disease controls from the VUMC with genetic data available. A weighted genetic risk score (GRS) for RDW was constructed for each individual by summing the effect alleles multiplied by the effect size at each RDW QTL (detailed description in *Methods* 0). The validity of the genetic instrument was assessed by regressing the first available RDW measurement on the RDW GRS which provided the variance explained (R^2) in RDW levels by the RDW GRS in this cohort.

The R^2 for the RDW GRS alone was calculated by subtracting the R^2 of the model not containing the GRS (null model) from the R^2 of the full model. The models were as follows;

Full model: $\text{RDW} \sim \text{GRS} + \text{sex} + \text{age} + \text{principal components (1st and 2nd)} + \text{batch (study specific)}$

Null model: $\text{RDW} \sim \text{sex} + \text{age} + \text{principal components (1st and 2nd)} + \text{batch (study specific)}$

The R^2 for the combined RDW QTL in the RDW GWAS discovery and replication populations (UK Biobank – UKB and INTERVAL) was calculated from the summary statistics. The individual R^2 estimates were then summed to give the overall variance explained by the RDW instrument:

$$R^2 = \sum_{k=1}^{K} \frac{Nsample_k + 1}{Nsample_k} \times \frac{Z_k^2}{Z_k^2 + Nsample_k} - \frac{1}{Nsample_k}$$

where:
k is an index for each of the variants used in the RDW genetic instrument
Z is the Z-statistic as reported in the RDW GWAS summary statistics
$Nsample$ is the sample size as reported in the RDW GWAS summary statistics.

The standard error for the R^2 estimate was calculated as shown below:

$$SE_{R^2} = \sum_{k=1}^{K} \sqrt{\left(\frac{2}{Nsample_k}\right) \times \left(2 \times R_k^2 + \frac{1}{Nsample_k}\right)}$$

4.2.3.4 Power calculation

Since genetic variants typically explain a small proportion of the variability in the associated trait, MR studies often require large sample sizes to detect a non-zero causal effect. The power to detect a causal association in the current MR analyses was calculated as described by Burgess et al., using the R^2 values estimated in VUMC and also those estimated in the RDW GWAS populations (Burgess, 2014).

4.3 Results

4.3.1 Defining the association of RDW and PAH

Within this observational analysis, each standard unit (1.4%) increase in RDW was associated with 90% higher odds of prevalent PAH after adjusting for the effects of age and sex (odds ratio, OR = 1.90, 95% CI = 1.80 - 2.01). There were no marked between-cohort differences in RDW levels (Figure 4.3).

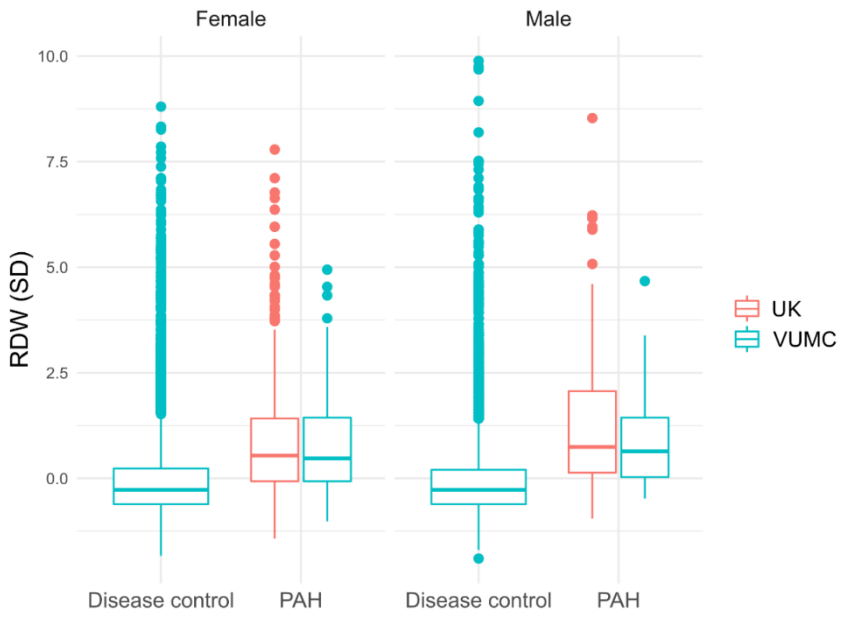


Figure 4.3 RDW levels in the merged cohort of PAH cases (N=642) and the disease control cohort (N=15,889). *The bottom and the top lines of the box indicate the 25th and 75th percentiles, while the centerline indicates the median value. The whiskers extend to 1.5 times the interquartile range from both ends of the box with individual points being more extreme observations. SD = standard deviation. Definition of abbreviations: RDW = red cell distribution width; SD = standard deviation; UK = United Kingdom; VUMC = Vanderbilt University Medical Centre.*

4.3.2 Genetic risk score for RDW

It was estimated that the 179 RDW QTL would explain over 12% of the variability (R^2 = 12.7%, 95% CI = 12.32% -12.99%) in RDW levels in the RDW GWAS population (UK Biobank and INTERVAL) in which they were discovered. The RDW GRS constructed for the VUMC controls explained 2.6% (95% CI = 2.17% - 3.19%) of the variability in the first available RDW measurement. The set of five variants specific to iron status explain an estimated 1.7% (95% CI = 1.62% – 1.87%) of the RDW variability in the UKB and INTERVAL populations whilst they explain 0.7% (95% CI = 0.43% - 0.92%) of the total variability in RDW in the VUMC controls. The RDW GWAS study populations had lower mean RDW

levels than the common disease controls in this study (UKB and INTERVAL = 13.45; VUMC = 13.60) and lower standard deviations (UKB and INTERVAL = 0.82; VUMC = 1.40) which could in part explain differences between the R^2 estimates in these studies.

4.3.3 Statistical power to detect causal effect

With a genetic instrument that explains a relatively high proportion (R^2 = 12%) of the total variation in RDW levels (Figure 4.4, red curve) there is 80% power to detect a causal effect as small as 1.25 (OR). If the true variance explained lies closer to that estimated in the VUMC controls (R^2 = 2.6%, Figure 4.4, green curve), this changes to 1.52. When the variance explained by the genetic instrument is small (R^2=1.7%, Figure 4.4, blue curve), the study is limited, with the current sample size, to an OR of 1.7 or above. However, if the effect of RDW calculated in this observational analysis (OR=1.80-2.01) was causal in nature, either of the two MR analyses, based on R^2 estimates from the RDW GWAS, would have over 80% power to detect an effect of that magnitude. Using the estimates based on the VUMC data, the analysis using all 179 RDW QTL (Figure 4.4, green line) would have sufficient (>80%) power with the current sample size, while the iron-specific model (Figure 4.4, purple line) would be underpowered.

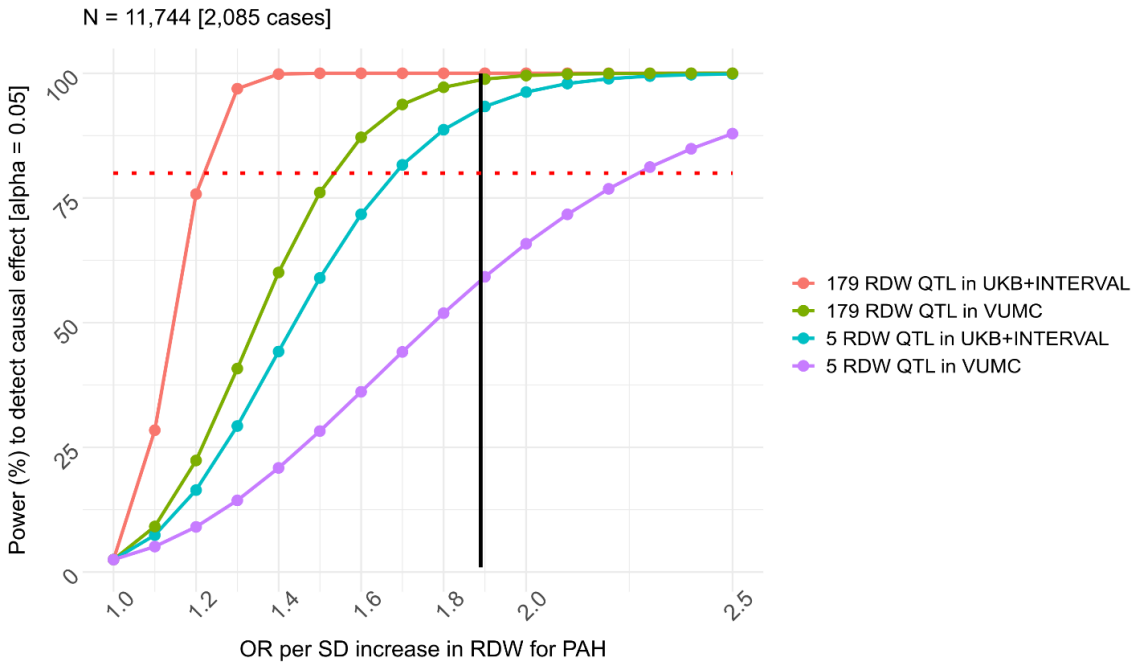


Figure 4.4 Power to detect a causal association in the RDW-PAH Mendelian Randomisation analyses. *X-axis: size of the true underlying causal effect of one standard unit increase in red cell distribution width (RDW) on pulmonary arterial hypertension (PAH) risk. The effect estimate obtained from the observational study is indicated with the vertical black line at OR = 1.90 whilst the red dotted line marks the desired power of 80%. Red curve: Mendelian Randomisation (MR) using all overlapping genome-wide significant variants from the RDW genome-wide association study (GWAS), given the true R^2 = 12% as per estimated in the UKB and INTERVAL cohorts. Green curve: MR using all genome-wide significant RDW quantitative trait loci (QTL) from the RDW GWAS, given the true R_2 = 2.6% as per estimated in our VUMC control cohort. Blue line: MR using five genome-wide significant variants from the RDW GWAS reflecting systemic iron status, given the true R^2 = 1.7% as per estimated in the UKB and INTERVAL cohorts. Purple line: MR using five genome-wide significant QTL from the RDW GWAS reflecting systemic iron status, given the true R^2 = 0.7% as per estimated in our VUMC control cohort.*

4.3.4 RDW-PAH causal relationship

I tested for a causal effect of RDW on development of PAH in the primary MR analysis using 179 RDW QTL and found no significant relationship (IVW ORcausal = 1.07, 95% CI = 0.93 – 1.23, Q p-value = 0.57, Figure 4.5). A secondary MR analysis based on five RDW QTL provided no evidence for a causal effect of iron status on PAH (IVW ORcausal = 1.09, 95% CI = 0.77 – 1.54, Q p-value = 0.91, Figure 4.5). The weighted median method, which is more robust to violations of MR instrument assumptions, yielded similar estimates for both the primary (WM ORcausal = 1.11, 95% CI = 0.89 – 1.38) and the secondary (WM ORcausal = 1.04, 95% CI = 0.68 – 1.59) MR analyses.

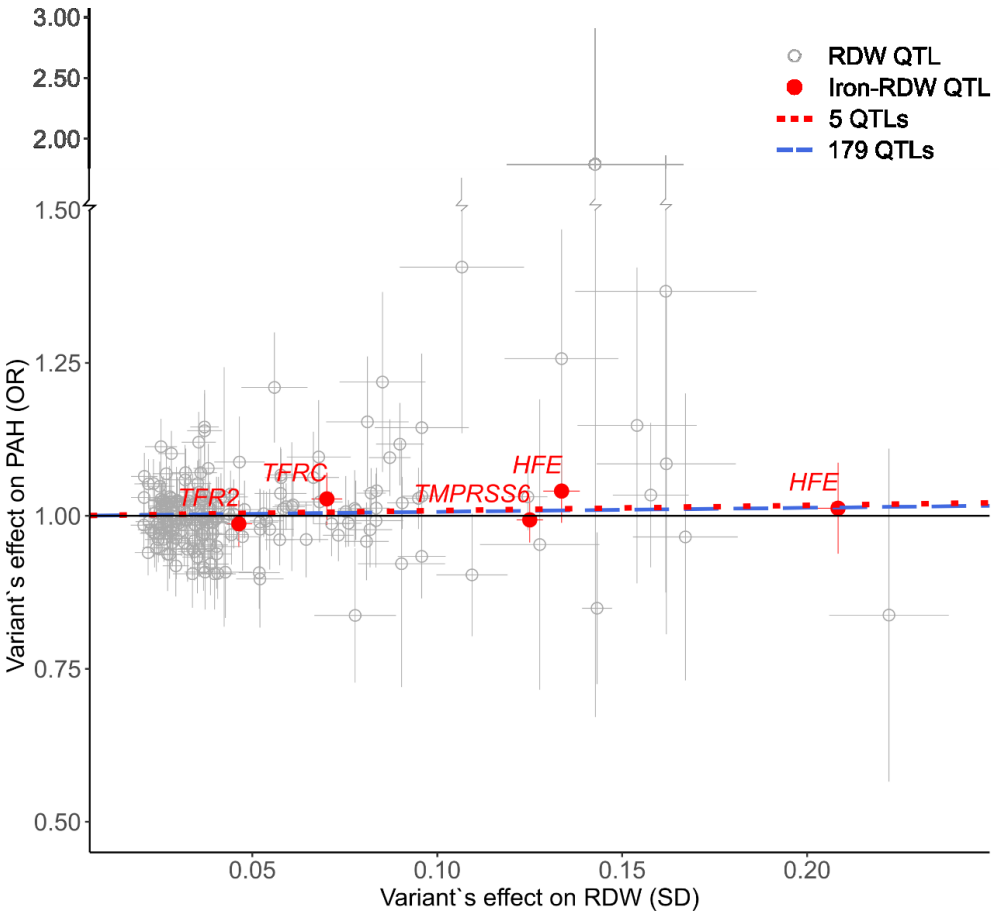


Figure 4.5 RDW-PAH Mendelian Randomisation analyses.
Scatterplot of variant – RDW associations (x-axis) plotted against variant – PAH associations (y-axis) where each dot represents a single RDW QTL. The effect estimates and their standard errors (grey bars) are given in standard units for RDW and in odds ratios for PAH. The solid black line denotes an OR of 1 (no effect) whilst the dashed blue line is the overall causal effect from the IVW regression using all 179 RDW QTL. The five iron-specific RDW QTL (red dots) used as instruments in the secondary Mendelian Randomisation analysis are labelled with their corresponding gene names and the red dotted line denotes the corresponding causal effect. Definition of abbreviations: RDW = red cell distribution width; SD = standard deviation; PAH = pulmonary arterial hypertension; OR = odds ratio; QTL = quantitative trait loci.

If the OR estimated in the primary MR analysis was indicative of the magnitude of a real causal effect, the number of PAH cases needed to detect such a causal effect with a genetic instrument that explains 10% of the variance in RDW would be ~ 20,600 (with the same 1:4.6 ratio of cases:controls as in this study) to achieve 80% power at a false positive rate of 5% (P = 0.05). Heterogeneity was tested between causal effects estimated in the four studies contributing to the PAH GWAS separately (Supplementary Figure 8.2.1 and Supplementary Figure 8.2.2) to assess if differences in the nature of their control populations yielded heterogenous effect estimates. Heterogeneity tests

on the IVW estimates (primary MR: Q = 0.83, df = 3, p = 0.84; secondary MR: Q = 0.55, df = 3, p = 0.91) did not detect considerable variability between the causal effects in the four contributing studies based on a random-effects model.

4.4 Discussion

I applied a two-sample MR approach to test whether the epidemiological relationship between elevated RDW levels, which are associated with iron deficiency, and PAH is causal in nature. The effect of RDW on PAH was estimated in a large sample of cases and common disease controls. By using genetic variants as instruments for RDW, I found no evidence for a causal effect of RDW on PAH of the magnitude suggested by observational studies.

Previous work has shown that iron deficiency is common in PAH and associated with a poor prognosis, reduced exercise capacity and worsening hemodynamics (Ruiter et al., 2011, Soon et al., 2011, Rhodes et al., 2011a). A physiological link has been described in healthy volunteers where iron infusion attenuated the rise in pulmonary artery pressure induced by acute hypoxia (Balanos et al., 2002, Smith et al., 2008) and in rats where chronic iron deficiency results in pulmonary hypertension (Cotroneo et al., 2015). This relationship could be driven by the role of iron in de-stabilizing the hypoxia-inducible factor, thereby deficiency can mimic the hypoxic state (Ivan et al., 2001). This study confirmed the association of raised RDW with PAH using controls from a hospital population and cases from multiple centres (with an effect size 1.90 for one standard unit RDW, 1.4%), supporting a recent hypothesis-free phenome-wide analysis which indicated, among several disease descriptors, PAH as the most strongly associated phenotype with RDW with a similar effect size (OR 2.0, 95% CI = 1.75-2.4 per % increase in RDW) (Thayer et al., 2019). This MR analysis was adequately powered to detect a causal role for RDW with an effect of this magnitude. The fact that no causal effect was detected suggests that at least some of the observed association is secondary to the disease.

The results of this study may appear to be at odds with previous clinical studies of the efficacy of iron supplementation in PAH patients, which have focused on functional capacity rather than disease pathology. An open-label study of twenty patients with idiopathic PAH with iron deficiency reported improved iron status, 6-minute walk distance (6MWD) and quality of life (QoL) two months after a single infusion of 1000 mg ferric carboxymaltose (Viethen et al., 2014). Another open-label study in fifteen iron deficient idiopathic PAH patients reported improvement of iron status, QoL and exercise endurance capacity on cardiopulmonary exercise testing after receiving 1000 mg of

intravenous iron (Ruiter et al., 2015). Neither of the clinical studies were placebo controlled, although Viethen et al. compared their intervention group to a group of matched iron-replete patients who did not receive iron infusion. It remains possible that iron supplementation in PAH could have benefits through mechanisms distinct from those driving the cardiovascular pathology, for example on muscle function (Charles-Edwards et al., 2019).

MR studies using data from large consortia support a causal effect of iron status in other diseases. The genetic instruments (two variants in *HFE* and one variant in *TMPRSS6*) used in these studies were also used in the secondary MR analysis. Gill and colleagues found iron to have a protective effect against coronary artery disease (IVW causal OR = 0.94 per standard deviation (SD) change in serum iron) (Gill et al., 2017) but increase the risk of cardioembolic stroke (IVW causal OR = 1.16 per SD iron) (Gill et al., 2018). The authors suggest the opposing effects of iron status on CAD and stroke might be due to different underlying mechanisms. Pichler et al. (Pichler et al., 2013) have reported that iron protects against the risk of developing Parkinson`s disease (IVW causal OR = 0.88 per SD iron). Iron deficiency is a common risk factor and these causal effect estimates in common diseases are modest; this study was not powered to detect an effect if this is also true of PAH.

In the light of this MR analysis, alternative explanations for the association of RDW and PAH have to be considered; namely, that PAH causes raised RDW (reverse causation, for example reduced oxygen delivery and/or haemolysis related to PAH may stimulate reticulocytosis which would increase RDW) or that PAH and elevated RDW are caused by an independent common mechanism. One such mechanism is chronic inflammation, which is a common feature of PAH (Hassoun, 2014) and leads to intracellular sequestration of iron. Other mechanisms which may modify RDW such as folate or vitamin B12 could be studied for their association with PAH. To directly test whether PAH is causal for raised RDW levels a stronger genetic instrument for PAH is required than currently known common variation identified by PAH GWAS.

An important strength of this study lies in the sample size available with phenotype and genetic data achieved through extensive collaboration. Given the rarity of PAH, data had to be pooled from several centres to allow the investigation of common genetic variation in PAH and to test causal relationships. This study also has some limitations. The control population for the observational study was not specifically selected to represent a population at risk of developing PAH. An example of a population at risk of PAH are relatives of patients with pathogenic *BMPR2* mutations (and other rare pathogenic variants). Given the reduced penetrance of familial/heritable PAH (~42% in women and ~14% in men carrying known mutations) (Machado et al., 2001), following up the families of

affected individuals, especially relatives harbouring mutations associated with PAH, would be an invaluable source to identify environmental triggers of PAH.

Despite my efforts to exclude controls with conditions that likely affect RDW and to use the first available RDW measurement, genetic effects of RDW are expected to differ between individuals with common diseases and a cohort of healthy volunteers. The R^2 of a genetic variant describes the variance explained in the phenotype in a given population at a given time. Therefore, there can be no single population parameter that applies to multiple populations or the same population at multiple timepoints. Although the R^2 estimated in the RDW GWAS is likely upwardly biased (since it contains the discovery as well as the replication samples), it might reflect better the extent to which genetic variation influences long-term RDW levels in disease-free populations than the R^2 estimate from the VUMC disease controls, who may be expected to have more variable RDW levels due to comorbidities. This highlights a potential challenge in estimating power for MR studies. I estimated that upwards of 20,000 PAH patients would be required to detect any likely causal effect of iron deficiency.

It is important to note that in these MR analyses, causal effects were estimated using the results of the PAH GWAS (Rhodes et al., 2018). Four independent studies contributed to the overall results of the PAH GWAS comparing allele frequencies of their PAH cases to control cohorts selected according to different criteria. The NIHRBR study used a control population with a mixture of rare diseases while the other major contributing study from the United States used mixed common disease controls. The two smaller contributing studies used population-based controls; a preferable design for estimating the effects of surrogate genetic variants for common risk factors. Selection bias can affect the estimates of the instrumental genetic variants on disease susceptibility. This is especially true if the control cohort is enriched for conditions also affected by the risk factor of interest.

4.5 Conclusions

There is strong observational evidence for an association between elevated RDW, a surrogate for iron deficiency, and PAH. However, this Mendelian randomisation analysis does not indicate that RDW is causally linked to disease development. The study was powered to detect a causal effect similar in size to that observed. A more modest causal effect remains possible but would require a significantly larger study population to detect. Extending international collaborations and careful follow up of populations at risk will allow increasingly sophisticated study designs to investigate causal relationships, shared underlying mechanisms with other conditions and overall genetic susceptibility in PAH.

5 Supplementation of iron in idiopathic pulmonary hypertension: a randomized controlled crossover trial

5.1 Introduction

Randomized controlled trails (RCTs) represent the 'gold standard' for testing causality or treatment efficacy. Randomization of a sufficient number of study subjects will balance out any confounders between treatment groups allowing the observed differences in outcome between the groups to be attributed to the treatment received. Randomised intervention trials (RIT) can be used to assess the role of putative causal factors in disease development and identify targets for prevention strategies in a study design that reduces biases that conventional observational trials suffer from. However, RITs require large sample sizes (depending on disease frequency in the population) and long follow-up periods which make this trial design very expensive. Effects of dietary intervention on disease risk are often investigated with RITs. An example of this is the Women`s Health Initiative study that recruited and randomized a population of 48,835 women to an intervention and a control dietary group and followed them up for a median of 8.1 years to investigate the effects of a low-fat diet on developing common diseases such as diabetes (Howard et al., 2018), cardiovascular diseases (Howard et al., 2006) and certain types of cancers (Prentice et al., 2006, Thomson et al., 2014). RITs can be affected by selection bias, as willing participants and those sticking to the treatment regime tend to be more interested in health and are generally healthier; therefore they may not be completely representative of their underlying populations. Selection bias can be a factor in all clinical trials, but its impact on the generalizability of treatment effect is believed to be minimal in RITs and not comparable to the magnitude of selection bias that affect observational studies, where participants select their own treatments. Another caveat in RITs is compliance or in other words adherence to the treatment regimen, which being a behavioral trait is shown to be predicted by factors such as socioeconomic status, ethnicity and marital status (Charles et al., 2003, Van Horn et al., 1997, Brunner et al., 2009), and should be considered and addressed, for instance, by increasing participant contact (Brunner et al., 2009).

Investigating causal relationships with RITs is often not feasible, especially in rare diseases like PAH. Even though the achievable sample size for rare diseases excludes most epidemiological study designs apart from the case-control design for meaningful statistical analyses, the growing availability of genotyping and sequencing as well as rapid improvements in methodology focused on

utilization of summary-level data now enable strategies to investigate these rare diseases in much larger sample sizes to answer questions about genetic susceptibility, causal relationships and shared etiologies with other diseases.

As discussed in previous chapters, this thesis focuses on iron deficiency as a causal factor in IPAH. Open-label studies without randomization or placebo control, summarized in Table 1.3 in the Introduction section, suggested that iron supplementation in IPAH patients with iron deficiency improved patient quality of life and exercise capacity (Viethen et al., 2014, Ruiter et al., 2015). However, it is unclear whether iron deficiency influences the natural history of IPAH directly or is an epiphenomenon. Chapter four of this thesis applied the technique of Mendelian Randomisation (MR), a form of instrumental variable analysis, to test for causality between iron status and IPAH using genetic variants for red cell distribution width and systemic iron status as instruments. There was no evidence of a causal relationship between iron deficiency and IPAH in this MR analysis powered to detect a causal association with an effect size of 1.25 (odds ratio) or greater using the largest to-date genome-wide association study of IPAH susceptibility (Ulrich et al., 2019). Possible roles for iron deficiency in IPAH include a causal role (MR analysis suggesting only with a modest effect), where iron deficiency might or might not influence IPAH once it developed, or iron deficiency simply being an epiphenomenon. However, even if iron deficiency did not contribute to development of IPAH, iron replacement therapy in IPAH patients with iron deficiency could improve metrics of severity of the disease.

In this chapter, I present the results from the first RCT of iron supplementation in IPAH patients with iron deficiency, designed to assess the clinical benefit and safety of intravenous iron replacement therapy in these patients. The trial is registered on ClinicalTrials.gov (Study Identifier: NCT01447628), was funded by the British Heart Foundation and sponsored by the Imperial College Academic Health Science Centre.

5.2 Methods

A phase II proof-of-concept study was designed to assess the safety and potential therapeutic benefit of intravenous ferric carboxymaltose in iron deficient IPAH patients. The trial took place across four specialist centers in the United Kingdom and one centre in Germany (Hammersmith Hospital, London; Papworth Hospital, Cambridge; Royal Hallamshire Hospital, Sheffield; and Giessen in Germany) where 42 patients were recruited in total and one centre in Beijing, China where 17 patients were enrolled. Patients were randomized in a 1:1 ratio in a 24-week, double-blind, placebo-controlled, crossover study. Iron was administered intravenously as a single 1000 mg (15 mg/kg if patient weight < 66.7 kg) infusion of Ferinject (ferric carboxymaltose, Europe) or 20mg/kg CosmoFer (iron dextran, China). Intravenous instead of oral iron supplementation was chosen in this trial to circumvent differences in dietary absorption of iron by abnormally high levels of hepcidin reported in IPAH (Soon et al., 2011, Rhodes et al., 2011a).

The primary outcome of choice in many PAH RCTs (Barst et al., 1996, Galiè et al., 2005, Rubin et al., 2002) to date has been 6-minute walk distance (6MWD), a noninvasive submaximal exercise test that is easy to perform, with prognostic significance in PAH (Miyamoto et al., 2000). However, 6MWD measurements exhibit large within-subject variability, not suiting trials with relatively small number of patients and which may be less informative in patients with milder disease (6MWD > 450 meters) (Frost et al., 2005). Outcome measures in this study were selected to have relatively low variability allowing the detection of an effect in a trial of ~60 patients. The initial primary outcome in this trial was resting pulmonary vascular resistance (PVR) measured at right heart catheterization which is an accurate and reproducible measure but also invasive. The primary endpoint was changed from PVR partway through the study in European centres after an interim analysis of a subset of patients indicated that the total sample size of the trial was too small for the detection of any potential effect of treatment on PVR, and hence the invasive catheterization procedure was removed from the protocol. While PVR remained the primary outcome in the China study, in Europe it was changed to endurance time at the end of the endurance bicycle cardiopulmonary exercise testing (CPET) at 80% peak work rate determined from the incremental exercise test. Secondary outcomes included exercise capacity and quality of life measures. Patients in the study received the iron treatment whilst being treated with approved PAH therapies, as recruiting treatment-näive patients would not be ethical.

Exercise capacity measured by cardiopulmonary exercise testing (CPET) is a reproducible and objective measurement designed to quantitate the ability of the cardiovascular system to deliver

oxygen to the exercising muscles and the ability of the muscles to extract oxygen. Uptake of oxygen (VO_2) during exercise is calculated using Fick`s equation:

$$VO_2 = (Stroke\ vol.\ \times Heart\ rate) \times (Arterial\ O_2 - Venous\ O_2)$$

,where stroke volume multiplied by the heart rate equals cardiac output, the amount of blood that is pumped from the two ventricles per unit time. Stroke volume defines the amount of blood per beat whilst heart rate gives number of beats per minute. Arterial and venous oxygen content reflect the level of oxygen saturation of the blood being pumped to the muscles and that returning to the heart. Maximal VO_2 reflects a person`s maximal ability to take up and utilize oxygen during exercise, or in other words a person`s functional aerobic capacity:

$$VO_2\max = (Stroke\ vol.\max\ \times Heart\ rate\ max) \times (Arterial\ O_2\max - Venous\ O_2\max)$$

Maximal VO_2 plateaus in healthy people near maximal exercise. In clinical conditions, this plateau might not be achieved and the peak VO_2 measurement is used instead. Abnormally low peak VO_2 indicates exercise intolerance. Peak O_2 pulse is calculated by dividing the peak VO_2 by the peak heart rate during exercise and reflects the amount of oxygen consumed per heartbeat. VO_2 at metabolic (anaerobic) threshold provides a measurement for the amount of oxygen required at the stage in exercise where oxygen supply cannot keep up with the increasing metabolic requirements of the muscles. The relationship between minute ventilation and carbon dioxide output called the VE/VCO_2 slope measures the amount of air breathed in to eliminate unit CO_2 and reflects ventilatory efficiency (Albouaini et al., 2007). CPET is a correlate of functional class (Sun et al., 2001) and has been successfully used as primary outcome by two clinical trials in PAH (Barst et al., 2003, Barst et al., 2004).

5.2.1 Participants

Idiopathic, heritable or anorexigen-induced PAH cases diagnosed by right heart catheterization with iron deficiency were eligible to participate in the study. Patients were identified and recruited on the basis of suspected iron deficiency, supported by measurements from the clinical laboratories at their Pulmonary Hypertension centres, which were later replicated for confirmation by the central study laboratory at Imperial College. Iron deficiency at recruitment was identified by laboratory biomarkers including ferritin, transferrin saturation (TSAT), serum iron and soluble transferrin receptors (sTfR) where available. Participants met either one of four stringent biomarker cut-offs: ferritin < 37 µg/L or TSAT < 16.4% or serum iron < 10.3 µmol/L or sTfR > 28.1 nmol/L. These cut-offs are based on previous work (Howard et al., 2013) demonstrating that 83% of IPAH patients with an

sTfR > 28.1 nmol/L could be identified correctly using the above cut-offs for the other three, more widely available biomarkers.

Eligible patients were between 16 and 75 years old and able to provide written consent, were stable on their PAH treatment regime for at least one month and had a minimum 6-minute walking distance of 50 meters at entry. Female participants of child-bearing potential agreed to using one of prespecified contraceptive methods. Exclusion criteria included renal or liver disease, severe anemia (haemoglobin <10 g/dL), haemoglobinopathy, hypophosphatemia (< 0.65 mmol/L), PAH-related hospital admission one month prior to recruitment, evidence of left ventricular or lung disease, pregnancy or lactation, acute or chronic infection or inflammation asthma, eczema, known hepatitis B or C infection, human immunodeficiency virus infection, disturbances in iron utilization, significant blood loss three months prior to recruitment and inability to perform Cardiopulmonary Exercise Test.

Three patients in the Europe study were not randomized and were withdrawn from the study before treatment and one person who received placebo in both periods due to an administrative error did not contribute data to the analyses. Participant characteristics in the Europe and China studies are shown in Table 5.1.

Characteristic	Iron - Placebo		Placebo - Iron	
	Europe	China	Europe	China
Demographic variables				
Female:male, N (female %)	14:5 (74%)	11:1 (92%)	14:5 (74%)	4:1 (80%)
Age, yr	50.0 (15.5)	35.0 (11.0)	49.0 (13.5)	29.0 (6.0)
BMI, kg/m^2	27.9 (6.3)	24.9 (4.5)	27.2 (5.8)	21.4 (2.2)
Ethnicity, N (%)				
Chinese	-	12 (100%)	-	5 (100%)
white	18 (95%)	-	18 (95%)	-
South Asian	1 (5%)	-	1 (5%)	-
Iron status				
Iron, umol/L	12.8 (10.3)	8.5 (6.5)	12.1 (10.4)	10.0 (6.0)
Ferritin, ug/L	17.0 (13.0)	11.0 (10.0)	18.0 (28.0)	11.0 (1.0)
sTfR, nmol/L	41.7 (22.8)	-	37.9 (21.3)	-
Hepcidin, ng/mL	3.5 (6.7)	-	1.3 (3.0)	-
Haemoglobin, g/dL	13.5 (2.8)	14.4 (2.4)	13.7 (1.6)	12.3 (1.1)
Functional status				
NYHA functional class, N (%)				
I	-	1 (8%)	1 (5%)	-
II	11 (58%)	9 (75%)	9 (47%)	3 (60%)
III	8 (42%)	2 (17%)	9 (48%)	2 (40%)
6MWD, m	427.0 (65.0)	427.5 (102.0)	450.0 (71.5)	480.0 (46.0)
Dyspnea score (Borg index)	2.5 (2.8)	2.5 (2.5)	2.5 (2.0)	0.5 (0.5)
NTproBNP, ng/L	72.9 (218.2)	-	222.5 (396.9)	-
Hemodynamic variables				
Systolic blood pressure, mmHg	112.0 (15.5)	104.0 (12.5)	110.0 (17.5)	108.0 (8.0)
Diastolic blood pressure, mmHg	67.0 (13.0)	70.0 (10.3)	68.0 (10.5)	70.0 (6.0)
PVR, Wood units	7.3 (2.4)	9.1 (3.5)	5.9 (4.1)	9.5 (1.0)
mPAP, mmHg	45.5 (14.3)	49.0 (16.0)	50.0 (4.0)	49.0 (16.0)
mRAP, mmHg	8.5 (7.8)	3.5 (5.3)	10.0 (3.0)	1.0 (7.0)
mPCWP, mmHg	11.0 (3.3)	6.5 (3.5)	12.5 (4.0)	11.0 (5.0)
Cardiac index, L/min/m^2	2.6 (0.6)	2.7 (0.4)	2.5 (1.7)	2.4 (0.7)
Treatment, N (%)				
PDE-5 inhibitors	15 (79%)	9 (75%)	17 (89%)	4 (80%)
Endothelin receptor antagonists	15 (79%)	4 (33%)	13 (68%)	3 (60%))
Prostacyclins	6 (32%)	1 (8%)	4 (21%)	-
Calcium channel blockers	1 (5%)	-	1 (5%)	-
Warfarin	13 (68%)	7 (58%)	12 (63%)	2 (40%)
Diuretics	11 (58%)	12 (100%)	12 (63%)	5 (100%)
Digoxin	1 (5%)	2 (17%)	3 (16%)	-
Sodium ferulate	-	12 (100%)	-	4 (80%)

Table 5.1 Baseline patient characteristics by randomization in Europe and China.
All continuous variables are median (interquartile range). Definition of abbreviations: BMI = body mass index; sTfR = soluble transferrin receptor; 6MWD = six-minute walk distance; NTproBNP = N-terminal -pro hormone brain natriuretic peptide; PVR = pulmonary vascular resistance; mPAP = mean pulmonary arterial pressure; mRAP = mean right atrial pressure; mPCWP = mean pulmonary capillary wedge pressure; PDE-5= phosphodiesterase type 5.

5.2.2 Study design

Participants were randomly assigned to arm A or B, where each subject assigned to arm A received iron infusion at baseline followed by placebo at week 12. In arm B, subjects received placebo at baseline followed by iron treatment at week 12. Assessments for all outcomes (primary and secondary) were carried out at baseline, 12 weeks and 24 weeks (Figure 5.1).

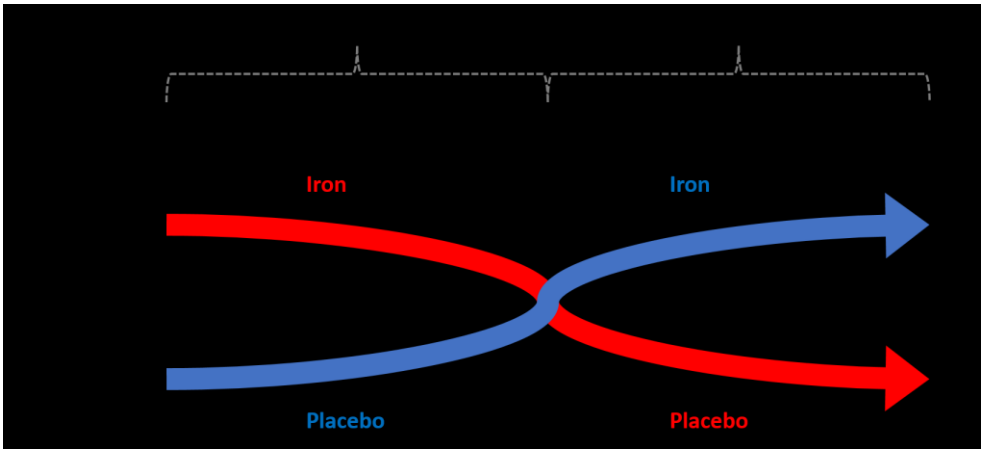


Figure 5.1 Study design of the Ferinject trial.
Arm A received iron (active treatment) first followed by placebo at the 12-week cross-over. Arm B received placebo at baseline followed by iron at 12 weeks.

The main analyses to test the effect of treatment on the primary and secondary outcomes were carried out by a trial statistician – Dr Les Huson – as stated in the statistical analysis plan available from ClinicalTrials.gov (Study Identifier: NCT01447628).

The primary outcome in the Europe study was changed after preliminary data analysis based on a subset of patients (N=17) at the end of period 1 to endurance time at the end of the endurance bicycle cardiopulmonary exercise testing at 80% peak work rate determined from the incremental exercise test (hereon referred to as endurance time).

5.2.3 Statistical analyses

The results from the Europe and the China studies were first analyzed separately followed by the meta-analysis of effect estimates for the overlapping efficacy endpoints. All endpoints were analyzed using linear mixed models appropriate for crossover design (Putt and Chinchilli, 1999).

Mixed models include both fixed and random effects. In this analysis, sequence of administration (arm A or B), period (first 12 weeks or second 12 weeks of the study) and treatment (active or placebo) were defined as fixed effects, whilst subject was included in the model as the random

effect. Baseline outcome measurements were found to be statistically significant covariates and were adjusted for in all models.

Additional analyses in the Europe study - not outlined in the statistical analysis plan of the trial - were conducted to test if the effect of iron supplementation differed with severity of baseline iron deficiency or was dependent on improvement in iron status on sixteen outcomes reflecting disease severity including CPET measures, N-terminal -pro hormone brain natriuretic peptide (NTproBNP), quality of life (QoL) score, 6 minute walk distance (6MWD) and Borg dyspnea score. To do this, the two arms of the study were aligned at the time of receiving treatment (iron infusion) which was at baseline for arm A and at 12 weeks for arm B. Patients in the two arms of the study were analyzed together as one group. These analyses will hereon be referred to as the 'treatment-only' analyses and baseline will refer to the timepoint when patients received iron infusion. Change in outcome from iron infusion (baseline) to 12 weeks after iron infusion was calculated by subtracting the baseline outcome measure from the outcome measured at 12 weeks post-infusion. Improvement in iron status was assessed by the reduction in the log-transformed ratio of sTfR and ferritin (sTfR/ferritin ratio). Change in outcome was modeled as a function of the change in sTfR/ferritin ratio, baseline sTfR/ferritin ratio, baseline outcome, age and sex:

$$\Delta outcome \sim \alpha + \Delta sTfR/ferritin + sTfR/ferritin_{Base} + outcome_{Base} + Age + Sex + \varepsilon$$

, where α = intercept and ε = residual variance.

Coefficients with a P-value below 0.00312, adjusted for sixteen tests were considered significant. Models showing suggestive (P-value < 0.05) associations between Δoutcome and ΔsTfR/ferritin were inspected for meeting modeling assumptions. Residuals plotted against fitted values of the regression model allows the detection of outliers and the presence of heteroscedasticity (unequal variance of errors) which may have the effect of giving too much weight to observations with large error variance when estimating regression coefficients. In homoscedasticity, residuals versus fitted values should be scattered around a horizontal line, with a roughly constant variance.

Treatment-only analyses were conducted on the whole cohort in the Europe study (N = 38) and on a subset of those patients (N = 24) with severe iron deficiency at baseline defined by the combined cut-off of ferritin (< 100 µg/L), TSAT (< 20%) and sTfR levels (> 28.1 µmol/L). The ferritin and TSAT thresholds represent standard thresholds for iron deficiency, less conservative than those used in the study description (see 'Participants' above) as they were used in combination.

5.3 Results

5.3.1 Baseline iron deficiency of participants

In the Europe study, 37 study participants met at least one of iron, ferritin or TSAT thresholds for iron deficiency (met: one=14, two=12, three=11), while 28 patients reached the sTfR threshold as well. All 17 participants in the China study were iron deficient at baseline defined by at least one of iron, ferritin or TSAT levels where 10 (59%) patients met all three biomarker cut-offs. sTfR levels were not measured in China. The two arms of the trial were similar with respect to the baseline clinical and laboratory characteristics and use of medications at the time of enrollment (Table 5.1).

There were no serious adverse events or allergic reactions reported during the trial and iron treatment was well tolerated in both studies indicated by phosphate levels which remained in the normal range (Figure 5.2).

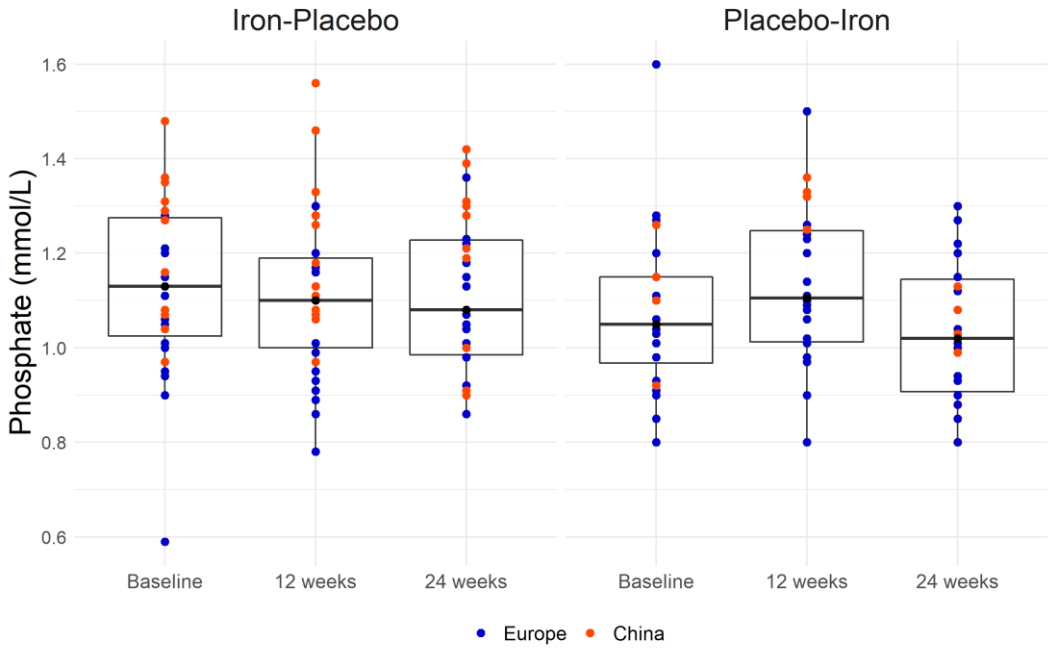


Figure 5.2 Phosphate levels in the Ferinject trial by randomization.
Normal phosphate levels for adults: 0.8-1.5 mmol/L. The bottom and the top lines of the box indicate the 25th and 75th percentiles, while the centre line indicates the median value. The whiskers extend to 1.5 times the interquartile range from both ends of the box with individual points being more extreme observations.

5.3.2 Primary and secondary outcomes by treatment

All primary and secondary endpoints measured at baseline, week 12 and week 24 of the trial are tabulated in Supplementary Table 8.3.1 and 8.3.2 for Europe and China, respectively. The main results of iron treatment in IPAH patients with iron deficiency on the primary and secondary

outcomes are presented in Table 5.2. Pulmonary vascular resistance (PVR) was initially the primary outcome in both studies, however after an interim analysis in the Europe study, it was changed to endurance time in the Europe study only as results indicated that any effect on PVR would be too small to detect with the study size. There was no significant effect of iron treatment observed for either endurance time in the Europe study ($Effect_{Iron}$ = -11.26 seconds, 95% CI = -70.28-47.76), or for PVR ($Effect_{Iron}$ = 1.57 Wood units, 95% CI = -0.22-3.36) (Table 5.2 and Figure 5.3/A and B).

Twenty-one out of 38 (55%) patients no longer met any of the iron deficiency criteria used for trial entry (see 5.2.1 Participants) 12 weeks after receiving iron infusion in the Europe study. In the China study, 13 out of 17 (76%) patients experienced iron status improvement to a degree that they did not meet any of the trial entry criteria 12 weeks after iron infusion.

Iron status significantly improved 12 weeks after iron infusion in both studies shown by the increase in ferritin levels (Europe: $Effect_{Iron}$ = 52.59 µg/L, 95% CI = 27.17-78.01, China: $Effect_{Iron}$ = 78.52 µg/L, 95% CI = 52.53-104.51, meta-analyzed P-value < 10^{-4}) in both studies and the decrease in sTfR in the Europe study ($Effect_{Iron}$ = -8.28 nmol/L, 95% CI = 4.63-11.93) (Table 5.2 and Figure 5.3/C and D). There was no significant difference in ferritin levels in the Iron-Placebo arm between week 12 and week 24 ($t_{(df=60)}$ = 0.84, P-value = 0.41), suggesting the effect of iron infusion lasted beyond 12 weeks. This observation was robust to excluding four individuals in the Europe study whose ferritin levels increased from 12 weeks to 24 weeks after receiving placebo ($t_{(df=49)}$ = 1.54, P-value = 0.13). TSAT levels exhibited a trend of moderate increase with iron treatment, but the difference between the placebo and active groups did not reach significance in either study (Europe: $Effect_{Iron}$ = 3.76 %, 95% CI = -0.04-7.56, China: $Effect_{Iron}$ = 5.54 %, 95% CI = 0.23-10.85). Serum iron levels did not increase significantly in either study (Europe: $Effect_{Iron}$ = 1.48 µmol/L, 95% CI = -0.76-3.72, China: $Effect_{Iron}$ = 2.46 µmol/L, 95% CI = -1.12-6.04). No significant change was observed in any other secondary endpoint by iron treatment, including those reported by intervention studies (Viethen et al., 2014, Ruiter et al., 2015) to increase in PAH patients after iron treatment such as 6MWD and QoL score (Table 5.2).

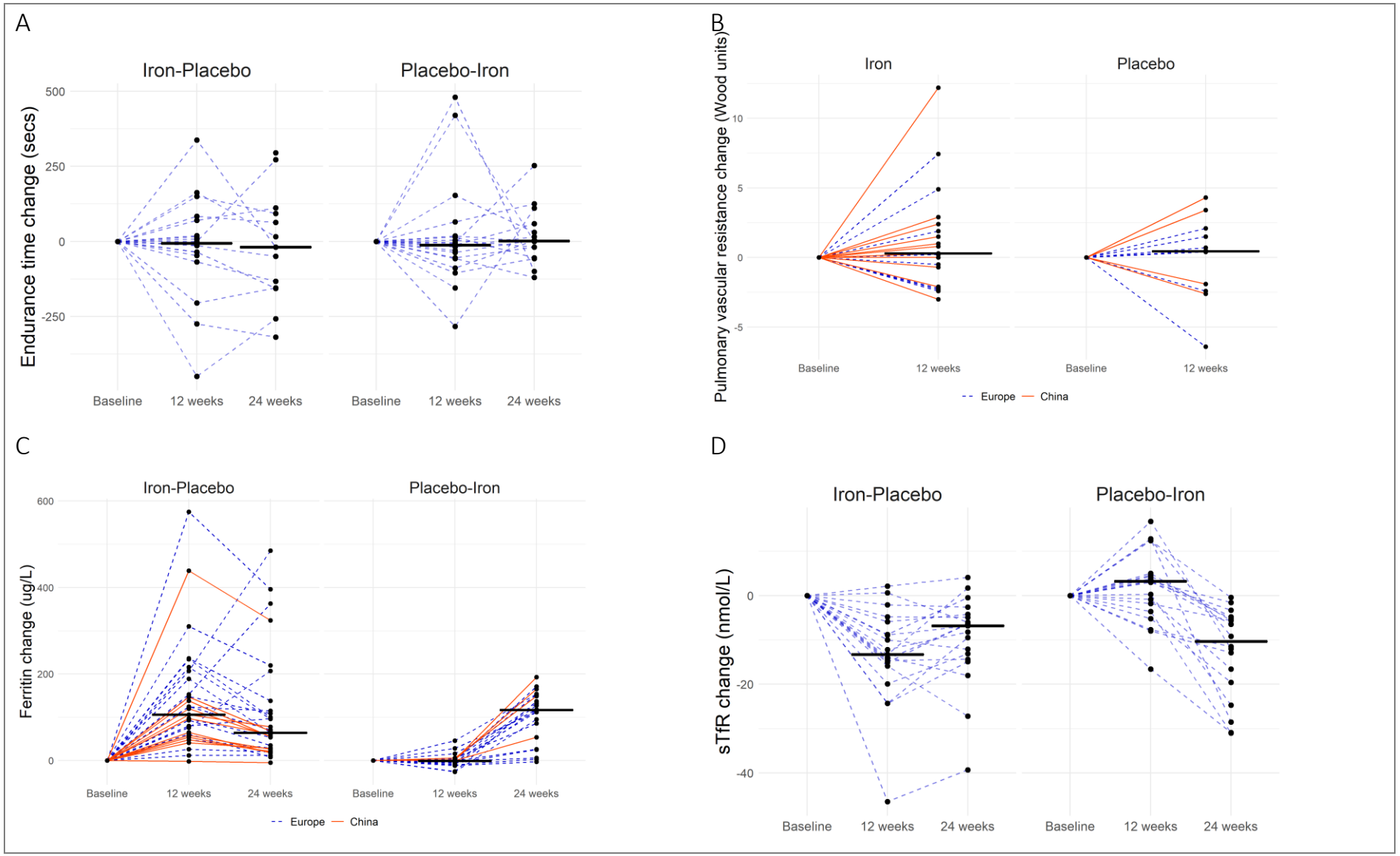


Figure 5.3 Changes in primary outcomes and iron status in the Ferinject trial.
Baseline measurement starts at zero, horizontal black lines are for the median change at 12 and 24 weeks. ***A:*** *Primary outcome - endurance time change in Europe.* ***B:*** *Primary outcome - pulmonary vascular resistance change (PVR) in China. PVR measurements were discontinued in Europe after period 1 before the crossover.* ***C:*** *Ferritin change in Europe and China.* ***D:*** *Soluble transferrin receptors (sTfR) in Europe.*

Endpoint	Europe					China					Meta
	Baseline mean, all patients	Placebo group after 12 weeks, LSM	Iron group after 12 weeks, LSM	Iron - Placebo 12 week effect	P-value	Baseline mean, all patients	Placebo group after 12 weeks, LSM	Iron group after 12 weeks, LSM	Iron - Placebo 12 week effect	P-value	P-value
Primary outcomes											
Endurance time, sec	292.51	312.2 (264.82-359.58)	300.94 (253.96-347.92)	-11.26	0.7109	-	-	-	-	-	-
Pulmonary Vascular Resistance, Wood unit	6.64	5.44 (3.61-7.27)	7.44 (5.71-9.17)	2	0.1143	10.24	10.03 (6.25-13.81)	11.6 (9.16-14.04)	1.57	0.471	0.2111
Secondary outcomes											
Iron indices											
Iron, umol/L	14.33	15.56 (13.94-17.17)	17.03 (15.42-18.64)	1.48	0.205	8.88	13.31 (9.83-16.79)	15.77 (12.29-19.25)	2.46	0.1993	0.1715
Ferritin, ug/L	26.28	94.48 (66.08-122.88)	147.07 (118.63-175.51)	52.59	0.0003	18.88	65.52 (44.27-86.77)	144.04 (122.79-165.29)	78.52	<0.0001	<0.0001
sTfR, nmol/L	42.75	37.01 (34.02-40)	28.73 (25.73-31.72)	-8.28	<0.0001	-	-	-	-	-	-
TSAT, %	17.07	22.17 (19.21-25.12)	25.92 (22.96-28.89)	3.76	0.061	12.88	20.64 (15.27-26)	26.18 (20.81-31.54)	5.54	0.0641	-
Functional status											
6MWD, m	421.79	424.38 (410.06-438.7)	425.89 (411.57-440.2)	1.5	0.8093	461.82	481.05 (443.73-518.37)	486.27 (448.96-523.59)	5.23	0.63	0.8533
Dyspnea score (Borg index)	3.53	3.67 (3.18-4.17)	3.25 (2.75-3.75)	-0.42	0.0725	1.65	1.3 (0.66-1.93)	1.36 (0.72-2)	0.06	0.8572	0.2348
NTproBNP, ng/L	394.94	422.95 (297.52-548.39)	312.04 (186.47-437.62)	-110.91	0.1115	-	-	-	-	-	-
Symptom score	8.41	7.96 (7.02-8.89)	7.94 (6.99-8.89)	-0.01	0.979	-	-	-	-	-	-
QoL score	7.44	7.04 (5.91-8.16)	7.54 (6.38-8.69)	0.5	0.2219	-	-	-	-	-	-
Activity score	8.47	9.2 (8.25-10.16)	9.1 (8.14-10.07)	-0.1	0.7702	-	-	-	-	-	-

Endpoint	Europe					China					Meta
	Baseline mean, all patients	Placebo group after 12 weeks,	Iron group after 12 weeks, LSM	Iron - Placebo 12 week effect	P-value	Baseline mean, all patients	Placebo group after 12 weeks,	Iron group after 12 weeks, LSM	Iron - Placebo 12 week effect	P-value	P-value
Hemodynamics (completed in N=36)											
mPAP, mmHg	-	47.65 (41.33-53.97)	46.7 (40.71-52.69)	-0.95	0.8212	-	47.73 (41.51-53.96)	50.05 (46.04-54.07)	2.32	0.5241	0.7932
mRAP, mmHg	-	9.19 (7.17-11.21)	7.32 (5.4-9.23)	-1.87	0.1777	-	5.84 (3.21-8.48)	6.23 (4.55-7.91)	0.39	0.7889	0.4156
mPCWP, mmHg	-	13.1 (10.34-15.87)	10.98 (8.36-13.59)	-2.13	0.2654	-	8.99 (7.14-10.84)	6.88 (5.72-8.05)	-2.11	0.0643	0.0865
Cardiac output, L/min	-	6.15 (5.31-6.99)	5.65 (4.86-6.44)	-0.5	0.3966	-	4.46 (3.44-5.48)	4.02 (3.37-4.68)	-0.43	0.4595	0.4925
PA Sat, %	-	70.84 (67.73-73.94)	71.91 (69-74.82)	1.07	0.6415	-	69.74 (65.39-74.08)	66.01 (63.2-68.81)	-3.73	0.1516	0.3239
Stroke volume, mL	-	88.94 (75.86-102.02)	89.73 (77.41-102.05)	0.79	0.8666	-	54.19 (45.22-63.17)	55.36 (49.57-61.15)	1.16	0.8104	0.9504
ICPET											
Peak VO2, L/min	1.22	1.14 (1.08-1.21)	1.16 (1.1-1.22)	0.02	0.6424	0.94	0.91 (0.84-0.97)	0.93 (0.87-1)	0.03	0.4695	0.6631
Peak VO2, mL/min/kg	15.51	14.94 (14.2-15.68)	14.84 (14.1-15.58)	-0.1	0.7625	15.27	13.15 (11.21-15.09)	14.76 (12.83-16.68)	1.61	0.2262	0.4756
VO2 at metabolic threshold, L/min	0.84	0.77 (0.73-0.82)	0.77 (0.73-0.82)	0	0.9967	0.67	0.68 (0.59-0.77)	0.67 (0.58-0.76)	-0.01	0.9144	0.9959
VO2 at metabolic threshold, mL/min/kg	10.45	10.02 (9.52-10.53)	9.96 (9.45-10.46)	-0.06	0.7711	10.82	10.81 (9.77-11.86)	11.02 (9.98-12.06)	0.21	0.7806	0.9074
VO2/WR slope	8.8	8.07 (7.46-8.68)	8.11 (7.5-8.71)	0.04	0.8688	6.39	5.96 (5.15-6.77)	5.96 (5.15-6.77)	0	0.9999	0.991
VE/VCO2 slope	41.17	42.32 (40.58-44.07)	41.18 (39.43-42.93)	-1.15	0.1788	34.86	35.49 (31.21-39.78)	34.99 (30.66-39.31)	-0.5	0.7795	0.414
Peak O2 pulse, L/beat	8.92	8.62 (8.17-9.08)	8.82 (8.36-9.28)	0.2	0.4997	6.56	6.29 (5.79-6.79)	6.52 (6.02-7.01)	0.23	0.4298	0.5451
ECPET (completed in N=38)											
VO2 at 3mins at 80% peak WR, L/min	1	1.26 (0.96-1.56)	1.08 (0.77-1.39)	-0.18	0.389	-	-	-	-	-	-
VO2 at 3mins at 80% peak WR, mL/min/kg	12.89	15.8 (12.39-19.22)	14.15 (10.63-17.67)	-1.66	0.4864	-	-	-	-	-	-

VO2 at 80% peak WR, L/min	1.19	1.13 (1.07-1.2)	1.18 (1.11-1.24)	0.04	0.2433	-	-	-	-	-	-
VO2 at 80% peak WR, mL/min/kg	15.26	14.61 (13.78-15.43)	15.29 (14.47-16.11)	0.69	0.1394	-	-	-	-	-	-

Table 5.2 Effect of iron treatment on primary and secondary outcomes in Europe and China.

Definition of abbreviations: LSM = least square mean; PVR = pulmonary vascular resistance; sTfR = soluble transferrin receptor; TSAT = transferrin saturation; 6MWD = six-minute walk distance; NTproBNP = N-terminal -pro hormone brain natriuretic peptide; QoL = quality of life; mPAP = mean pulmonary arterial pressure; mRAP = mean right atrial pressure; mPCWP = mean pulmonary capillary wedge pressure; ICPET = incremental bicycle cardiopulmonary exercise testing; ECPET = endurance bicycle cardiopulmonary exercise testing; sTfR = soluble transferrin receptor; 6MWD = six-minute walk distance; NTproBNP = N-terminal -pro hormone BNP; QoL = quality of life; VO2 = oxygen consumption; WR = work rate; VE = minute ventilation; VCO2 = carbon dioxide output

5.3.3 Primary and secondary outcomes by change in iron status

To remove the possible effect of the Iron-placebo group (arm A) still benefitting from the initial iron treatment in the second, placebo phase, I performed a simplified analysis of only the treatment phases. In the treatment-only analyses, patient outcomes in the two arms (Iron-Placebo or Placebo-Iron) of the Europe study were aligned so that baseline corresponds to the timepoint when iron was administered. Change in iron availability 12 weeks after iron infusion was assessed by reduction in the ratio of sTfR/ferritin. Figure 5.4 shows the change in sTfR/ferritin ratio from baseline to 12 weeks in the two arms of the study, with improvements in iron availability observed in both study arms. No difference was observed between the two arms in the ratio of sTfR/ferritin at either of the two timepoints (at baseline: $t_{(df=35)}$= -0.25, P-value = 0.8; at 12 weeks: $t_{(df=34)}$= -1.43, P-value = 0.16). The strongest predictor of change in sTfR/ferritin ratio was baseline levels of sTfR/ferritin ratio, with higher levels at baseline (i.e. more iron deficient) associated with a greater decrease at 12 weeks ($Effect_{Base_sTfR/Ferr}$ = -0.37; 95% CI = -0.09 – -0.65) adjusted for age, sex and baseline NTproBNP levels. This association suggests more severe iron deficiency at baseline resulted in a steeper improvement in iron status in response to iron supplementation.

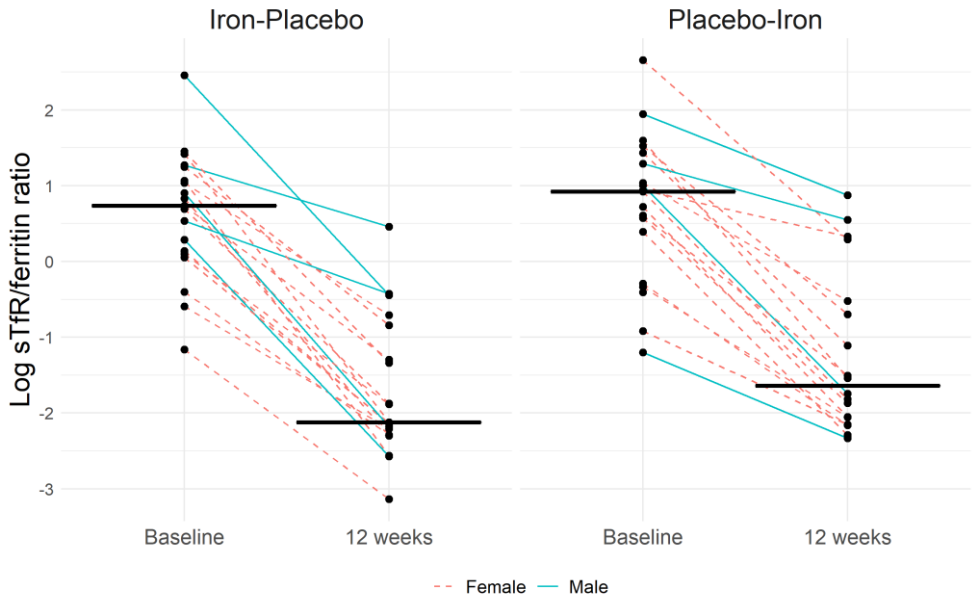


Figure 5.4 Ratio of soluble transferrin receptor and ferritin levels at baseline and twelve weeks after iron administration in the treatment-only analyses.
The two arms (Iron-Placebo or Placebo-Iron) of the study were aligned so that baseline corresponds to the timepoint when iron was administered. Black horizontal lines correspond to median levels. Definition of abbreviations: log = natural logarithm; sTfR = soluble transferrin receptors.

Twenty-four of the thirty-seven patients formed a subset with severe iron deficiency at baseline. Treatment-only analyses were conducted on both the whole cohort and the severely iron-deficient subset of the cohort, where significant improvements in iron availability were observed in both arms (Figure 5.5).

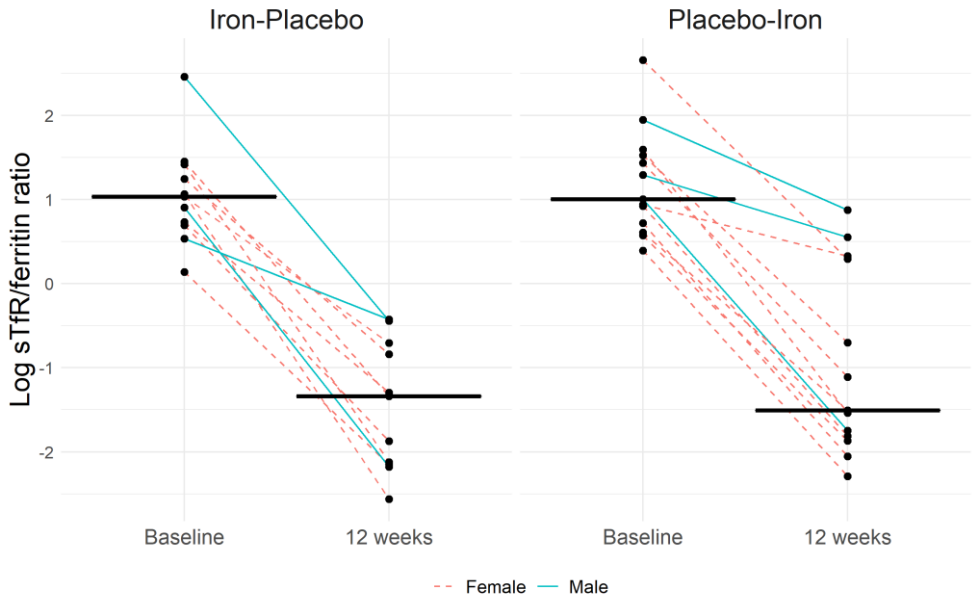


Figure 5.5 Ratio of soluble transferrin receptor and ferritin levels at baseline and twelve weeks after iron administration in the treatment-only analyses in patients with severe iron deficiency at baseline.
The two arms (Iron-Placebo or Placebo-Iron) of the study were aligned so that baseline corresponds to the timepoint when iron was administered. Solid blue lines = men, dashed red lines = women, black horizontal lines = median. Definition of abbreviations: log = natural logarithm; sTfR = soluble transferrin receptors.

Treatment-only analyses results for sixteen outcome measures (including the primary outcome in the Europe study) tested for the association with reduction in the ratio of sTfR/ferritin are shown in Figure 5.6. All models were adjusted for baseline sTfR/ferritin ratio, baseline outcome measurement, age and sex. Results for all coefficients in the sixteen regression models are shown in Supplementary Table 8.3.3 and Supplementary Table 8.3.4. The association of change in peak VO_2 levels (from incremental bicycle CPET) with change in sTfR/ferritin ratio in the severely iron deficient subset ($Effect_{sTfR/Ferr\text{-}change}$ = -0.11 L/min; 95% CI = -0.19 – -0.02) reached suggestive significance (P-value < 0.05). Change in peak O_2 pulse – derived from peak VO_2 (see Introduction above) – was associated with change in sTfR/ferritin ratio in the severely iron deficient subset of patients ($Effect_{sTfR/Ferr\text{-}change}$ = -0.88 mL/beat; 95% CI = -1.4 - -0.36) below the significance threshold adjusted for multiple testing (P-value < $3.12x10^{-3}$) and reached suggestive significance in the whole cohort ($Effect_{sTfR/Ferr\text{-}change}$ = -0.58 mL/beat; 95% CI = -1.01 - -0.14). These negative associations between reduction in sTfR/ferritin ratio and change in peak O_2 pulse and peak VO_2 would suggest a larger increase in exercise capacity

with iron therapy in patients who experienced a smaller reduction in sTfR/ferritin ratio (i.e. less improvement in iron availability).

Significant (peak O_2 pulse) and suggestive (peak VO_2 from the incremental and endurance CPET) associations from the treatment-only analyses (Figure 5.6) were evaluated to see how linear regression models represent the data and whether assumptions for inference and prediction are met.

The association of change in peak VO_2 and reduction in sTfR/ferritin ratio is shown in Figure 5.7/C, whilst residuals and fitted values from the regression model are plotted in Figure 5.7/D. Two extreme observations (observations 26 and 31) fall more than 2 standard deviations (SD) away from the mean change in peak VO_2 ($mean_{peakVO2change}$ = 0.01; $SD_{peakVO2change}$ = 0.18) and exhibit large error variance (residuals) in the full association model (Figure 5.7/C,D). Excluding observations 26 and 31 attenuates the association ($Effect_{sTfR/Ferr\text{-}change}$ = -0.05 L/min; 95% CI = -0.12 – 0.008) of peak VO_2 change and reduction in sTfR/ferritin ratio.

Peak O_2 pulse change was negatively associated with reduction in sTfR/ferritin ratio in the severely iron deficient cohort suggesting an increased effect on exercise capacity in patients with less improvement in iron status. Two observations (observations 2 and 31) are further than 2 SD from the mean change in peak O_2 pulse ($mean_{peakO2pulse\text{-}change}$ = 0.29; $SD_{peakO2pulse\text{-}change}$ = 1.18) shown in Figure 5.7/A. Excluding these two (2 and 31) observations attenuates the association of change in peak O_2 pulse and change in sTfR/ferritin ratio ($Effect_{sTfR/Ferr\text{-}change}$ = -0.58 mL/beat; 95% CI = -1.14 – -0.03) that no longer meets the significance threshold adjusted for multiple testing.

Recorded adverse events were checked for any coinciding reports on patient health that could explain the large differences in exercise capacity between baseline and 12 weeks after iron infusion. For patient 31 (Figure 5.7), the 12-week assessment, where the patient performed considerably worse compared to the baseline exercise test, coincided with an episode of respiratory cold and cough with sputum production treated with two courses of antibiotics. The onset of the respiratory cold was recorded for three days prior to the 12-week assessment and ended three months later.

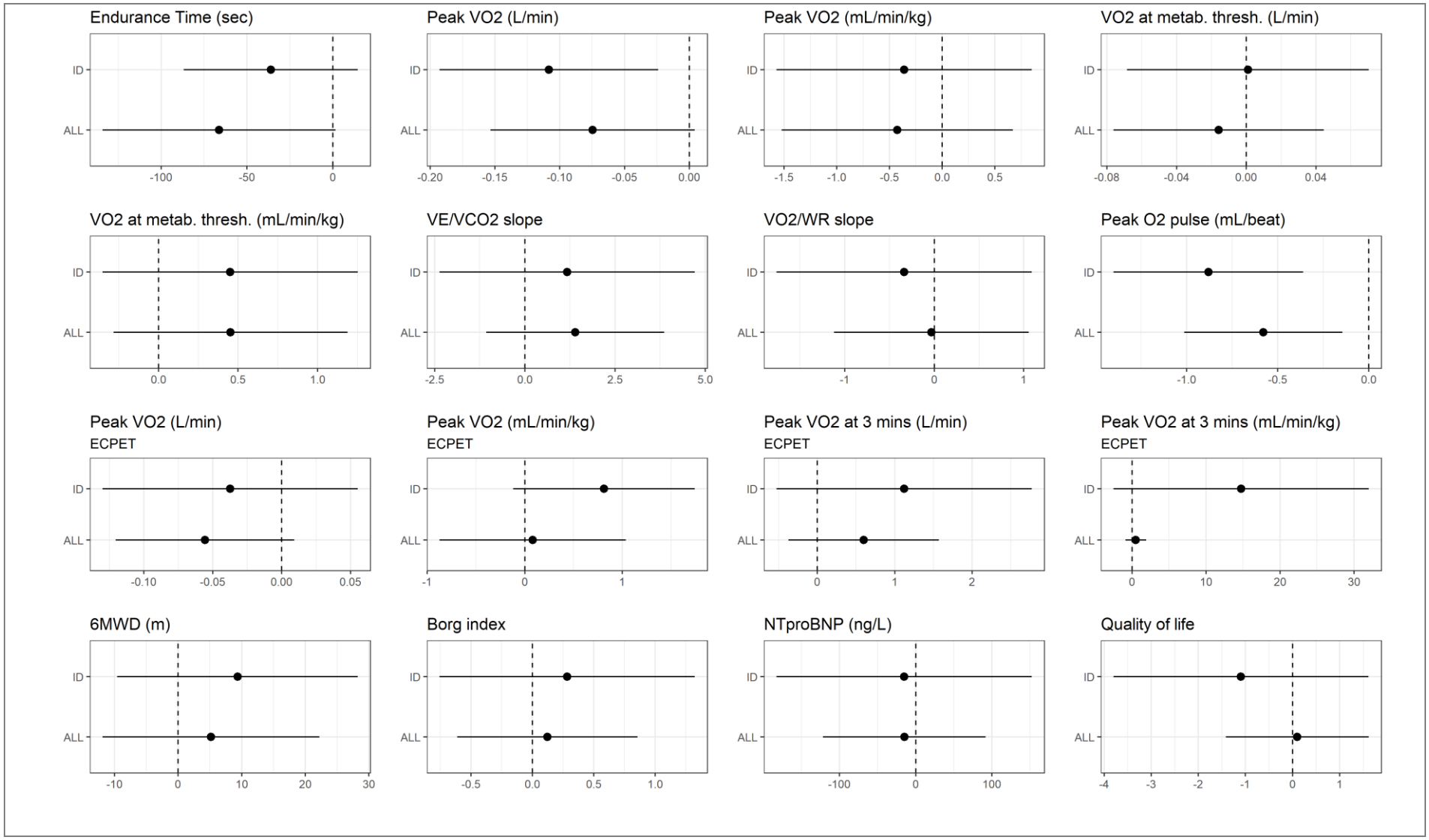


Figure 5.6 Results of the treatment-only analyses: effect of iron status improvement as estimated by sTfR/Ferritin ratio on change in endpoint.
Analyses were run in the whole of the cohort (ALL = 37) and a subset of severely iron deficient patients (ID = 24). Effect estimates on the x-axes are in units of change in the outcome per unit increase in sTfR/ferritin ratio change. All models are adjusted for baseline sTfR/ferritin ratio, baseline outcome, age and sex. Definition of abbreviations: ECPET = endurance bicycle cardiopulmonary exercise testing; 6MWD = six-minute walk distance; NTproBNP = N-terminal -pro hormone BNP; VO2 = oxygen consumption; WR = work rate; VE = minute ventilation; VCO2 = carbon dioxide output.

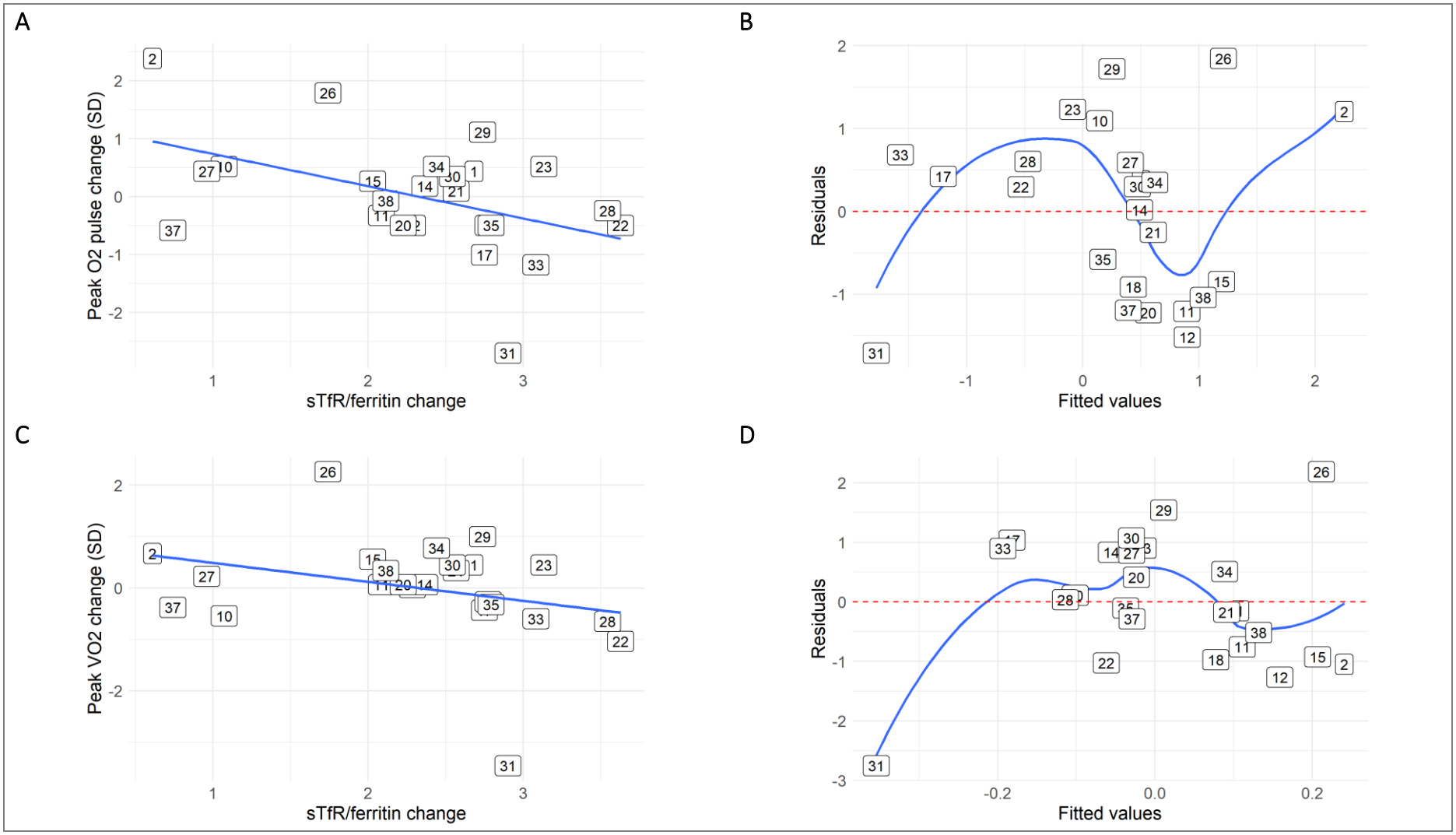


Figure 5.7 Peak O2 pulse and peak VO2 in the treatment-only analyses.

A: *Change in peak O2 pulse as a function of change in sTfR/ferritin ratio in the severely iron deficient cohort.* ***B:*** *Fitted values plotted against standardized residuals from the prediction model of change in peak O2 pulse in the severely iron deficient cohort.* ***C:*** *Change in peak VO2 as a function of change in sTfR/ferritin ratio in the severely iron deficient cohort.* ***D:*** *Fitted values plotted against standardized residuals from the treatment-only prediction model of change in peak O2 pulse in the severely iron deficient cohort. Straight blue lines in A and C: correlation coefficients. Blue lines in B and D: Smoothed localized regression lines show how residuals are scattered around the regression line. Red horizontal line in B and D: zero distance from the regression line of the full model.*

There is marked heteroscedasticity, or in other words unequal variance of residuals across predicted values of the dependent variable, in the peak O_2 pulse and peak VO_2 models (Figure 5.7/B,D) with the majority of observations set around the center of the regression lines, which is especially true of the peak O_2 pulse model (Figure 5.7/B). Interestingly, higher baseline iron availability (lower sTfR/ferritin ratio) is associated with improved exercise capacity measures at 12 weeks for three exercise capacity measures below suggestive significance (P-value < 0.05) (Supplementary Table 8.3.4). These include change in peak VO_2 adjusted for body weight from the endurance CPET ($Effect_{Base_sTfR/Ferr}$ = -1.53 mL/min/kg; 95% CI = -2.91 - -0.16), change in VO_2 at metabolic threshold adjusted for body weight from the incremental CPET ($Effect_{Base_sTfR/Ferr}$ = -1.43 mL/min/kg; 95% CI = -2.46 - -0.4) and change in peak VO_2 adjusted for body weight from the incremental CPET ($Effect_{Base_sTfR/Ferr}$ = -1.57 mL/min/kg; 95% CI = -2.61 - -0.53). Furthermore, there is a trend in some endurance CPET exercise measures for beneficial effects of iron availability improvement on exercise capacity (Figure 5.6: peak VO_2 and peak VO_2 at 3 mins both adjusted and unadjusted for body weight from ECPET).

5.4 Discussion

Treatment with a single infusion of intravenous iron (1000 mg ferric carboxymaltose) in the form of Ferinject (Europe) or Cosmofer (China) in idiopathic PAH (IPAH) patients with iron deficiency without anemia did not improve cardiopulmonary haemodynamics, functional status, exercise capacity or quality of life. Iron supplementation with the single maximal dose of ferric carboxymaltose significantly improved iron status as shown by increased ferritin and decreased sTfR levels and was well tolerated and not associated with adverse side effects in IPAH. The results of this trial are in agreement with the findings of the genetic instrumental variable analyses presented in chapter four which found no evidence for a causal relationship between iron status and PAH susceptibility.

The primary outcome of choice in the original study design was resting pulmonary vascular resistance (PVR) measured by right heart catheterization in both studies (Europe and China). PVR`s relatively low variability makes it a favorable efficacy measure, allowing potential effects to be detectable in small patient populations. It also reports more directly on the primary problem in PAH which is perturbed cardiopulmonary haemodynamics. However, an interim analysis of a subset of patients indicated that the total sample size of the trial would be too small for the detection of any potential effect of iron supplementation on pulmonary haemodynamics; therefore, subjecting all patients to invasive cardiac catheterization was not acceptable. The primary outcome was changed to endurance time from the endurance CPET test in the Europe study, while in China, where no interim analysis took place, PVR remained the primary outcome.

Single infusion of ferric carboxymaltose significantly improved the iron status of iron deficient IPAH patients in both studies as indicated by ferritin and soluble transferrin receptor (sTfR) levels measured 12 weeks after treatment. The majority of patients (55% in the Europe study and 76% in the China study) no longer met any of the iron deficiency criteria used for trial entry 12 weeks after receiving iron infusion. No adverse events related to iron treatment were recorded during the trial and iron supplementation was well tolerated as shown by phosphate levels that remained in the normal range.

I carried out exploratory analyses to investigate if change in primary and secondary outcomes was dependent on the degree of improvement in iron status or if iron supplementation had an effect in severely iron deficient patients only. These analyses were conducted in the Europe study as iron improvement was measured with the ratio of sTfR and ferritin (sTfR/ferritin) and sTfR was not measured in the China study. Changes in sixteen outcomes reflecting disease severity were assessed in the whole cohort (N=38) and a severely iron deficient subset (N=24) that met cut-offs for all three

of ferritin, TSAT and sTfR levels. Improvement in peak O2 pulse (volume of oxygen per heartbeat) from the incremental CPET test was significantly associated with less improvement in iron availability in the severely iron deficient subset. This effect would suggest, somewhat counter-intuitively, that exercise capacity is improved in those patients who exhibit less improvement in iron status in response to iron supplementation, even after correcting for baseline iron status. Peak O2 pulse is derived by dividing the peak VO2 by the peak heart rate during exercise and reflects the amount of oxygen consumed per heartbeat. Peak VO2 showed the same direction of effect, but only reached suggestive significance (P-value < 0.05). These associations were not robust to outlier exclusion, although peak O2 pulse remained below suggestive significance, and inspection of model residuals suggested the unequal variance of residuals warranted caution when making inferences to independent variables in these models. In contrast to these negative associations of exercise capacity and improvement in iron availability, three models indicated higher baseline iron availability (lower sTfR/ferritin ratio) could be associated with improved exercise capacity at 12 weeks below suggestive significance. Since the strongest predictor of change in sTfR/ferritin ratio was baseline sTfR/ferritin ratio - with more improvement being associated with lower baseline levels – a possible explanation of negative association between exercise capacity and more pronounced improvement in iron status would be that it is driven by the poorer baseline iron status. In line with this, higher baseline sTfR/ferritin levels were associated with better outcomes at 12 weeks for three exercise variables below suggestive significance.

The apparent lack of a clinically significant effect of iron supplementation on pulmonary haemodynamics, which are standard efficacy measures in clinical trials of PAH, limits conclusions of an effect that could be drawn to secondary outcome measures which should be interpreted with caution.

Iron deficiency with or without anemia is common in IPAH (Ruiter et al., 2011, Soon et al., 2011, Rhodes et al., 2011a, Krasuski et al., 2011) as well as in other forms of pre- and post-capillary pulmonary hypertension patients from the National Pulmonary Hypertension Service at Hammersmith Hospital, described in the first chapter of this thesis. Prior to this RCT, results from two intervention studies of iron supplementation in idiopathic and associated PAH patients with intravenous ferric carboxymaltose (1000 mg) suggested improvement in well-being indicated by quality of life scores, various exercise capacity measures, including 6-minute walking distance, and skeletal muscle oxygen handling (Viethen et al., 2014, Ruiter et al., 2015). However, these studies were not randomized or placebo controlled and therefore prone to the biases such as inappropriate control selection and lack of blinding, which might have resulted in spurious associations between iron supplementation and improvement in outcomes.

RCTs have shown that intravenous iron supplementation improves the functional status and well-being of patients with iron deficiency and chronic heart failure (CHF), a cardiovascular disease in which iron deficiency is also common and predicts prognosis (Jankowska et al., 2010). Two smaller (N ≤ 40) (Toblli et al., 2007, Okonko et al., 2008) and one large (N = 459) RCT (Anker et al., 2009) reported reduced levels of the myocardial stress biomarker, NTproBNP, and inflammatory biomarkers, improved left ventricular function, improved functional class, 6-minute walking distance and self-reported quality of life in CHF patients after intravenous iron therapy. Apart from differences in iron dosing between RCTs in CHF and this trial (Okonko et al. and Anker et al. adjusted dosing to achieve iron replenishment assessed by either ferritin levels or the Ganzoni formula), other differences between CHF and PAH may be relevant. The mechanisms by which iron supplementation improves outcomes in CHF are yet to be understood. If pulmonary vascular remodeling, the hallmark pathophysiological feature of PAH, does not improve with iron therapy, right heart failure cannot be prevented by iron in PAH.

This pilot study had some limitations. Even though there was a significant improvement in iron status on the study level, some patients would still have met iron deficiency criteria after treatment. Also, patients who were randomized to the arm that received iron first were not iron deficient at the beginning of the placebo phase (shown by similar ferritin levels between the 12-week and 24-week endpoints in this arm of the trial). Treatment -only analyses involving the active treatment phase from both arms aimed to address these caveats (i.e. differences in iron replenishment and lasting effect of iron treatment in the iron-placebo arm), although this had less power compared to the primary analysis of the whole trial. The single infusion of iron and lack of washout period were decided upon to minimize the number hospital visits to patients who live throughout the United Kingdom. Also, the relatively short duration of the study did not allow for any long-term effects of iron replenishment to be detected, such as improved pulmonary hemodynamics caused by the reversal of pulmonary vascular remodeling. However, this study was of similar duration and had a larger sample size than the two open label studies of iron supplementation in PAH mentioned above (Viethen et al., 2014, Ruiter et al., 2015) and therefore could have reproduced their results.

5.5 Conclusions

No evidence of a clinically significant effect of iron supplementation in iron deficient IPAH patients was found in this proof-of-concept RCT. There was no convincing evidence for the effect of varied improvement in iron availability on functional capacity. The results of this trial suggest that iron deficiency does not influence the natural history of IPAH directly and is likely an epiphenomenon which is in line with the findings of the Mendelian Randomisation study presented in chapter two where no evidence of causality between iron status and IPAH was found using genetic instruments.

6 Expression quantitative trait locus mapping in pulmonary arterial hypertension

6.1 Introduction

The relationships between genomic variability and disease risk and endophenotypes has been a focus of many research groups with access to suitable disease cohorts in recent years. This has been aided by the increasing affordability of genotyping and sequencing methodologies. Interpretation of the mechanisms behind effects of genetic variants discovered to associate with phenotypes poses a challenge as many are located in the non-coding space of the genome (Maurano et al., 2012). Non-coding variants exert no direct effect on protein structure, making the biological link to the disease or phenotype more difficult to discern. Variants in the non-coding space may instead exert their effect by influencing gene expression (Maurano et al., 2012). Characterising the genomic control of gene expression can offer a handle on understanding the role of disease-associated variants by linking them to disrupted pathways (Gamazon et al., 2018). Variants associated with the expression of a gene, commonly called expression quantitative trait loci (eQTL) have been described, primarily in population-based studies in the past few years (Westra et al., 2013, Joehanes et al., 2017, Zhernakova et al., 2017).

Differential gene expression analyses between cases and controls may be useful for diagnostic and/or prognostic purposes or in identifying genes which may have a causal role in developing the disease (Gonorazky et al., 2019, Dancik, 2015, Parker et al., 2009). In the case of the latter, distinguishing between differentially expressed genes that are secondary to the disease and those that contribute to disease development is a big and incredibly important challenge. However, if the frequency of an eQTL (or set of eQTLs) for a differentially expressed gene is also different between cases and controls, a causal association may be established, given the assumptions of causal inference, described in *Methods* 2.6, are met. In brief, for a causal relationship to be established between a gene product proxied by an eQTL and a disease, the eQTL has to be independent of the confounders of the association between the gene and the disease and it has to have no direct effect on the disease via a pathway that does not involve the gene it is instrumenting.

Currently one of the main limiting factors for causal inference analysis using genomic data is the number of known eQTLs. Certain eQTL effects which go undetected in healthy cohorts commonly used for eQTL mapping may be unmasked in clinical conditions and diseases. The diagnostic value of

gene expression profiles, i.e. that they can be used to distinguish affected individuals from non-affected ones, underlines the uniqueness of the set of genes expressed in a disease state (Gonorazky et al., 2019, Dancik, 2015). Differential gene expression in diverse states (carcinogenesis, inflammation, etc.) or at diverse stages (such as stages of embryonic development) becomes evident when comparing the gene correlation matrices of a biological system under different conditions or states. Such techniques are called 'differential co-expression networks' (Singh et al., 2018) and have been used successfully to identify genes and gene sets that are important in the state or stage under investigation in a network of genes (Gov and Arga, 2017, Zhu et al., 2016). By principle, functionally involved nodes or modules rewire more frequently than uninvolved ones in states they create and/or maintain and therefore can be identified (Hou et al., 2014).

With the exception of housekeeping genes, the majority of genes in the human transcriptome are tissue-specific. Therefore, tissue selection for gene expression studies of disease is of importance. The lung is the most relevant tissue in the pathogenesis of PAH, although whole blood also has a relevance to PAH and might capture more than just systemic effects given the immune component of this disease. The role of inflammation in PAH has been extensively studied since the observation that inflammatory cells infiltrate various components of the remodeled vascular wall (Tuder et al., 1994). Since PAH can be the complication of many autoimmune diseases such as connective tissue disease, thyroiditis, scleroderma, systemic lupus erythematosus, human immunodeficiency virus infection and schistosomiasis, it might be reasonable to suspect that inflammatory processes, even in the absence of a diagnosed comorbidity, contribute to disease development and maintenance. Another advantage of using whole blood instead of lung tissue is the non-invasiveness of sampling and availability. Lung samples are often only obtainable from the explanted organ after lung transplantation or from post-mortem tissues. Appropriate control samples for differential gene expression analyses are equally hard to come by.

In this study, I characterized the genetic variability of transcriptome-wide gene expression in whole blood samples from 276 consecutively sampled PAH patients in the PAH Cohort Study to define potentially novel eQTL active in this disease state, which could be used in elucidating known genotype-phenotype associations or causal analyses.

6.2 Methods

6.2.1 Study participants and sample processing

A total of 276 patients with idiopathic, heritable or drug-induced pulmonary arterial hypertension (PAH) recruited from expert centres across the UK were whole genome sequenced as part of the UK National Institute for Health Research BioResource (NIHRBR) study and had their transcriptomes profiled through RNA sequencing (RNAseq) as part of the PAH Cohort study (both studies are described in *Methods* 2.1 and 2.4). Table 6.1 shows the demographic characteristics and white blood cell counts of the individuals in this study.

Whole blood (3ml) was collected in Tempus™ Blood RNA Tubes, which were stored at -80 ^{o}C until required. RNA was extracted using a Maxwell robotic system (Promega). Samples with a 260/230 ratio >1.5 and a 260/280 ratio in the range 1.9-2.1 were further quality checked by Bioanalyser and those achieving a minimum RNA Integrity Number (RIN) of 7 were submitted for sequencing. Globin-Zero Gold rRNA Removal Kits (Illumina Inc, San Diego, CA) were used to remove ribosomal RNA contamination from whole blood RNA samples. 75bp paired-end sequencing on a Hiseq4000 was performed on pooled libraries of ~80 samples.

6.2.2 Gene expression data

RNA sequencing and transcript abundance estimation procedures are described in the *Methods* (2.4). White blood cell composition was quantified using quanTIseq, a novel deconvolution algorithm (Finotello et al., 2019). Predicted white blood cell fractions from quanTIseq correlated with clinical white cell fractions available in a subset of PAH patients (Spearman correlation = 0.44-0.73). Well-detected transcripts with a minimum of two reads in at least 95% of samples (n = 26,050) were taken forward to the eQTL analyses.

6.2.3 Genetic data

Whole-genome sequencing in the NIHRBR study is described in *Methods* 2.1.Genetic variants with a minimum minor allele frequency (MAF) of 5%, call rate of 95% and those with a p-value greater than 10^{-5} for Hardy Weinberg equilibrium were considered, leaving a total of 7,362,566 markers for analyses. Genomic positions are reported on the Genome Reference Consortium Human Build 37. Variants were annotated with the dbSNP database (version 150) using the software tool ANNOVAR (Wang et al., 2010).

Characteristic	Men	Women
N (%)	82 (30)	194 (70)
Age at diagnosis, yr	45.6 (24.5)	47.7 (22.1)
Age at sampling, yr	54.4 (22.8)	52.9 (21.8)
B cells	0.028 (0.014)	0.028 (0.016)
Alternatively activated macrophages	0.022 (0.008)	0.020 (0.011)
Neutrophils	0.265 (0.098)	0.285 (0.100)
Natural killer cells	0.020 (0.017)	0.020 (0.020)
CD8+ T cells	0.050 (0.057)	0.050 (0.038)
Regulatory CD4+ T cells	0.150 (0.128)	0.160 (0.128)
Other uncharacterized cells	0.397 (0.117)	0.392 (0.141)

Table 6.1 Patient characteristics and white blood cell fractions in the PAH Cohort study.
Age and white blood cell fractions are presented as median (interquartile range).

6.2.4 eQTL effect estimation

Raw gene count data was first variance stabilized to achieve homoskedasticity (constant variance along the range of mean values) using the DeSeq2 R package (Love et al., 2014). The variance stabilized transcript count data was first quantile normalized to the median distribution. The sample means were then centered to zero and sample variance was linearly scaled such that each sample had a standard deviation of one.

A total of 7,362,566 markers were tested for their effects on the expression of all 26,050 well detected genes. To adjust for possible cryptic relatedness, I computed multidimensional scaling components based on 108,574 independent common (MAF ≥ 5%) genetic markers using PLINK 1.90 (parameters: --indep-pairwise 50 5 0.05). There were no outliers observed based on the first and second components (Figure 6.1).

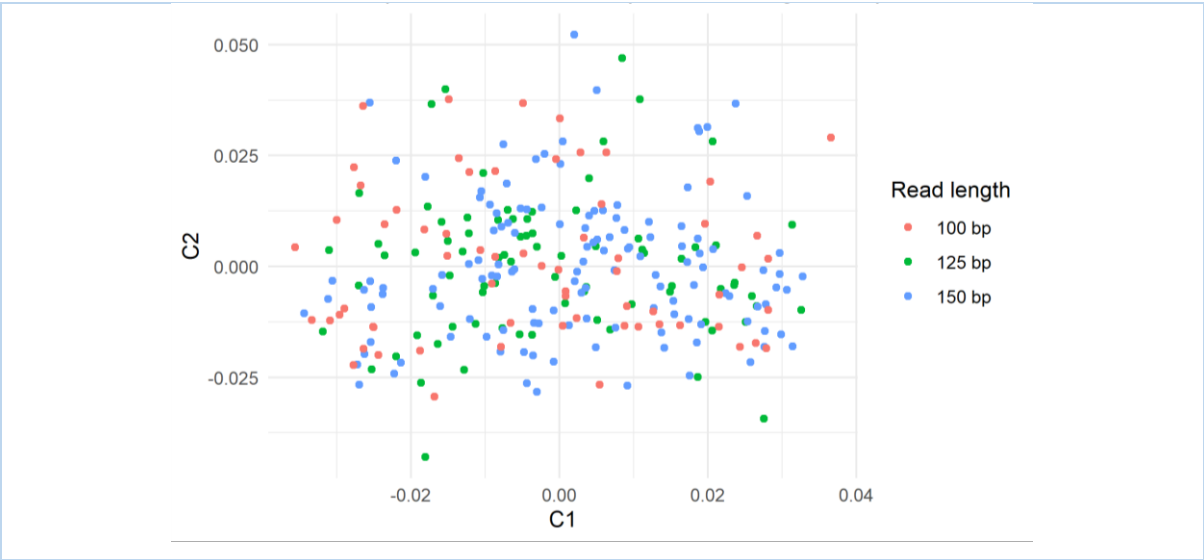


Figure 6.1 Population substructure in the PAH Cohort study.
First (x-axis) and second (y-axis) components from multidimensional scaling (MDS) in 276 PAH patients with gene expression data. Samples were whole-genome sequenced using three different read length chemistries: 100 bp (n=65), 125 bp (n=80), 150 bp (n=131). C1 and C2 = first and second MDS components; bp = base pair.

Variants within 1 megabase (Mb) on either side of the gene`s transcription start site (TSS) were regarded cis-acting or *cis*-eQTL whilst all other markers farther on the same chromosome or on different chromosomes were regarded *trans*-eQTL.

The eQTL analysis was conducted using the QUICKTEST software (Johnson et al., 2009). Other bioinformatics software tools used in data processing and analyses are listed in *Methods* 0. I modelled each well-detected transcript as a function of genotype (coded 0, 1 or 2 reflecting the number of alternate alleles) in a linear regression framework while adjusting for the following covariates: sex, first four components from multidimensional scaling and white blood cell fractions listed in Table 6.1. The genome-wide results were subsequently pruned using the 'clump' command in PLINK 1.90 (parameters: --clump-p1 1.9x10-12; --clump-r2 0.01; --clump-kb 1000) to pick the lead variant with the lowest p-value in a block of variants in linkage disequilibrium, and thus obtain independent effects. I report eQTL reaching a Bonferroni-adjusted significance threshold of $1.9x10^{-12}$ given by dividing the genome-wide significance threshold ($P \leq 5x10^{-8}$) by the number of transcripts tested (n=26,050).

6.2.5 Validation rate

To assess the extent to which eQTL-transcript pairs in PAH overlap with previously reported eQTL-transcript pairs described in healthy populations, I calculated the validation rate of my findings in the two largest published eQTL studies to date and the Genotype-Tissue Expression (GTEx) Project (Westra et al., 2013, Joehanes et al., 2017, Aguet et al., 2019). Validation rate was defined as the

number of significant eQTL-transcript pairs in this study validated by the published study divided by the total number of significant eQTL-transcript pairs that were tested by the published study and multiplied by one hundred.

I extracted all eQTL with effects below the study specific significance threshold (see details below as to how these were defined in the published studies) from the published studies` results for all significant transcripts in this study. I used Ensembl identifiers for the transcripts and genomic coordinates (chromosome and base pair position) on the Genome Reference Consortium Human Build 37 for eQTL when matching my eQTL-transcript pairs to those published by the external studies.

I considered an eQTL-transcript pair to be validated if the lead variant was in linkage disequilibrium ($r^2 \geq 40\%$) with the lead eQTL of the same transcript in the published study. For this purpose, a list of correlated variants was obtained from the European population of the 1000 Genomes Project (Durbin et al., 2010) - using the R package 'proxysnps' - for each lead eQTL reported by the studies used for validation. Each eQTL-transcript pair that reached the study-specific significance threshold in at least one of the studies was considered validated.

Since the complete list of variants or transcripts that passed study-specific quality control is not usually made available by published studies, I had to make assumptions when determining the total number of my significant eQTL-transcript pairs tested in the external studies. This did not apply to GTEx results where all tested eQTL-transcript pairs could be retrieved.

All transcripts present on the expression array used in the other two studies were assumed to have been tested (see below for details). Annotation files were downloaded from the manufacturer`s website for the complete list of transcripts present on the expression arrays (see below for details). Additionally, all eQTL in this study were assumed to have been available for analysis in the studies used for validation which used genotyping array data imputed to high-density reference panels (Table 6.2). Imputed variants are routinely filtered based on some sort of imputation quality score, usually depending on the software used for imputation. Successful imputation in a region depends on the "backbone" of high-quality genotyped variants available as well as the allele frequency of the imputed variant. I restricted my analyses to common variants with MAF ≥ 5% which is at least as high as those of the external studies (Table 6.2).

6.2.6 Studies used for validation

I selected two of the largest published eQTL studies and the GTEx Project to compare my results to. Westra et al. (Westra et al., 2013) meta-analysed eQTL effects from seven studies totaling 5,311 individuals to identify cis-acting effects genome-wide as well as trans-acting effects of 4,542 variants implicated in diseases and traits from the GWAS Catalog (Buniello et al., 2019) at the time of their study. The other eQTL study used for validation was published by Joehanes et al. (Joehanes et al., 2017) who conducted the largest to-date single cohort transcriptome-wide analysis testing both cis- and trans-acting elements genome-wide in the whole blood samples of 5,257 individuals from the Framingham Heart Study. The GTEx Project aims to create a database of genotype and gene expression correlations in multiple human tissues of consenting donors to aid the scientific community in understanding inherited disease susceptibility. The GTEx website provides a browser (https://gtexportal.org/home/testyourown) for retrieving results for any given variant and transcript pair based on Ensembl identifiers in the tissue of interest.

The current release (V8) of the GTEx Project has 838 post-mortem donor samples with genetic data of which 670 contributed to the eQTL mapping in whole blood (GTExPortal, Aguet et al., 2019). Only tissue samples that passed histological examination were accepted for the project, however tissue exclusions were not made based on cause of death. Demographics and cause of death statistics can be found on the GTEx Portal (GTExPortal).

Westra et al. meta-analysed seven studies that measured gene expression on one of Illumina`s whole-genome Expression BeadChips (HT12v3, HT12v4 or H8v2 arrays). Since most of the cohorts used the HT12v3 array, the analyses were restricted to transcripts present on this array (Westra et al., 2013). Joehanes et al. used the Affymetrix Human Exon ST 1.0 array for the whole cohort (Joehanes et al., 2017). The complete list of transcripts was obtained from the manifest files, namely HumanHt-12_V3_0_R3_11283641_A.bgx for Illumina and HuEx-1_0-st-v2.na33.1.hg19.probeset.csv for Affymetrix, available on the manufacturers` websites.

The genotype data from Westra et al. and Joehanes et al. were imputed to the largest haplotype reference panels available (Table 6.2) at the time of conducting their analyses. Westra et al. mapped cis-eQTL differently from our and the other two studies` approach, using a 250 kilobases (kb) maximum distance from the probe midpoint to demarcate cis effects, while eQTLs with a distance greater than 5 Mb were defined as trans-eQTLs. The validation rate of my trans eQTL was not assessed using Westra`s trans eQTL results as they have only tested a small set of selected variants (see above). Also, my cis-eQTL were restricted to the ones not farther than 250 kB from the transcript`s TSS when validating my signals with Westra`s results.

	Westra et al.	Joehanes et al.	GTEx	This study
Gene expression array	Illumina HumanHT-12 v4.0	Affymetrix HuEx 1.0 ST	RNAseq	RNAseq
Imputation panel	HapMap2	1000-Genomes	WGS	WGS
MAF threshold	≥5%	≥1%	≥1%	≥5%
***N* variants in analysis**	not reported	8,510,936	10,008,325	7,362,566
***Cis* eQTL**	≤250 kb from probe midpoint	≤1Mb from the TSS	≤1Mb from the TSS	≤1Mb from the TSS
***Trans* eQTL**	≥5 Mb from probe midpoint	>1 Mb from the TSS	>1 Mb from the TSS	>1 Mb from the TSS

Table 6.2 Characteristics of the PAH Cohort eQTL mapping study and the eQTL studies used for validation. *Comparison of gene expression and genetic data and expression quantitative trait locus definitions of this study and the studies used for validation. MAF = minor allele frequency; TSS = transcription start site; RNAseq = RNA sequencing; WGS = whole genome sequencing.*

6.2.7 Overlap with common variants affecting PAH susceptibility

Our group has previously investigated the genome-wide common (MAF ≥ 2%) genetic variation of PAH susceptibility in 2,085 idiopathic or heritable PAH cases and 9,659 controls in a meta-analysis of four PAH genome wide association studies (hereon referred to as the 'PAH GWAS') (Rhodes et al., 2018). This largest-to-date PAH GWAS observed two loci above genome-wide significance ($5x10^{-8}$), one within *HLA-DPA1/DPB1*, which encodes a member of the major histocompatibility complex (MHC) class II family, and one (containing two independent associated signals) upstream of *SOX17*, encoding a transcription factor. The *HLA-DPA1/DPB1* genotype was shown by the same study to strongly affect survival in PAH patients, while the variants near *SOX17* were functionally linked to altered expression of *SOX17.* Apart from these two loci, there were thirty-two further variants that reached suggestive significance (p-value $<10^{-5}$) and passed the visual assessment of regional association plots ensuring each of them is a good quality signal in a cluster of variants linked by linkage disequilibrium (LD). All PAH susceptibility lead variants reaching suggestive significance and nearby variants in LD ($r^2 \geq 40\%$) were retrieved to check for overlapping loci with the significant eQTL in this study.

6.2.8 Overlap with variants from the GWAS Catalog

An important aim of this study is to identify eQTL in PAH with effects that have not been observed in population-based studies, and to assess the clinical relevance of these PAH-specific eQTL in PAH susceptibility itself (see above), but also in a wide range of diseases.

The NHGRI-EBI GWAS Catalog of published genome-wide association studies (GWAS) (Buniello et al., 2019) provides a curated database of genetic markers associated with a wide range of traits and

diseases. All variant-phenotype pairs below an association P-value threshold of $9x10^{-6}$ were downloaded from the GWAS Catalog website on March 5, 2020. The GWAS Catalog at the time contained 113,510 unique variants (including variants in LD) reported as lead variants for 4,314 phenotypes. The list of GWAS variants were intersected with each eQTL locus (including the lead eQTL and variants in LD with r2 ≥ 40%) identified in this study.

6.2.9 Functional enrichment analysis of genes with novel eQTL

An important aim of this eQTL study is to differentiate previously reported eQTL that influence gene expression in disease-free populations from eQTL with unmasked or augmented effects in PAH. Potentially novel eQTL-eGene pairs in this study are the ones eligible to be validated by at least one study and not validated by any of the studies that tested them.

Functional enrichment analysis of genes with novel eQTL in PAH using the WEB-based GEne SeT AnaLysis Toolkit (WebGestalt) (Wang et al., 2017) was conducted to determine if these genes are predominantly from pathways with relevance to the pathobiology of PAH. A form of pathway analysis called over-representation analysis (ORA) (Khatri et al., 2012) was run using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa et al., 2007) database to group these novel e-genes into a smaller number of gene-sets relating to biological processes. In pathway analysis, genes are grouped together based on their involvement in common pathways.

ORA assesses the fraction of genes in a pathway found among the list of query genes (Khatri et al., 2012), commonly a list of differentially expressed genes in a disease or condition. Each pathway in the database is assessed for the query genes as well as a list of appropriate background genes to estimate the probability of finding the number of genes shared between the query list of size n and the pathway in a random sample of n genes from the background list. ORA implemented in WebGestalt is based on Fisher`s test and therefore the hypergeometric distribution.

The query list in this study was the list of novel e-genes whilst all the tested genes that are also present on the expression array (see above) of at least one of the studies used for the validation process constituted the list of background gene. Pathways below an FDR corrected significance threshold of 0.05 were considered significant.

6.3 Results

6.3.1 eQTL discovery

I observed 2,354 significant eQTL-eGene pairs in total accounting for 9% of all genes tested in this transcriptome- and genome-wide eQTL mapping in PAH. 147 of these pairs were novel transcripts not yet annotated in the Ensembl database (Figure 6.2). 90% of all significant eQTL-eGene pairs were cis-acting. The top 50 lead eQTL-eGene pairs, ranked by percent of variance explained (R^2), are presented in Table 6.3. The proportion of variation explained (R^2) in gene expression levels was generally lower for trans-acting eQTL than for cis-acting eQTL ($median_{cis\text{-}eQTL}$ = 23.1%, $IQR_{cis\text{-}eQTL}$ = 12.5%; $median_{trans\text{-}eQTL}$ = 21.5%, $IQR_{trans\text{-}eQTL}$ = 11.5%; $t_{(df=1332)}$ = 4.5, p-value = $8.3x10^{-6}$).

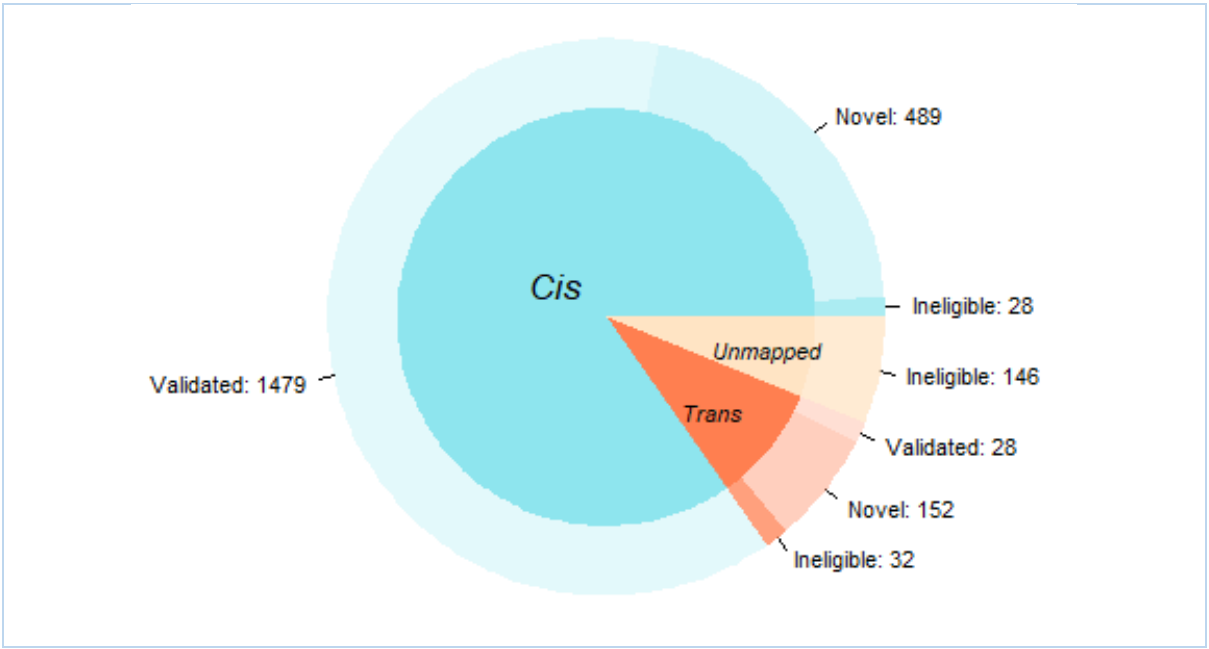


Figure 6.2 Results of the PAH Cohort eQTL discovery.
Pie chart showing the proportions of cis-, trans- and unmapped eQTL identified in the PAH Cohort eQTL mapping. Counts of eQTL within each category are shown next to the category name. An eQTL counts 'novel' if it was eligible to be validated in at least one of the previously published studies and was not among their significant eQTL-eGene pairs. Those eQTL-eGene pairs that reached the study-specific significance threshold in at least one of the published studies were considered 'validated'. Unmapped transcripts not yet annotated in the Ensembl database could not be validated. Ineligible eQTL-eGene pairs were not tested by any of the three studies used for validation.

rs ID	chr:pos	ref	alt	Gene	Ensembl gene ID	cis/tran	P-value	R^2 (%)	validat
rs3793638	9:123582697	G	T	*RP11.27I1*	ENSG00000226752	cis	$1.17x10^{-97}$	61.4	Y
rs10876864	12:56401085	G	A	*RPS26*	ENSG00000197728	cis	$7.6x10^{-89}$	59.1	Y
rs1055337	10:16555431	A	T	*C1QL3*	ENSG00000165985	cis	$2.88x10^{-86}$	58.4	Y
rs2732317	8:33330687	A	C	*FUT10*	ENSG00000172728	cis	$1.36x10^{-83}$	57.6	Y
rs10771798	12:31223322	A	G	*DDX11*	ENSG00000013573	cis	$5.47x10^{-83}$	57.4	Y
rs2295768	6:2954602	C	T	*HTATSF1P*	ENSG00000271361	cis	$7.1x10^{-83}$	57.4	N
rs1458255	4:39446549	G	A	*RPL9P9*	ENSG00000237550	trans	$1.74x10^{-82}$	57.3	NA
rs2240101	15:31115709	G	C	*RP5.1086*	ENSG00000259845	cis	$4.5x10^{-79}$	56.2	N
rs11231152	11:62351984	G	A	*RP11.371*	NA	unmap	$1.62x10^{-78}$	56	NA
rs2302559	7:105903904	T	C	*NAMPTL*	ENSG00000229644	trans	$6.35x10^{-78}$	55.9	N
rs11053539	12:10154580	G	T	*CLEC12A*	ENSG00000172322	cis	$1.02x10^{-77}$	55.8	Y
rs9308807	2:109173930	T	C	*LIMS3*	ENSG00000256977	trans	$6.63x10^{-78}$	55.8	NA
rs4554383	7:150457308	T	C	*TMEM17*	ENSG00000002933	cis	$1.36x10^{-77}$	55.7	Y
rs2536690	12:9618649	G	A	*OVOS*	ENSG00000214776	cis	$4.8x10^{-77}$	55.6	Y
rs2927608	5:96252432	G	A	*ERAP2*	ENSG00000164308	cis	$1.53x10^{-75}$	55.1	Y
rs1628955	15:65187220	T	C	*ANKDD1A*	ENSG00000166839	cis	$1.58x10^{-74}$	54.7	Y
rs7520185	1:213016007	G	A	*FLVCR1.A*	ENSG00000198468	cis	$2.22x10^{-74}$	54.7	Y
rs6720375	2:132440073	G	C	*AC093724*	ENSG00000197927	cis	$3.62x10^{-74}$	54.6	Y
rs9890200	17:48624523	A	C	*SPATA20*	ENSG00000006282	cis	$2.26x10^{-73}$	54.3	Y
rs79379134	16:30292500	A	T	*RP11.347*	ENSG00000183604	cis	$6.28x10^{-73}$	54.2	Y
rs9349050	6:37663624	T	C	*MDGA1*	ENSG00000112139	cis	$8.74x10^{-73}$	54.1	Y
rs74827019	19:10403603	C	T	*AC114271*	NA	unmap	$1.99x10^{-72}$	54	NA
rs745666	17:72744798	G	C	*RP11.452I*	ENSG00000266036	cis	$1.9x10^{-70}$	53.3	N
rs7635	11:721570	C	T	*RPLP2*	ENSG00000177600	cis	$2.05x10^{-70}$	53.3	N
rs4554383	7:150457308	T	C	*TMEM17*	ENSG00000106565	cis	$1.48x10^{-69}$	53	Y
rs2431530	5:102118007	C	T	*PAM*	ENSG00000145730	cis	$1.48x10^{-69}$	53	Y
rs58288565	8:144631031	A	G	*RP11.661*	ENSG00000254859	cis	$2.67x10^{-69}$	52.9	Y
rs10948059	6:42928461	C	T	*PEX6*	ENSG00000124587	cis	$1.94x10^{-69}$	52.9	Y
rs4654746	1:21772167	C	G	*NBPF3*	ENSG00000142794	cis	$6.18x10^{-69}$	52.7	Y
rs471521	12:10115647	A	T	*CLEC12B*	ENSG00000256660	cis	$1.72x10^{-68}$	52.6	Y
rs6511291	19:21950402	T	C	*ZNF100*	ENSG00000197020	cis	$1.94x10^{-68}$	52.5	N
rs7314691	12:661981	G	A	*B4GALNT*	ENSG00000139044	cis	$3.86x10^{-68}$	52.4	Y
rs13197384	6:135818897	C	A	*AHI1*	ENSG00000135541	cis	$3.96x10^{-68}$	52.4	Y
rs2932102	12:48665720	A	G	*ZNF641*	ENSG00000167528	cis	$6.65x10^{-68}$	52.3	Y
rs2688874	10:52122258	G	A	*RP11.564*	ENSG00000231345	cis	$1.08x10^{-67}$	52.3	Y
rs9392465	6:3162378	C	A	*TUBB2A*	ENSG00000137267	cis	$1.77x10^{-67}$	52.2	Y
rs756553	21:45538647	T	C	*CH507.9B*	NA	unmap	$1.76x10^{-67}$	52.2	Y
rs1332018	1:110282972	G	T	*GSTM3*	ENSG00000134202	cis	$3.68x10^{-67}$	52.1	NA
rs6858440	4:47917474	A	G	*NFXL1*	ENSG00000170448	cis	$1.12x10^{-66}$	51.9	Y
rs2151511	20:60776444	A	G	*MTG2*	ENSG00000101181	cis	$2.01x10^{-66}$	51.8	Y
rs964611	15:48597514	C	A	*SLC12A1*	ENSG00000074803	cis	$2.99x10^{-66}$	51.7	Y
rs4582	2:27604279	A	G	*PPM1G*	ENSG00000115241	cis	$3.44x10^{-66}$	51.7	Y
rs10483877	14:77182302	C	T	*VASH1*	ENSG00000071246	cis	$9.56x10^{-66}$	51.5	Y
rs12904576	15:101094127	T	C	*RP11.526I*	ENSG00000270127	cis	$2.81x10^{-65}$	51.3	Y
rs55794721	1:25561667	G	A	*RHD*	ENSG00000187010	cis	$3.82x10^{-65}$	51.3	Y
rs1059307	6:86387888	G	T	*SNHG5*	ENSG00000203875	cis	$7.83x10^{-65}$	51.2	Y
rs67933574	12:11226237	A	G	*TAS2R43*	ENSG00000255374	cis	$9.15x10^{-65}$	51.1	Y
rs11079015	17:39975164	C	T	*NT5C3B*	ENSG00000141698	cis	$8.58x10^{-65}$	51.1	N
rs1256839	15:90454757	G	C	*ARPIN*	ENSG00000242498	cis	$2.28x10^{-64}$	51	N
rs11637626	15:32810421	G	A	*RP11.632*	ENSG00000254912	cis	$1.86x10^{-64}$	51	Y

Table 6.3 Lead variants for the top 50 expression quantitative trait loci in the PAH Cohort ordered by variance explained in gene expression.
Column names: rs ID = dbSNP (150) variant identifier; chr:pos = chromosome and genomic position (Ensembl GRCh37 coordinates); ref = reference allele in the human genome reference GRCh37; alt = alternative allele in the human genome reference GRCh37; Ensembl gene ID = Ensembl stable gene identifiers; variant type = dbSNP version 150 variant annotation; R2 (%) = percent variance explained in gene expression by the genetic variant; validated = validated (Y) by at least one of the published studies, or not validated (N) by any of the studies, or ineligible to be validated (NA).

6.3.2 Validation rate

To assess how well my results overlap with previously reported eQTL in healthy populations, I calculated the validation rate of significant cis-eQTL in the studies of Westra et al., Joehanes et al. and the GTEx Project while the validation rate of significant trans-eQTL were validated in the studies of Joehanes et al. and the GTEx Project. An eQTL-eGene pair was validated if the lead variant was in linkage disequilibrium (r2 ≥ 40%) with the lead eQTL (FDR < 0.05) of the same transcript in the opposite study. The number of validated eQTL-eGene pairs was divided by the total pairs eligible to be validated.

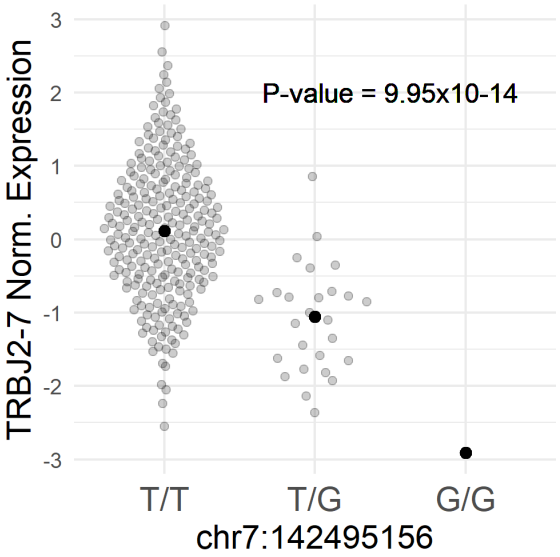
1,509 out of the 1,986 (76%) unique *cis*-eGenes could be mapped to the Illumina HT12v3 array used by Westra et al., 2,022 out of 2,134 (95%) unique eGenes were mapped to the Affymetrix Human Exon chip used by Joehanes et al., whilst results from the GTEx Portal could be obtained for 73% of eQTL-eGene pairs in this study. 28% of my cis-eQTL were validated by Westra et al., 31% by Joehanes et al. and 90% of tested cis-eQTL validated in GTEx. Overall, 75% of significant *cis*-eQTL-transcript pairs were validated in at least one of the studies (Figure 6.2). The overall validation rate for *trans*-eQTL-transcript pairs reached 16%, with all but one being validated by the GTEx project alone. Joehanes et al. validated only one *trans*-eQTL-transcript pair (chr1:248039451 (C/T) - JAM3).

6.3.3 Overlap with common variants affecting PAH susceptibility

One of the significant *cis*- eQTL in PAH was associated with PAH susceptibility and met a suggestive threshold of significance (P value < 10^{-5}) in a recent meta-analysis of PAH GWAS (Rhodes et al., 2018). The eQTL chr7:142495156 (T/G) associated with the expression of *TRBJ2-7* which encodes the J region of the variable domain of the T cell receptor beta chain, is validated in GTEx (P-value = $1.4x10^{-6}$) in whole blood and is directionally consistent between the two studies with the major allele 'T' increasing *TRBJ2-7* transcript levels. However, it is worth noting that the effect of chr7:142495156 (T/G) on gene expression in the GTEx population ($Effect_{T\ allele}$ = 0.27 SD, 95% CI = 0.16-0.38) is much weaker than its effect in this PAH cohort ($Effect_{T\ allele}$ = 1.16 SD, 95% CI = 0.87-1.45).

Chr7:142495156 (T/G) is a missense variant that changes the seventh amino acid coded by the single exon of *TRBJ2-7* from a phenylalanine to a valine residue. Figure 6.4 illustrates the normalized expression levels of *TRBJ2-7* in PAH patients in this eQTL study and genotype counts in PAH cases and controls in the biggest contributing cohort (880 PAH and 5,048 controls) of the PAH GWAS (Rhodes et al., 2018). The effect of chr7:142495156 (T/G) on PAH susceptibility was tested in three out of the four cohorts in the PAH GWAS where the G allele – associated with lower *TRBJ2-7* expression - was found in all three GWASs to increase the odds of developing PAH, resulting in a

meta-analysed P-value of $1.1x10^{-6}$ (OR = 1.74; 95% CI = 1.39-2.17). Chr7:142495156 (T/G) is the lead variant in both this eQTL study and the PAH GWAS. *TRBJ2-7* is primarily expressed in whole blood (median transcripts per million [TPM] = 11, N=755), in the spleen (median TPM = 8, N=241), in the small intestine (median TPM = 3.5, N=187) as well as in the lung (median TPM = 2, N=578), however chr7:142495156 (T/G) is reported in GTEx as a significant eQTL in whole blood only. Tissue-specific gene expression data was derived from the GTEx Analysis Release V8 (https://gtexportal.org/home/gene/TRBJ2-7).



Genotype counts for chr7:142495156 in the NIHRBR cohort				
	T/T	T/G	G/G	MAF
PAH	767	111	2	6.5%
Controls	4648	393	7	4%

TRBJ2-7		
Start	End	Length
142,495,140	142,495,186	47
CTCCTACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACAG		

Figure 6.3 PAH susceptibility variant chr7:142495156 (T/G) is also an eQTL for TRBJ2-7.
Top: *Normalized TRBJ2-7 expression levels in 273 PAH patients by genotype at the eQTL chr7:142495156.*
Bottom: *Table of genotype counts and minor allele (G) frequency in PAH cases and controls from the UK National Institute for Health Research BioResource (NIHRBR) cohort. Genomic coordinates with the nucleotide sequence of TRBJ2-7 from the GRCh37 Ensembl browser. Highlighted nucleotides represent single nucleotide polymorphisms (SNP) in TRBJ2-7. Red SNPs are stop gained variants, yellow SNPs are missense and green SNPs are synonymous. The eQTL chr7:142495156 (T/G) is a missense SNP shown in bold in the nucleotide sequence.*

6.3.4 Overlap with variants from the GWAS Catalog

In order to assess the relevance of eQTL identified in this study to broad clinical phenotypes, I intersected the list of 2,173 unique eQTL with the database of published genotype-phenotype associations from the GWAS Catalog. In total, 929 eQTL were reported for at least one trait previously (median $_{\text{N GWAS traits / eQTL}}$ = 2; IQR $_{\text{N GWAS traits / eQTL}}$ = 3). Ninety-eight (11%) of the 929 eQTL had at least 10 or more unique GWAS phenotypes associated with them. Figure 6.4 shows the number of overlapping GWAS phenotypes per eQTL in the validated and novel eQTL categories. The proportion of cis-acting eQTL overlapping with at least one published GWAS association in the validated subset

was 47%, whilst in the novel subset it was 37%. Among *trans*-eQTL, the proportion of loci overlapping at least one GWAS trait was 50% in the validated, and 27% in the novel group. The two-sample equality of proportions test was significant for both cis-, and trans-acting eQTL (Cis: $\chi^2_{(df=1)}$ = 13.9; 95% CI = 0.05-0.15; p-value = 0.0002. Trans: $\chi^2_{(df=1)}$ = 4.2; 95% CI = -0.003-0.45; p-value = 0.04), suggesting a higher proportion of overlapping loci in the validated eQTL groups. Interestingly, the proportion of novel eQTL previously reported for lung-related phenotypes (11%) is twice as high as the proportion of validated eQTL associated with lung-related phenotypes (5.5%) ($\chi 2_{(df=1)}$ = 6.9; 95% CI = 0.006-0.1; p-value = 0.009). Lung-related phenotypes reported for the novel eQTL include chronic obstructive pulmonary disease (COPD), lung function (forced vital capacity, forced expiratory volume), low versus high forced expiratory volume, interstitial lung disease, emphysema, lung cancer and lung adenocarcinoma.

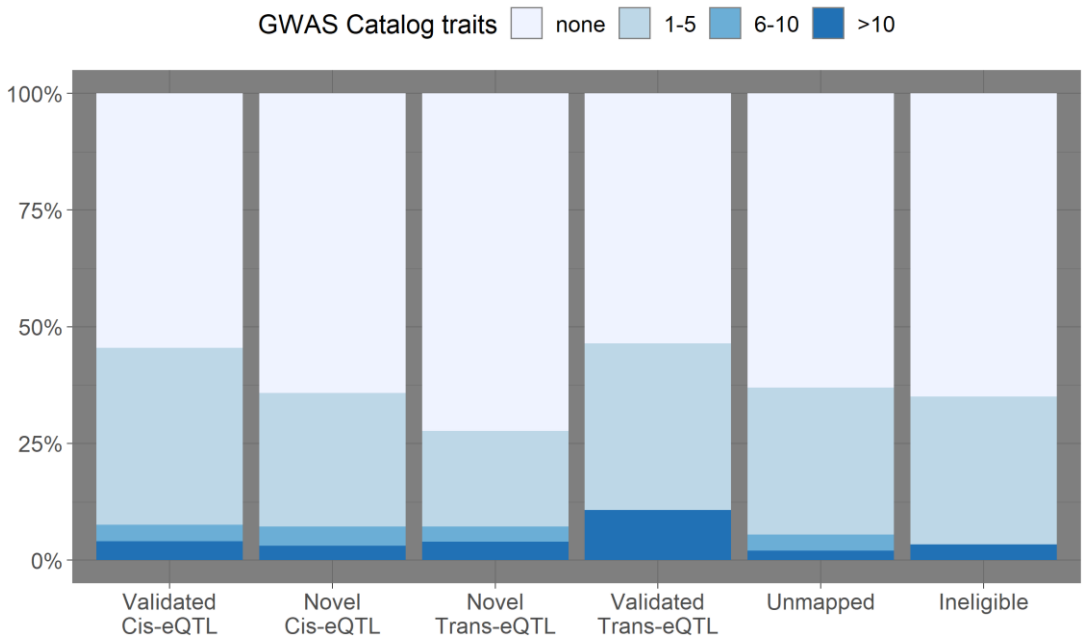


Figure 6.4 Reported GWAS traits for eQTL in the PAH Cohort.
Percentage of binned numbers of overlapping GWAS traits (y-axis) per eQTL in each eQTL category (x-axis). Reported genotype-trait associations were downloaded from the NHGRI-EBI Catalog of published genome-wide association studies.

6.3.5 Functional enrichment analysis of genes with novel eQTL

To assess if novel e-Genes in the PAH Cohort are enriched for certain biological processes or pathways active in diseases, I ran ORA in WebGestalt using the KEGG knowledge-base. A list of 606 novel e-Genes were queried using a list of 25,128 background genes.

Five pathways were found to be significantly (FDR < 0.05) enriched, with the pathways taste transduction, graft-versus-host disease and autoimmune thyroid disease being the three most significant (Table 6.4). Supplementary Table 8.4.2 in the *Appendix* lists the overlap between these novel e-Genes with their corresponding eQTL in PAH and each of the five pathways.

Gene set	Description	Size	Overlap	Expectation	Enrichment ratio	P-value	FDR
hsa04742	Taste transduction	48	8	0.87	9.24	1.78×10^{-6}	5.72×10^{-4}
hsa05332	Graft-versus-host disease	32	6	0.58	10.40	1.85×10^{-5}	2.86×10^{-3}
hsa05320	Autoimmune thyroid disease	34	6	0.61	9.79	2.66×10^{-5}	2.86×10^{-3}
hsa05330	Allograft rejection	33	5	0.59	8.40	2.75×10^{-4}	0.02
hsa04940	Type I diabetes mellitus	36	5	0.65	7.70	4.19×10^{-4}	0.03

Table 6.4 Pathways enriched for genes with novel eQTL in the PAH Cohort.
Results of the over-representation analysis using the KEGG pathway database. Column names: Gene set = searchable identifier of the gene set in KEGG; Size = number of genes participating in the gene set and overlap gives the number of genes in the gene set that overlap with the queried gene list; Expectation = number of genes that are expected to be shared between the queried gene list and the gene set if the queried list was a random sample of the background gene list; Enrichment ratio = fold enrichment of the gene set for the queried gene list; P-value = enrichment significance of the Fisher`s exact test; FDR = false discovery rate adjusted significance.

6.4 Discussion

Gene expression serves as an intermediate phenotype between genetic variants and associated phenotypes such as clinical diagnoses-, presentation and biomarkers, anthropometric- and behavioural traits, etc. Previous eQTL studies provided support for the idea that genetic effects on gene expression have phenotypic consequences and that eQTL aid in the biological interpretation of associated genetic markers in disease (Zhernakova et al., 2017, Westra et al., 2013, Joehanes et al., 2017). However, eQTL effects are not only dependent on the tissue or cell type under investigation (Powell et al., 2012, Aguet et al., 2019), but also on the environment or context (Zhernakova et al., 2017, Barreiro et al., 2012, Fairfax et al., 2014, Çalışkan et al., 2015) in which gene expression is measured. This implies that a more comprehensive eQTL map could be constructed by extending mapping efforts beyond population-based studies. I ran a transcriptome-wide eQTL analysis in a cohort of 276 PAH patients to characterize the genetic control of gene expression variability in this condition, uncovering potentially novel eQTL not detected in healthy populations. The resulting 2,354 significant eQTL-eGene pairs were intersected with the results of two previously published population-based eQTL studies and the results of the GTEx Project. 25% of cis-acting eQTL-eGene pairs and 84% of trans-acting eQTL-eGene pairs in this study were not found in any of the three studies used for validation. Novel and validated eQTL were investigated for their pathobiological relevance to both PAH - using the largest-to-date PAH GWAS published by our group (Rhodes et al., 2018) - and a wide range of diseases from the NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog).

Nearly half (N=929, or 43%) of all eQTL identified in this study colocalized with at least one trait or disease in the GWAS Catalog. Even though the proportion of colocalizing eQTL was higher for validated *cis*- and *trans* -eQTL than for novel eQTL, the proportion of eQTL associated with lung-related phenotypes was twice as high among novel eQTL than validated eQTL previously detected in population-based studies (p-value = 0.009). This indicates that these novel blood eQTL identified in PAH might be highly informative for pulmonary diseases such as COPD, interstitial lung disease and emphysema. The higher proportion of GWAS-trait associated eQTL in the validated subset might be explained by validated eQTL mainly consisting of cis-acting variants (98.1%), whilst the proportion of *cis*-eQTL among novel variants is 76.3 %. Trans-acting eQTL usually explain a lower proportion of variation (R^2) in gene expression levels than cis-acting eQTL, which was observed in this study too. The higher variance explained by the predominantly cis-acting eQTL in the validated subset means that these eQTL are more likely to be discovered - due to better statistical power - in genetic association studies. Also, novel eQTL in this study were not replicated in an independent PAH cohort,

therefore it is expected that a proportion of these novel association signals is spurious which could contribute to the lower colocalization rate with previously reported GWAS loci.

As an example for the overlap between lung-related GWAS traits and eQTL in PAH , the novel *cis*-eQTL chr7:74069645 (A/G) associated with expression of the gene *GTF2IRD2B* overlaps with multiple phenotypes including COPD (Zhu et al., 2019), lung function (Kichaev et al., 2019), cardiovascular disease (Kichaev et al., 2019) and diastolic blood pressure (Evangelou et al., 2018). Deletion of *GTF2IRD2B* leads to a rare congenital disease called Williams-Beuren syndrome which frequently presents with supravalvular aortic stenosis (SVAS; OMIM 185500), a congenital heart defect characterized by the narrowing of the aorta, pulmonary and coronary arteries and other blood vessels (Merla et al., 2012).

The list of eQTL was also intersected with susceptibility loci for PAH including two loci reaching genome-wide significance (p-value < $5x10^{-8}$) and 34 loci below the threshold for suggestive significance (p-value < 10^{-5}). The lead variant of one of these possible PAH susceptibility loci was found to colocalize with a lead *cis*-eQTL associated with the expression of the gene *TRBJ2-7*. The same allele that lowers *TRBJ2-7* expression increases the risk of developing PAH. *TRBJ2-7* - T Cell Receptor Beta Joining 2-7 - codes for the J region of the variable domain of the T cell receptor beta chain which participates in antigen recognition. T-cell receptors are present on the surface of T-lymphocytes and are essential in the normal functioning of immune responses. The complex pathophysiological mechanisms behind PAH involve multiple driving factors of which immune dysfunction and inflammation are suspected to be among the major contributors (Rabinovitch et al., 2014). Certain subtypes of T-cells have been proposed to be detrimental, whilst others to attenuate PAH progression. Whilst T-helper (Th) cells (Th1, Th2 and Th17) have been implicated in causing autoimmune response through production of interleukins (Steiner et al., 2009, Maston et al., 2017, Chen et al., 2018), regulatory T-cells (Treg) were found to have an immunosuppressive role in PAH and attenuate PAH development (Tamosiuniene et al., 2011). Tamosiuniene et al. showed that Treg deficient rats develop severe PAH and that immune-reconstitution with Tregs prevents early inflammation and attenuates development of PAH in response to vascular injury (Tamosiuniene et al., 2011). The marked difference in the magnitude of effect of chr7:142495156 (T/G) on *TRBJ2-7* expression between the PAH cohort and the GTEx population may indicate that the effect of this eQTL becomes amplified when pathways and processes in which T-cells play a role become activated. The importance of antigen presenting and recognition in PAH is underlined by the most significant genetic variant discovered in the PAH GWAS in the *HLA-DPA1/DPB1* locus encoding class II major histocompatibility complex (MHC) molecules (Rhodes et al., 2018) which associated with three HLA alleles all containing a glutamic acid (Glu^{69}) at amino acid residue 69. MHC molecules at the

surface of antigen presenting cells are recognized by T cell receptors which are restricted to sensing antigens bound to class I or class II molecules of the MHC. The *TRBJ2-7* and *HLA-DPA1/DPB1* PAH susceptibility loci emphasize the role of immune dysregulation in PAH development (Rabinovitch et al., 2014, Kumar and Graham, 2018) which warrants further investigation. Colocalization of a disease susceptibility variant with a *cis*-eQTL may provide more robust evidence for the direction of the relationship (if there is one, assuming no horizontal pleiotropy at the eQTL) between the disease and gene expression, i.e. altered gene levels caused by the *cis*-eQTL are more likely to influence the disease phenotype than vice versa. However, *in silico* analysis without *in vitro* or *in vivo* experiments cannot provide definitive evidence for a genetic variant`s effect on gene expression.

Pathway analysis of eGenes associated with novel eQTL identified five biological processes from the KEGG database enriched for this list of genes which included four immune-related phenotypes. Six genes from the graft-versus-host disease and allograft rejection pathways overlap with the list of genes with novel eQTL, five of which belongs to the Human Leukocyte Antigen (HLA) class. Furthermore, all genes overlapping with the type I diabetes and allograft rejection pathways come from the HLA class. Genes encoded by the more than two hundred *HLA* genes in the highly polymorphic HLA region play an important role in antigen processing and presentation. In total, 20 eGenes in this study come from the HLA classes I and II (N=18) and from the HLA class IB (N=2) genes. Half of these eQTL-*HLA* gene associations were found to be novel. Out of the 5 *trans*- and 17 *cis*-eQTL, 3 *trans*- and 8 *cis*-eQTL were validated by published studies. The immune-related pathways enriched for novel eQTL-eGene pairs in this study demonstrate that novel eQTL could be identified in disease populations since their gene expression profiles differ from the profiles of healthy individuals.

Overall, the validation rate of eQTL in this study was comparable to that seen in published studies (Joehanes et al., 2017, Westra et al., 2013). A third of *cis*-eQTL validated in the two population-based studies by Westra et al. and Joehanes et al. whilst 90% of *cis*-eQTL were validated by the GTEx Project. In comparison, 25% to 58% of *cis*-eQTL reported by Joehanes et al. were validated by other eQTL studies, with the lowest validation rate coming from the validation between their study that used expression array data (Joehanes et al., 2017) and one that used RNA sequencing (Battle and Montgomery, 2014). Fifteen percent of *trans*-eQTL-eGene pairs validated in either the GTEx project or the study of Joehanes et al. *Trans*-eQTL validation rates are generally much lower (under 10%) than validation rates of *cis*-eQTL (Westra et al., 2013, Joehanes et al., 2017). However, the validation rate reported by this study and possibly other studies could underestimate the true validation rate between eQTL studies, since the list of variants and genes that passed quality control and were tested by any one study are usually not made available. Consequently, the number of eQTL-eGene

pairs eligible to be validated is overestimated when all genetic variants and genes present on the array are assumed to have been available for analysis. Apart from overestimating the number of eQTL-eGene pairs eligible for validation, imperfect validation rates between two eQTL studies can arise due to other factors such as the difficulty in accurately matching gene expression levels measured with different platforms, suggested by the higher validation rates between two array-based studies than between array-based- and RNA sequencing – based studies (Joehanes et al., 2017). This was true of this study and validation rate was higher with the RNA sequencing-based GTEx Project than with the other two, array-based studies. Differences between the study populations being compared can also give rise to imperfect validation and also to novel discoveries, as all genetic effects depend on both the genetic (epistasis) and environmental context of the population they were estimated in and therefore do not necessarily apply to another population or the same population at a different time. A more accurate way of identifying novel eQTL in this study would have been to contrast PAH eQTL effects with eQTL effects estimated in a control population assayed on the same platforms and in one batch with the PAH samples to reduce the variability due to technical factors. Also, novel eQTL are yet to be replicated in an independent population of PAH patients to identify true positives.

The effects of eQTL can vary by the tissue and even cell type under investigation (Powell et al., 2012, Andiappan et al., 2015) and may be modified by external and environmental factors (Barreiro et al., 2012, Fairfax et al., 2014). Previous studies have successfully identified so-called 'response' eQTL that are associated with variation in gene expression levels in cells under one of two experimental conditions, but not both. For example, in the study of Barreiro nearly two-hundred eQTL were identified in primary dendritic cells from sixty-five individuals with effects specific to either the Mycobacterium tuberculosis-infected cells or the uninfected ones (Barreiro et al., 2012). These response eQTL were argued to constitute natural regulatory variation that likely affect host - Mycobacterium tuberculosis interaction and account for interindividual variation in the immune response and susceptibility of tubercolosis. The authors of the study found when integrating their data with genome-wide genetic susceptibility of pulmonary tuberculosis that these response eQTL were more likely to be associated with the disease, uncovering potential susceptibility genes in pulmonary tubercolosis. Another study by Fairfax et al. (Fairfax et al., 2014) investigated the effect of innate immune stimuli on eQTL effects by exposing primary CD14+ cells from 432 European volunteers to the inflammatory cytokine interferon-γ or the endotoxin lipopolysaccharide. Inflammatory stimulation revealed hundreds of eQTL specific to either stimulus, which were enriched for disease-risk loci. In this study, the proportion of eQTL also associated with lung diseases or lung function was twice as high in the novel eQTL subset than in the validated subset, highlighting

the importance of genotype – environment interaction in understanding genetic variation of disease susceptibility characterized by genome-wide association studies.

6.5 Conclusions

To date, large-scale eQTL-mapping have been done in healthy individuals from population-based studies (Joehanes et al., 2017, Westra et al., 2013, Zhernakova et al., 2017) providing a valuable knowledge base for understanding the mechanisms behind associations of genetic variation and disease. It has been shown that recapitulating the environmental context relevant to disease, it is possible to more extensively decipher genetic variation of disease susceptibility (Fairfax et al., 2014, Barreiro et al., 2012, Çalışkan et al., 2015). Genome- and transcriptome-wide eQTL-mapping in this cohort of idiopathic and heritable PAH patients identified hundreds of potentially novel eQTL with twice the proportion of lung disease associated genetic variants than eQTL validated by population-based studies. Apart from pulmonary conditions, these novel eQTL specific to or active in PAH could be useful in understanding genetic risk factors for other diseases that share common mechanisms with PAH, such as those with immune dysregulation.

7 Conclusions and future work

The role of iron deficiency in the development of pulmonary arterial hypertension (PAH) and disease severity was investigated in a cross-sectional observational study, an instrumental variable analysis using genetic variants as proxy measures and a RCT; all of which found no evidence for a clinically significant effect of iron availability on PAH. This thesis emphasized the need for robust criteria for diagnosing iron deficiency in PAH and other PH subgroups. QTL for intermediate phenotypes, such as gene expression, proteins and metabolites, can offer a handle on understanding the mechanisms behind genetic associations of disease susceptibility. Genetic effects of gene expression in PAH largely overlapped with those described in disease-free populations; however, the relevance of the set of novel eQTL to lung-related phenotypes and diseases supports the rationale behind mapping QTL for intermediate phenotypes in disease populations.

Previous investigations into the phenomenon of iron deficiency in PAH focused on IPAH and heritable PAH cases and used healthy volunteers (Rhodes et al., 2011a) or patients with chronic thromboembolic pulmonary hypertension (CTEPH) (Soon et al., 2011) as controls. In the chapter titled 'Iron deficiency and anaemia in pulmonary hypertension diagnostic groups: a cross-sectional study in the National Pulmonary Hypertension Service, Hammersmith Hospital, iron deficiency prevalence and clinical outcomes were investigated using three different iron deficiency criteria and iron status biomarkers in a heterogenous cohort of patients representing five PH diagnostic groups. Iron deficiency was highly prevalent in IPAH and heritable PAH, but also in other PH classes including CTEPH patients and disease controls, contrary to the findings of previous studies (Soon et al., 2011, Rhodes et al., 2011a). However, it is important to note, that the control group in this study comprised of patients who were referred to the pulmonary hypertension clinic but were not diagnosed with PH. This, possibly heterogenous, group of patients could have had conditions that confounded effect estimates of diagnosis in these analyses.

The Mendelian Randomisation (MR) analysis of RDW, a correlate of iron status biomarkers, and IPAH and heritable PAH in chapter four showed no evidence of a causal relationship with RDW using all known RDW QTL or specifically those with strong evidence of involvement in regulating systemic iron status discovered in population-based studies as instruments for iron status. This analysis was powered to detect a causal effect (OR = 1.25 - 1.52) smaller than the observed estimate of the RDW-PAH association in this study (OR = 1.80 – 2.01 per 1.4% RDW) and another recent phenome-wide association study of RDW (OR = 1.75-2.4 per 1% RDW) (Thayer et al., 2019). The MR causal estimates

of RDW-PAH suggest either no effect or a much more modest causal effect of iron status than those observed in standard epidemiological investigations. However, the sample size required to detect smaller causal effects, like those reported for iron status by other MR studies (Gill et al., 2017, Gill et al., 2018, Pichler et al., 2013), is nearly ten times that of the largest PAH genome-wide association study (Rhodes et al., 2018) used in the MR analyses in this work. In agreement with the MR study, the first RCT presented in chapter five found no clinically significant impact of intravenous iron supplementation in iron deficient IPAH patients on markers of disease severity at 12 weeks. The treatment was well tolerated and effective at restoring iron status, suggesting it was well performed and that positive effects of iron treatment in previous studies are not likely to be through direct effects on pulmonary vascular tension or remodeling. Peripheral effects such as improved muscle function cannot be excluded (Charles-Edwards et al., 2019). Despite the potential weaknesses of the RCT and the MR study (i.e. lack of washout period in the clinical trial, no power to detect moderate effects in the MR study), the two approaches finding no evidence of a causal effect makes it more convincing that iron deficiency is not a therapeutic target in PAH and therefore not a priority for future clinical studies.

While the rapid development of MR methodologies provided robust ways of estimating causal effects between an exposure and an outcome (Burgess et al., 2020, Sanderson et al., 2018, Zheng et al., 2017), a prerequisite to MR is the availability of well-defined, genetically controlled phenotypes to serve as exposures and outcomes. The genetic control of a rich variety of complex phenotypes and biomarker measurements has been characterized in large populations utilizing electronic health records or deidentified databases (Bycroft et al., 2018, Crawford et al., 2014). Gene expression QTL have been extensively described in recent years in unaffected individuals from population-based studies (Westra et al., 2013, Joehanes et al., 2017, Zhernakova et al., 2017), whereas eQTL mapping in disease cohorts lags behind. The molecular environment specific to certain disease states could mask or unmask eQTL effects and specific environmental stimuli (such as viral and bacterial infections or inflammatory cytokines) have been shown to alter the effects of certain QTL in human cell studies . In the last chapter of this thesis, genetic determinants of gene expression in IPAH and heritable PAH were mapped and compared to eQTL from population-based studies to identify PAH-specific eQTL. Even though the number of GWAS and disease susceptibility loci continually increases with increasing affordability of genotyping and sequencing technologies, linking these common susceptibility variants to mechanisms of disease remains a major challenge (Visscher et al., 2012). Intermediate phenotypes such as gene expression, protein levels, metabolites and more are being integrated with genomic data to identify molecular patterns associated with disease (Hasin et al., 2017) and have been proven useful in translating disease susceptibility loci into biological insight

(Westra et al., 2013, Battle et al., 2015, Nath et al., 2017). In this PAH eQTL mapping study, the proportion of potentially novel eQTL – those not reported by population-based eQTL studies – that overlap loci reported for lung-related phenotypes was twice as high as that of the validated eQTL set. This highlights the importance of environment and genetic background whereby genetic variants contributing to disease susceptibility often exert their effect only under certain environmental and/or genetic conditions. Expression QTL studies have demonstrated that so-called 'response' eQTL that only show an effect in response to viral (Çalışkan et al., 2015), bacterial (Barreiro et al., 2012) or immune (Fairfax et al., 2014) stimuli are abundant and colocalize with susceptibility loci of the diseases related to the stimuli.

There could be alternative ways of identifying response eQTL relevant to PAH and related diseases. Even though mimicking the whole array of mechanisms that constitutes PAH is not possible, mapping eQTL in healthy individuals before and after exposing them to hypoxia (e.g. in populations living at altitude who change altitude at different times of year for work or study) could identify genetic variants that only have an effect on gene expression in the context of the molecular environment typical of low oxygen conditions. This could also be achieved using human cells where hypoxia can be induced with Cobalt (II) Chloride hexahydrate (Wu and Yotnda, 2011), a chemical inducer of hypoxia-inducible factor-1.3. Since loss-of-function mutations in *BMPR2* increase susceptibility to PAH and *BMPR2* levels are reduced in the lungs of heritable, idiopathic and – to a smaller extent – secondary PAH patients (Atkinson et al., 2002), transcriptome-wide gene expression measured before and after *BMPR2* knockdown in a cell type that expresses *BMPR2* could also identify eQTL active in a molecular environment that is characteristic of PAH. Since eQTL generally explain a large proportion of variance in gene expression, hundreds of response eQTL may be identified in under a hundred samples (Barreiro et al., 2012). Such context dependent eQTL could help in linking disease susceptibility loci with unclear functions to candidate causal genes and biological mechanisms.

Diagnosing iron deficiency in patients with chronic disease is challenging as routine iron status biomarkers are affected by iron deficiency and anaemia of chronic disease in a similar way. Also, iron deficiency and anaemia of chronic disease may simultaneously be present, further complicating the diagnosis. Testing the performance of iron status biomarkers and their combinations, ferritin and soluble transferrin receptor (sTfR) in particular, in diagnosing true iron deficiency identified by bone marrow staining in PAH patients would provide a useful reference for future investigations. However, bone marrow biopsy is an invasive procedure and might not be acceptable in PAH patients without additional justification for its need. In the absence of such evidence in PAH, I relied on the results of studies that evaluated iron biomarkers in rheumatoid arthritis, other chronic diseases and

malignancies (Means Jr et al., 1999, Suominen et al., 2000, Bultink et al., 2001) and used the ratio of sTfR and ferritin shown to maximize sensitivity and specificity in disease populations. Interestingly, higher than normal sTfR levels (> 28.1 nmol/L) together with normal ferritin (> 100 µg/L) and transferrin saturation (> 20%) as defined in the 'sTfR-only' iron deficiency group was associated with much worse prognosis, especially in the H/IPAH diagnostic group, but sTfR levels alone were not. These results suggest that a high-risk patient group could be identified based on higher than normal sTfR together with normal iron biomarker levels - which could result from inflammation if residual confounding remains after controlling for interleukin-6 levels -, although the mechanisms underlying elevated sTfR in the apparent absence of iron deficiency in this patient group should be further investigated. A possible explanation could be worsening hypoxia (i.e. reduced oxygen delivery) as the disease progresses and/or haemolysis related to PAH (Barnett et al., 2008), both of which can stimulate reticulocytosis and thus lead to increased sTfR. Reticulocytosis could also explain the strong association of high RDW and PAH susceptibility reported in previous studies (Thayer et al., 2019, Rhodes et al., 2011b) and in chapter four of this thesis. Reticulocyte counts and/or erythropoietin measurements could be used to test whether reticulocytosis is prominent in patients in the sTfR-only group and if it separates them from the other sTfR-based, but severely iron deficient group. Further measurements on arterial blood gases, such as arterial oxygen and carbon dioxide pressures, could also be useful in assessing hypoxemia/tissue hypoxia in these two groups.

PAH is a rare but complex disease with genetic and environmental risk factors. Even though much progress has been made in recent years to uncover genetic determinants of PAH (Graf et al., 2018, Rhodes et al., 2018), epidemiological investigations into the environmental component of the disease are limited by the available sample size and the biases of the case-control design and the difficulty to identify appropriate controls that represent the population at risk of developing PAH. For example, healthy controls are often recruited from individuals visiting primary care and living in the neighborhood of the hospital which is also the referral hospital for patients with a rare disease. A large proportion of these referred patients with the rare disease travel from other areas of the country and could differ with regards to, for instance, socioeconomic factors from the recruited control population, leading to spurious associations (Morabia, 1997). Given the reduced penetrance of heritable PAH and IPAH (~42% in women and ~14% in men carrying known mutations) (Machado et al., 2001), following up families of affected individuals, especially relatives harbouring mutations associated with PAH, could identify environmental modifiers of genetic risk. Examples of populations at risk of developing PAH are relatives of patients with pathogenic *BMPR2* mutations, individuals with HIV or schistosomiasis infection or systemic sclerosis. The longitudinal follow-up of such cohorts would be an invaluable source to identify environmental triggers of PAH.

As genotyping is becoming increasingly accessible to the wider research community, extending international collaborations and sharing summary-level data in the future could increase the precision of common genetic effects of PAH susceptibility, allowing the investigation of shared genetic aetiologies and causal mechanisms between PAH and a wide range of phenotypes characterized by large population-based studies like that of the UK Biobank. Well-designed Mendelian Randomisation studies could be applied to other phenotypes associated with PAH, such as metabolic disturbances and diabetes (Poms et al., 2013), proteins including vascular endothelial growth factor (Voelkel and Gomez-Arroyo, 2014), and inflammation (Tian et al., 2019, Kumar and Graham, 2018), which may require genetic effects to be defined in a larger than currently available PAH population and improved definitions of QTL. The relevance of PAH eQTL to other lung-related phenotypes suggest it could be worth exploring the genetic variablilty of gene expression (and other intermediate phenotypes) in disease populations to unmask regulatory QTL which cannot easily be found in healthy, population-based studies.

The main conclusion of this thesis is that although iron deficiency is prevalent in IPAH and other PH subclasses, no evidence for a causal relationship or a clinically meaningful effect of iron supplementation in IPAH patients was found. Approaches using genetic epidemiology, such as Mendelian Randomization and estimating genetic correlation between PAH and other candidate phenotypes should be applied to investigate causal and epidemiological relationships in the future, which will be made more powerful by increasing the accuracy of PAH genetic effects through collaboration, and by improving QTL definitions of phenotypes relevant to PAH. Genetic variants influencing gene expression in IPAH but not in healthy individuals were highly relevant for lung-associated traits and diseases, highlighting the usefulness of mapping quantitative trait loci in disease cohorts.

8 Appendix

8.1 Iron deficiency and anaemia in pulmonary hypertension diagnostic groups: a cross-sectional study in the National Pulmonary Hypertension Service, Hammersmith Hospital

Supplementary Table 8.1.1 Characteristics of patients referred to the National Pulmonary Hypertension Service at Hammersmith Hospital.
Values reported are median and interquartile range unless otherwise specified. Definition of abbreviations: No PH = disease controls; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; PAH CTD = pulmonary arterial hypertension associated with connective tissue disease; PAH CHD = pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension; BNP = B-type natriuretic peptide; NTproBNP = N-terminal pro b-type natriuretic peptide; IL-6 = interleukin-6; 6MWD = six-minute walk distance test; PVR = pulmonary vascular resistance; mPAP = mean pulmonary arterial pressure

	No PH		H/IPAH		PAH CTD		PAH CHD		PH LHD		CTEPH	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
N (%)	116 (72)	45 (28)	131 (73)	48 (27)	40 (78)	11 (22)	19 (59)	13 (41)	49 (60)	32 (40)	120 (48)	128 (52)
Age at diagnosis, mean (SD)	57.95 (16.72)	57.22 (14.55)	49.25 (18.19)	57.61 (18.02)	56.9 (14.92)	64.62 (13.23)	46.3 (15.88)	52.95 (20.25)	70.4 (11.95)	73.67 (8.35)	61.85 (16.51)	62.43 (13.83)
Age at sample, mean (SD)	58.04 (16.87)	57.36 (14.41)	52.12 (17.43)	59.65 (16.26)	58.96 (13.97)	65.47 (12.88)	49.99 (16.08)	55.28 (19.36)	70.91 (11.14)	73.64 (8.35)	64.27 (16.17)	64.63 (13.73)
WHO class, N (%)												
available	88 (76)	36 (80)	107 (82)	39 (81)	34 (85)	7 (64)	18 (95)	11 (85)	41 (84)	28 (88)	108 (90)	119 (93)
I	9 (10)	8 (22)	3 (3)	3 (8)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2)	1 (4)	5 (5)	10 (8)
II	24 (27)	8 (22)	19 (18)	7 (18)	11 (32)	0 (0)	3 (17)	4 (36)	5 (12)	3 (11)	17 (16)	37 (31)
III	54 (61)	20 (56)	75 (70)	26 (67)	19 (56)	4 (57)	12 (67)	6 (55)	32 (78)	24 (86)	73 (68)	66 (55)
IV	1 (1)	0 (0)	10 (9)	3 (8)	4 (12)	3 (43)	3 (17)	1 (9)	3 (7)	0 (0)	13 (12)	6 (5)
Ethnicity, N (%)												
available	69 (59)	28 (62)	110 (84)	34 (71)	32 (80)	10 (91)	14 (74)	10 (77)	32 (65)	21 (66)	105 (88)	118 (92)
white	50 (72)	23 (82)	95 (86)	31 (91)	25 (78)	9 (90)	12 (86)	9 (90)	29 (91)	18 (86)	93 (89)	107 (91)
South Asian	7 (10)	0 (0)	12 (11)	2 (6)	3 (9)	1 (10)	2 (14)	1 (10)	2 (6)	2 (10)	0 (0)	2 (2)
black	12 (17)	5 (18)	3 (3)	1 (3)	4 (12)	0 (0)	0 (0)	0 (0)	1 (3)	1 (5)	12 (11)	9 (8)
Anticoagulant at the time of sample, %	28	47	64	58	37	67	47	60	64	52	100	100
BNP *ng/L*	46.05 (19.08-88.22)	28.8 (10.6-83.18)	85.3 (29.9-199.6)	268.1 (126.8-614.8)	136.5 (48.3-423.6)	242.25 (78.43-964.12)	120.3 (71.45-229.43)	169.4 (64.9-303.32))	253 (163.52-392.5)	378 (174.65-601.55)	89.3 (38.03-265)	82.6 (42.95-335.1)
NTproBNP *pmol/L*	186.2 (77.2-797.2)	290.9 (79.8-792.4)	403.98 (102.23-1056.1)	666.65 (103.4-1677.2)	387.3 (193.78-1576.2)	1190.95 (367.12-2112)	670.3 (268.55-1013.4)	944.17 (216.7-1603)	1278.15 (704.9-1701.13)	1200.55 (546.03-1653.8)	392.75 (84.92-1306.33)	477.2 (135.83-1555.9)

	No PH		H/IPAH		PAH CTD		PAH CHD		PH LHD		CTEPH	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
6MWD *m*	336 (213-408)	397.5 (323.25-450)	336 (120-390)	284 (108-419.5)	240 (105-362.25)	228 (39-413.25)	336 (120-387)	320 (240-340)	125 (48-312)	297 (156-356.5)	240 (96-372)	321.5 (198-427)
Haemodynamics available, %	35	40	86	94	78	82	42	54	78	72	83	90
PVR Wood units	1.74 (1.34-2.41)	1.16 (0.77-1.54)	11.46 (5.9-15.33)	8.4 (5.67-12.3)	7.38 (4.94-10.55)	10.14 (7.76-11.52)	4.47 (2.05-10.32)	2.86 (2.59-11.26)	3.88 (2.55-7.85)	2.33 (1.71-5.2)	8.29 (5.42-12.43)	7.85 (4.36-10.97)
mPAP mmHg	20 (18-23)	23 (20.25-24)	53 (42-60)	50 (40-60)	43 (32.5-50)	40 (39-47)	40.5 (28.8-54.8)	47 (35.5-54.5)	35 (28-44.5)	31 (29.5-42.5)	45 (34-51.25)	44 (35-55)

	Replete	Ferr-TSAT	sTfR	Ferr-TSAT + sTfR
N	258	93	193	208
Female, N (%)	143 (55)	77 (83)	107 (55)	148 (71)
Age at sample, mean (SD)	59.6 (15.72)	59.66 (18.56)	61.15 (14.9)	61.73 (17.7)
WHO class, N (%)				
available	215 (83)	69 (74)	175 (91)	177 (85)
1	19 (9)	4 (6)	11 (6)	6 (3)
2	51 (24)	14 (20)	43 (25)	30 (17)
3	132 (61)	49 (71)	106 (61)	124 (70)
4	13 (6)	2 (3)	15 (9)	17 (10)
Ethnicity, N (%)				
available	198 (77)	69 (74)	159 (82)	157 (75)
white	183 (92)	65 (94)	123 (77)	130 (83)
South Asian	5 (3)	3 (4)	11 (7)	15 (10)
black	10 (5)	1 (1)	25 (16)	12 (8)
Anticoagulant, %	65	55	63	70
Iron µmol/L	16.5 (14-20)	11 (9-12)	15 (12-19)	8.5 (6-11)
sTfR nmol/L	22.4 (19.3-25.5)	23.9 (21.6-25.7)	35.8 (31.3-44.1)	42.2 (33.72-52.35)
Log sTfR/ferritin ratio	-1.61 (-2.15--0.91)	-0.81 (-1.16--0.35)	-1.17 (-1.72--0.66)	0.34 (-0.22-1.04)
Transferrin g/L	2.6 (2.38-2.8)	2.8 (2.6-3.1)	2.6 (2.23-2.9)	3.1 (2.7-3.4)
TSAT %	26 (22-33)	16 (14-18)	24 (20-31)	12 (8-16)
Ferritin µg/L	104 (55.25-183)	54 (35-72)	128 (81-212)	27 (17-50.25)
Hepcidin ng/mL	17.15 (9.2-32.7)	10.6 (5-18.5)	16.2 (6.6-36.6)	2.55 (0.9-6.82)
Haemoglobin g/dL	14.2 (13.03-15.28)	13.2 (12.2-14.5)	13.8 (12.1-15.2)	12.7 (11.4-14.03)
Haematocrit %	0.43 (0.39-0.45)	0.41 (0.38-0.43)	0.42 (0.38-0.45)	0.4 (0.36-0.43)
RDW %	14 (13.3-14.6)	14.4 (13.6-15)	15.1 (14.2-16.5)	15.9 (14.5-17.1)
BNP ng/L	79.6 (32.02-253.6)	78.1 (41.95-242.8)	86.6 (35.02-268.4)	102 (48.45-309.5)
NTproBNP pmol/L	472.5 (100-1302.35)	316.24 (101.87-945.85)	585 (148.92-1480.6)	459.24 (122.7-1567.9)
IL-6 pg/mL	3 (1.6-5.65)	3.3 (1.7-5.1)	4.68 (2.52-8.25)	4.9 (3.08-9)
6MWD m	354 (221.5-432)	320 (144-393)	288 (132-384)	240 (96-364)
PVR available, N (%)	183 (71)	71 (76)	137 (71)	157 (75)
PVR Wood units	6.74 (2.58-11.22)	6.3 (2.56-12.38)	7.38 (3.4-12.13)	7.76 (3.54-11.52)
mPAP available, N (%)	183 (71)	72 (77)	138 (72)	157 (75)
mPAP mmHg	44 (31.5-52)	40.5 (28-54)	44 (30.25-53)	44 (32-53)

Supplementary Table 8.1.2 Characteristics of patients referred to the National Pulmonary Hypertension Service at Hammersmith Hospital by iron status.

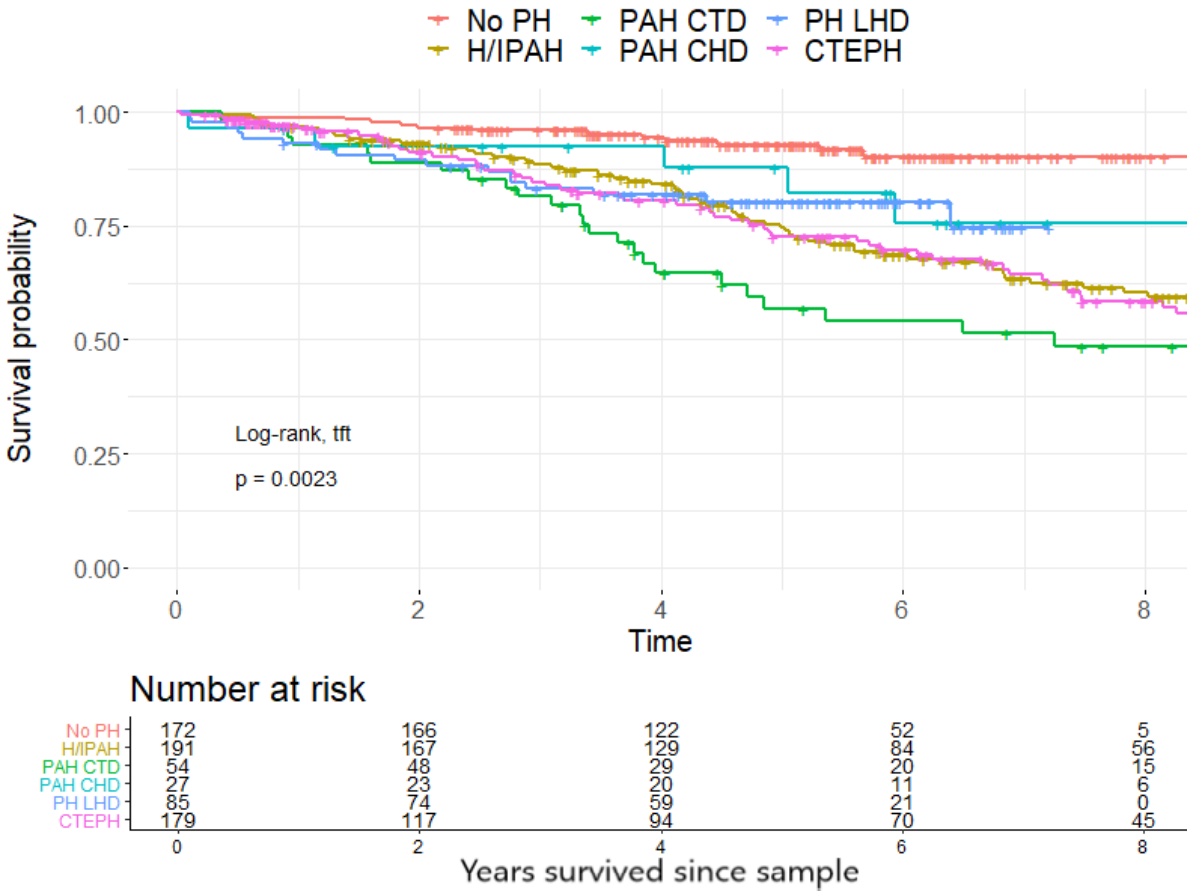
Values reported are median and interquartile range unless otherwise specified. Definition of abbreviations: Ferr = ferritin; sTfR = soluble transferrin receptor; TSAT = transferrin saturation; RDW = red cell distribution width; BNP = B-type natriuretic peptide; NTproBNP = N-terminal pro b-type natriuretic peptide; IL-6 = interleukin-6; 6MWD = six-minute walk distance test; PVR = pulmonary vascular resistance; mPAP = mean pulmonary arterial pressure.

	Replete	Ferr-TSAT	sTfR	Ferr-TSAT + sTfR
N	183	65	123	130
Female, N (%)	95 (52)	55 (85)	65 (53)	89 (68)
Age at sample, mean (SD)	59.79 (15.97)	59.04 (18.64)	63.36 (13.81)	65.21 (16.65)
available	164 (0.9)	56 (0.86)	115 (0.93)	118 (0.91)
1	16 (0.1)	3 (0.05)	5 (0.04)	6 (0.05)
2	37 (0.23)	10 (0.18)	28 (0.24)	18 (0.15)
3	99 (0.6)	41 (0.73)	72 (0.63)	85 (0.72)
4	12 (0.07)	2 (0.04)	10 (0.09)	9 (0.08)
Anticoagulant, %	69	62	70	76
Iron µmol/L	16 (13.5-21)	11 (9-12)	14 (11-19)	9 (6-11)
sTfR nmol/L	22.17 (19.32-25.5)	24 (20.8-26.3)	35.6 (30.6-40.45)	40.89 (32.2-50.38)
Log sTfR/ferritin ratio	-1.53 (-2.13--0.8)	-0.7 (-1.15--0.34)	-1.22 (-1.76--0.74)	0.28 (-0.28-0.99)
Transferrin g/L	2.6 (2.3-2.8)	2.8 (2.6-3)	2.6 (2.3-2.9)	3.05 (2.7-3.3)
TSAT %	26 (22-33)	17 (14-19)	24 (19-31)	13 (8-17)
Ferritin µg/L	96 (51-183)	49 (32-66)	127 (76-211)	27 (18-51.75)
Hepcidin ng/mL	15.8 (8.53-31.08)	9.2 (3.9-15.9)	17 (6.05-37.05)	2.5 (0.9-6.38)
Haemoglobin g/dL	14.4 (13.4-15.5)	13.3 (12.4-14.2)	14.1 (12.85-15.5)	12.8 (11.6-14.17)
Haematocrit %	0.43 (0.4-0.46)	0.41 (0.38-0.43)	0.43 (0.4-0.47)	0.4 (0.37-0.43)
RDW %	14 (13.3-14.5)	14.4 (13.8-15.1)	15 (14.3-15.9)	15.8 (14.6-16.9)
BNP ng/L	79.4 (36.6-252.2)	78.1 (39.75-216.55)	126.2 (41.8-412.75)	112.8 (54.4-299.5)
NTproBNP pmol/L	464.8 (99.73-1199.85)	396.42 (136.46-1178.6)	577.5 (159.1-1480.6)	458.9 (103.75-1466.4)
IL-6 pg/mL	2.98 (1.65-5.6)	3.2 (1.6-5.7)	5.4 (3-9.8)	5.5 (3.45-9.2)
6MWD m	360 (240-444)	304 (127.5-397)	280 (120-381)	240 (90-355)
PVR available %	135 (74)	53 (82)	97 (79)	105 (81)
PVR Wood units	6.9 (2.65-10.5)	6.39 (2.94-11.62)	7.95 (3.05-12.14)	7.76 (3.8-11.65)
mPAP available %	135 (74)	53 (82)	97 (79)	105 (81)
mPAP mmHg	44 (32.5-52)	43 (30-54)	45 (31-54)	43 (32-53)

Supplementary Table 8.1.3 Characteristics of patients from a white background referred to the National Pulmonary Hypertension Service at Hammersmith Hospital by iron status.
Values reported are median and interquartile range unless otherwise specified. Ferr = ferritin; sTfR = soluble transferrin receptor; TSAT = transferrin saturation; RDW = red cell distribution width; BNP = B-type natriuretic peptide; NTproBNP = N-terminal pro b-type natriuretic peptide; IL-6 = interleukin-6; 6MWD = six-minute walk distance test; PVR = pulmonary vascular resistance; mPAP = mean pulmonary arterial pressure.

	Replete	Ferr-TSAT	sTfR	Ferr-TSAT + sTfR
N	15	4	36	27
black	10	1	25	12
South Asian	5	3	11	15
Female, N (%)	10 (67)	3 (75)	24 (67)	21 (78)
Age at sample, mean (SD)	55.65 (15.8)	49.85 (18.07)	52.91 (15.84)	46.63 (14.43)
available	15 (1)	4 (1)	34 (0.94)	25 (0.93)
1	1 (0.07)	1 (0.25)	4 (0.12)	0 (0)
2	4 (0.27)	2 (0.5)	11 (0.32)	8 (0.32)
3	10 (0.67)	1 (0.25)	16 (0.47)	12 (0.48)
4	0 (0)	0 (0)	3 (0.09)	5 (0.2)
Anticoagulant, %	60	0	50	39.13
Iron µmol/L	15 (13.5-17.5)	12 (11.5-12)	16 (13.75-18.25)	8 (6-11)
sTfR nmol/L	22.6 (20.77-24.75)	25.82 (24.71-26.45)	48.21 (32.66-64.85)	45 (34.72-57.05)
Log sTfR/ferritin ratio	-1.41 (-1.95--1.01)	-1.07 (-1.18--0.69)	-0.93 (-1.36--0.57)	0.59 (0.09-1.31)
Transferrin g/L	2.35 (2.02-2.68)	2.85 (2.63-3.08)	2.5 (1.9-2.8)	3.1 (2.68-3.48)
TSAT %	30 (24.5-32.5)	15.5 (14.75-17)	25 (22.5-33.75)	12 (8-13.5)
Ferritin µg/L	104 (69.5-167.5)	70 (56.25-76)	121 (90.5-217.5)	21 (13-48)
Hepcidin ng/mL	15 (8.3-20)	15.68 (8.82-22.57)	11.55 (6.5-34.65)	2.8 (1.35-5.4)
Haemoglobin g/dL	12.5 (11.95-13.4)	13.15 (12.65-13.58)	12 (10.55-13.62)	12.5 (10.9-13.5)
Haematocrit %	0.38 (0.36-0.4)	0.39 (0.38-0.4)	0.37 (0.31-0.42)	0.38 (0.34-0.41)
RDW %	14.9 (13.5-15.6)	14.25 (13.95-14.53)	15.95 (14.5-18.02)	15.7 (14.5-17.8)
BNP ng/L	99.8 (25.7-436.17)	55.2 (32.6-116.4)	48 (23.82-123.55)	22 (13.5-63.3)
NTproBNP pmol/L	656.5 (110.95-1045.1)	113.4 (92.27-144.42)	915.9 (135.72-1551.33)	376.7 (115.45-1179.24)
IL-6 pg/mL	3.6 (2.4-6.7)	2.74 (1.33-4.35)	3.5 (2.5-6.7)	3.7 (2.25-4.9)
6MWD m	345 (172-360)	240 (164-345.75)	336 (194-420)	291 (176-412.5)
PVR available %	13 (86.67)	3 (75)	20 (55.56)	19 (70.37)
PVR Wood units	5.36 (2.73-7.69)	1.34 (0.67-7.36)	5.08 (2.9-10.15)	8.01 (4.83-13.88)
mPAP available %	13 (86.67)	4 (100)	20 (55.56)	19 (70.37)
mPAP mmHg	41 (28-50)	55 (37.75-63.75)	32 (27.5-47.75)	46 (43-54.5)

Supplementary Table 8.1.4 Characteristics of patients from a non-white background referred to the National Pulmonary Hypertension Service at Hammersmith Hospital by iron status.
Values reported are median and interquartile range unless otherwise specified. Ferr = ferritin; sTfR = soluble transferrin receptor; TSAT = transferrin saturation; RDW = red cell distribution width; BNP = B-type natriuretic peptide; NTproBNP = N-terminal pro b-type natriuretic peptide; IL-6 = interleukin-6; 6MWD = six-minute walk distance test; PVR = pulmonary vascular resistance; mPAP = mean pulmonary arterial pressure.



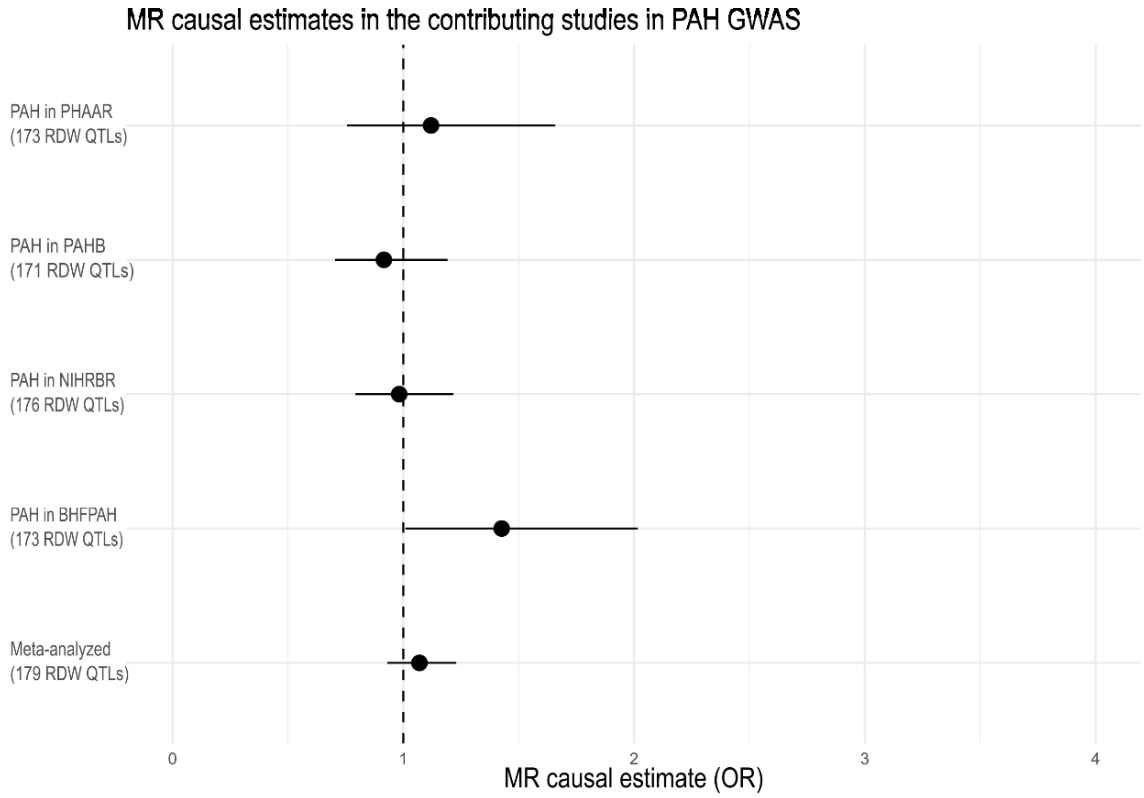
Supplementary Figure 8.1.1 Kaplan-Meier survival estimates stratified by diagnostic group.
Definition of abbreviations: No PH = disease controls; H/IPAH = heritable/idiopathic pulmonary arterial hypertension; PAH CTD = pulmonary arterial hypertension associated with connective tissue disease; PAH CHD = pulmonary arterial hypertension associated with congenital heart disease; PH LHD = pulmonary hypertension due to left heart disease; CTEPH = chronic thromboembolic pulmonary hypertension; tft = test for trend.

8.2 Mendelian randomisation analysis of red cell distribution width in pulmonary arterial hypertension

Variant information				RDW GWAS Astle et al.		Genetics of Iron Status GWAS Benyamin et al.								
Gene	Lead variant ID	Effect Allele	Allele Frequency in UKB and INTERVAL	beta RDW	*P* - value	Proxy variant ID*	beta iron	*P* – value iron	beta ferritin	*P* – value ferritin	beta TSAT	*P* – value TSAT	beta TF	P – value TF
HFE	rs144861591	C	0.92	0.21	6.50×10^{-216}	rs1800562	-0.37	4.0×10^{-77}	-0.21	1.4×10^{-29}	-0.58	1.5×10^{-178}	0.55	1.3×10^{-153}
TMPRSS6	rs855791	A	0.44	0.13	9.79×10^{-271}	-	-0.19	4.3×10^{-77}	-0.05	5.8×10^{-8}	-0.19	3.5×10^{-80}	0.04	1.3×10^{-4}
HFE	rs198851	G	0.85	0.13	2.55×10^{-161}	-	-0.19	1.6×10^{-40}	-0.06	3.6×10^{-6}	-0.23	4.7×10^{-59}	0.12	3.0×10^{-17}
TFRC	rs7619708	C	0.24	0.07	4.35×10^{-64}	rs6583288	0.03	1.2×10^{-2}	0.004	7.3×10^{-1}	0.05	3.8×10^{-6}	-0.06	3.8×10^{-8}
TFR2	rs9801017	G	0.37	0.05	9.40×10^{-37}	rs7385804	-0.06	7.2×10^{-8}	-0.02	2.5×10^{-2}	-0.05	1.8×10^{-7}	0.01	4.0×10^{-1}

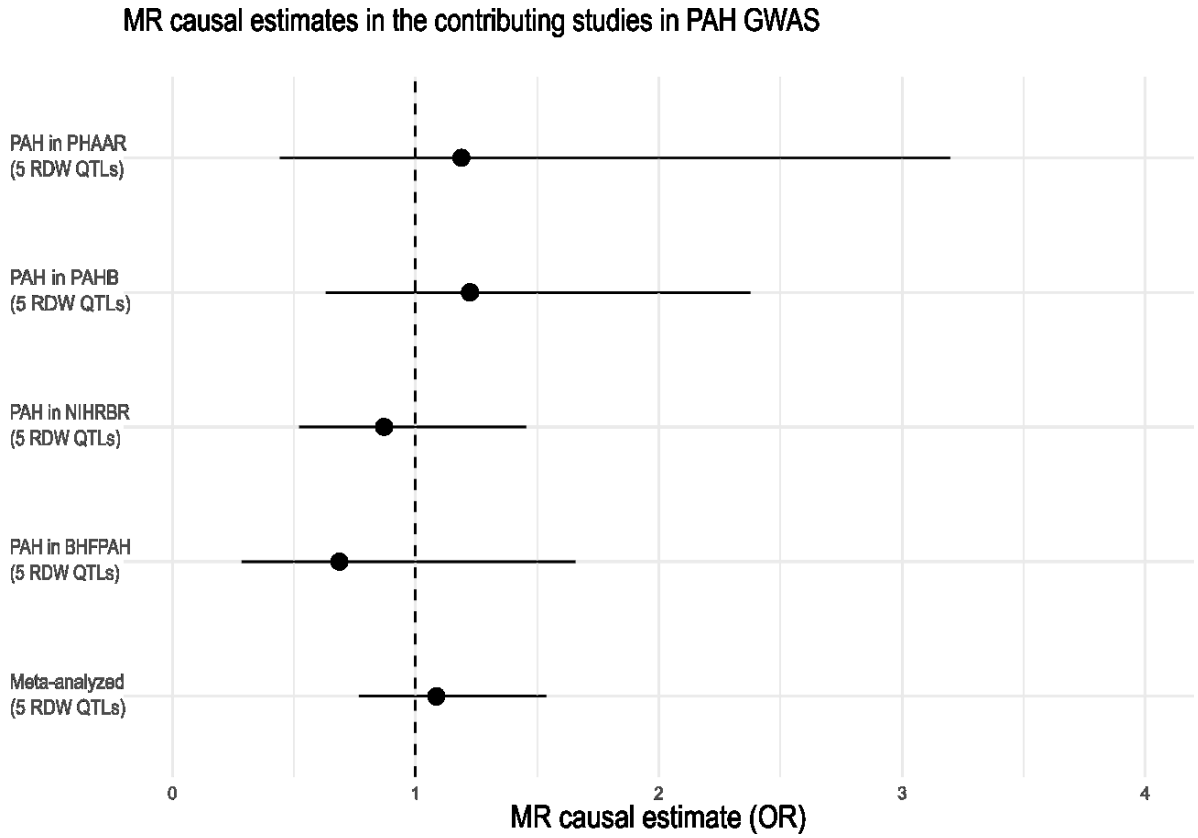
Supplementary Table 8.2.1 Five variants selected from the RDW GWAS based on their effects on systemic iron status.

*Elevated RDW can reflect iron deficiency which presents with decreased serum iron levels. Betas for RDW and the other iron biomarkers are reported in standard units. Definition of abbreviations: RDW = red cell distribution width; TSAT = transferrin saturation; TF = transferrin; GWAS = genome-wide association study. *Where the lead RDW variant for the locus is not available in the Genetics of Iron Status GWAS, the results of a suitable proxy variant in strong linkage disequilibrium (r2 ≥ 0.8) with the lead variant are listed.*



Supplementary Figure 8.2.1 RDW-PAH primary Mendelian Randomisation analysis in the individual contributing studies in the PAH GWAS.

Individual inverse variance weighted causal estimates for the primary MR analysis of association between RDW and PAH from the four contributing studies in PAH GWAS. The y-axis lists the study name and the number of available RDW quantitative trait loci (QTL) including suitable proxy variants with a minimum r^2 of 0.8 in the study. X-axis: PAH odds ratios per one standard unit increase in RDW (dot) with the corresponding lower and upper 95% confidence intervals (horizontal line). PHAAR: Pulmonary Hypertension Allele-Associated Risk (269 PAH cases, 1068 controls). PAHB: US National Biological Sample and Data Repository for Pulmonary Arterial Hypertension (694 PAH cases, 1560 controls). NIHRBR: UK National Institute for Health Research BioResource (847 PAH cases, 5048 controls). BHFPAH: British Heart Foundation Pulmonary Arterial Hypertension (275 PAH cases, 1983 controls).



Supplementary Figure 8.2.2 RDW-PAH secondary Mendelian Randomisation analysis in the individual contributing studies in the PAH GWAS.

Individual inverse variance weighted causal estimates for the secondary MR analysis of association between RDW acting via iron status and PAH from the four contributing studies in PAH GWAS. The y-axis lists the study name and the number of available RDW quantitative trait loci (QTL) including suitable proxy variants with a minimum r2 of 0.8 in the study. X-axis: PAH odds ratios per one standard unit increase in RDW (dot) with the corresponding lower and upper 95% confidence intervals (horizontal line). PHAAR: Pulmonary Hypertension Allele-Associated Risk (269 PAH cases, 1068 controls). PAHB: US National Biological Sample and Data Repository for Pulmonary Arterial Hypertension (694 PAH cases, 1560 controls). NIHRBR: UK National Institute for Health Research BioResource (847 PAH cases, 5048 controls). BHFPAH: British Heart Foundation Pulmonary Arterial Hypertension (275 PAH cases, 1983 controls).

8.3 Supplementation of iron in idiopathic pulmonary hypertension: a randomized controlled crossover trial

Variable		Iron - Placebo			Placebo - Iron		
	N	**Baseline**	**12 weeks**	**24 weeks**	**Baseline**	**12 weeks**	**24 weeks**
Iron indices							
Iron, umol/L	37	12.8 (10.3)	16.3 (9.3)	17.0 (8.7)	12.1 (10.4)	10.9 (8.3)	16.2 (5.4)
Ferritin, ug/L	37	17.0 (13.0)	146.0 (137.5)	115.0 (141.50)	18.0 (28.0)	14.0 (18.0)	134.5 (114.0)
TSAT, %	37	12.4 (15.3)	24.0 (17.8)	23.3 (23.3)	13.6 (12.5)	14.1 (10.4)	24.4 (11.6)
Haemoglobin, g/dL	38	12.4 (1.0)	15.0 (1.5)	15.0 (1.7)	12.5 (3.1)	13.6 (3.6)	14.6 (1.6)
sTfR, nmol/L	37	41.7 (22.8)	27.4 (17.5)	28.6 (18.8)	37.9 (21.3)	38.2 (16.0)	26.3 (8.5)
Hepcidin, ng/mL	37	3.5 (6.7)	43.6 (50.4)	31.8 (37.0)	1.3 (3.0)	1.0 (6.2)	20.5 (21.3)
Primary endpoint							
Endurance time, secs	29	272.0 (187.5)	253 (211.3)	384.0 (244.0)	255.0 (150.5)	216.0 (163.5)	273.0 (125.5)
Secondary endpoints							
Functional status							
6MWD, m	35	427.0 (65.0)	426.5 (71.5)	450.0 (92.0)	450.0 (71.5)	451.0 (107.5)	433.0 (119.3)
Dyspnea score (Borg index)	35	2.5 (2.8)	2.0 (5.25)	3.0 (5.0)	2.5 (2.0)	2.5 (1.5)	2.0 (2.4)
NYHA functional class, *no.*	37						
I		0	0	0	1	1	1
II		11	11	11	10	12	11
III		8	8	8	8	6	6
NTproBNP, ng/L	37	72.9 (218.2)	88.7 (96.6)	106.7 (233.4)	222.5 (396.9)	168.3 (677.0)	283.1 (300.9)
Hemodynamics							
Systolic blood pressure, mmHg	37	112.0 (15.5)	113.0 (12.0)	112.0 (10.0)	110.0 (17.5)	110.0 (18.5)	113.0 (18.2)
Diastolic blood pressure, mmHg	37	67.0 (13.0)	67.0 (16.5)	67.0 (19.0)	68.0 (10.5)	66.0 (8.5)	71.0 (15.8)
PVR, Wood units	17	7.3 (2.4)	6.2 (2.7)	-	5.9 (4.1)	4.7 (1.6)	-
mPAP, mmHg	17	45.5 (14.3)	45.0 (17.0)	-	50.0 (4.0)	54.0 (20.5)	-
mRAP, mmHg	17	8.5 (7.8)	6.0 (5.0)	-	10.0 (3.0)	9.5 (5.3)	-
mPCWP, mmHg	17	11.0 (3.3)	9.0 (7.0)	-	12.5 (4.0)	13.5 (7.5)	-
Cardiac index, L/min/m^2	17	2.6 (0.6)	2.7 (0.7)	-	2.5 (1.7)	2.9 (1.6)	-
Questionnaire (CAMPHOR)							
QoL score	31	6.0 (10.0)	5.0 (7.3)	5.0 (10.0)	4.5 (7.0)	5.0 (6.5)	6.0 (9.0)
Symptom score	31	8.0 (8.0)	7.0 (7.0)	9.0 (10.0)	8.0 (7.5)	8.5 (7.0)	6.0 (5.0)

Variable		Iron - Placebo			Placebo - Iron		
	N	Baseline	12 weeks	24 weeks	Baseline	12 weeks	24 weeks
Questionnaire (CAMPHOR)							
Activity score	31	8.0 (6.0)	8.0 (7.0)	10.0 (7.0)	9.5 (7.8)	8.0 (7.5)	8.0 (9.3)
ECPET							
VO_2 at 80% peak WR, L/min	29	1.14 (0.59)	1.15 (0.37)	1.19 (0.39)	1.14 (0.37)	1.07 (0.48)	1.14 (0.41)
VO_2 at 80% peak WR, mL/min/kg	29	14.82 (3.04)	15.67 (2.44)	15.62 (3.26)	14.97 (4.48)	14.61 (7.78)	15.23 (5.50)
VO_2 at 3mins at 80% peak WR, L/min	27	1.03 (0.39)	1.06 (0.50)	1.07 (0.26)	1.04 (0.29)	1.03 (0.34)	1.09 (0.21)
VO_2 at 3mins at 80% peak WR, mL/min/kg	27	13.12 (4.26)	14.53 (6.03)	14.22 (1.04)	13.50 (4.67)	13.80 (4.13)	15.15 (3.71)
ICPET							
Peak VO_2, L/min	33	1.17 (0.48)	1.14 (0.25)	1.14 (0.32)	1.10 (0.36)	1.14 (0.40)	1.20 (0.38)
Peak VO_2, mL/min/kg	33	14.95 (2.68)	15.24 (3.72)	14.56 (2.42)	14.49 (4.93)	15.28 (6.40)	15.38 (5.24)
VO_2 at metabolic threshold, L/min	33	0.72 (0.32)	0.75 (0.19)	0.74 (0.20)	0.89 (0.33)	0.87 (0.31)	0.78 (0.25)
VO_2 at metabolic threshold, mL/min/kg	33	10.06 (2.15)	9.78 (1.93)	9.64 (2.23)	10.13 (2.86)	10.18 (3.84)	9.92 (2.81)
VE/VCO_2 slope	34	40.90 (9.60)	38.70 (18.85)	40.00 (18.00)	37.80 (9.20)	40.00 (10.05)	36.70 (9.40)
VO_2/WR slope	29	9.55 (2.58)	8.80 (1.50)	9.20 (2.85)	8.45 (1.58)	8.05 (1.98)	7.80 (3.05)
Peak O_2 pulse	34	8.55 (4.23)	8.70 (3.60)	8.90 (2.50)	8.20 (2.80)	8.10 (3.05)	8.70 (2.20)

Supplementary Table 8.3.1 Primary and secondary endpoints of the Ferinject (Europe) clinical trial.
All continuous variables are median (interquartile range). Participants were randomized to receive either iron (Iron-Placebo) or saline (Placebo-Iron) at baseline in this crossover design. The Iron-Placebo group received intravenous iron at baseline and saline at 12 weeks. The Placebo-Iron group received saline at baseline and intravenous iron at 12 weeks. Column 'N' gives the total number of participants completing the assessment at every visit (baseline, 12 week, 24 week). Definition of abbreviations: TSAT = transferrin saturation; sTfR = soluble transferrin receptor; ECPET = endurance cardiopulmonary exercise test; 6MWD = six-minute walk distance; NTproBNP = N-terminal - pro hormone BNP; PVR = pulmonary vascular resistance; mPAP = mean pulmonary arterial pressure; mRAP = mean right atrial pressure; mPCWP = mean pulmonary capillary wedge pressure; QoL = quality of life; ICPET = incremental cardiopulmonary exercise test; VO2 = oxygen consumption; WR = work rate; VE = minute ventilation; VCO2 = carbon dioxide output.

.

Variable		Iron - Placebo			Placebo - Iron		
	N	**Baseline**	**12 weeks**	**24 weeks**	**Baseline**	**12 weeks**	**24 weeks**
Iron indices							
Iron, umol/L	16	8.5 (6.5)	14.0 (7.5)	12.5 (10.0)	10.0 (6.0)	12.0 (2.75)	16.5 (4.0)
Ferritin, ug/L	16	11.0 (10.0)	98.0 (86.8)	53.0 (47.5)	11.0 (1.0)	13.0 (3.3)	169.5 (45.5)
TSAT, %	16	11.0 (8.25)	24.5 (20.8)	21.5 (13.5)	12.0 (8.0)	14.5 (4.25)	26.0 (7.8)
Haemoglobin, g/dL	16	14.4 (2.4)	15.4 (2.3)	16.0 (1.8)	12.3 (1.1)	12.9 (1.7)	13.4 (1.1)
sTfR, nmol/L	-	-	-	-	-	-	-
Hepcidin, ng/mL	-	-	-	-	-	-	-
Primary endpoint							
PVR, Wood units	16	9.1 (3.5)	11.1 (4.9)	-	9.5 (1.0)	9.8 (7.6)	-
Secondary endpoints							
Functional status							
6MWD, m	16	427.5 (102.0)	463.5 (68.75)	465.0 (91.8)	480.0 (46.0)	515.5 (75.5)	505.0 (71.5)
Dyspnea score (Borg index)	16	2.5 (2.5)	1.0 (2.6)	1.5 (2.3)	0.5 (0.5)	0.8 (0.9)	0.8 (0.9)
NYHA functional class, *no.*	16						
I		1	1	1	0	1	1
II		9	9	9	3	3	3
III		2	2	2	2	0	0
NTproBNP, ng/L	-	-	-	-	-	-	-
Hemodynamics							
Systolic blood pressure, mmHg	16	104.0 (12.5)	110.0 (14.0)	100.0 (18.0)	108.0 (8.0)	110.0 (22.5)	100.0 (1.8)
Diastolic blood pressure, mmHg	16	70.0 (10.3)	70.0 (8.5)	69.5 (12.5)	70.0 (6.0)	65.0 (5.0)	70.0 (2.3)
mPAP, mmHg	16	48.0 (13.0)	51.0 (17.3)	-	49.0 (16.0)	43.0 (16.0)	-
mRAP, mmHg	16	3.5 (5.3)	6.5 (5.5)	-	1.0 (7.0)	5.0 (2.0)	-
mPCWP, mmHg	16	6.5 (3.5)	7.5 (2.3)	-	11.0 (5.0)	9.0 (0.8)	-
Cardiac index, L/min/m^2	16	2.7 (0.4)	2.5 (0.6)	-	2.4 (0.7)	2.7 (0.7)	-
Questionnaire (CAMPHOR)							
QoL score	-	-	-	-	-	-	-
Symptom score	-	-	-	-	-	-	-
Activity score	-	-	-	-	-	-	-

Variable		Iron - Placebo			Placebo - Iron		
	N	Baseline	12 weeks	24 weeks	Baseline	12 weeks	24 weeks
ECPET							
Endurance time, secs	-	-	-	-	-	-	-
VO_2 at 80% peak WR, L/min	-	-	-	-	-	-	-
VO_2 at 80% peak WR, mL/min/kg	-	-	-	-	-	-	-
VO_2 at 3mins at 80% peak WR, L/min	-	-	-	-	-	-	-
VO_2 at 3mins at 80% peak WR, mL/min/kg	-	-	-	-	-	-	-
ICPET							
Peak VO_2, L/min	15	0.83 (0.2)	0.83 (0.15)	0.93 (0.19)	1.05 (0.19)	0.98 (0.18)	1.04 (0.05)
Peak VO_2, mL/min/kg	15	13.9 (2.5)	13.7 (2.5)	14.6 (3.5)	18.2 (5.2)	14.1 (7.9)	18.4 (3.2)
VO_2 at metabolic threshold, L/min	15	0.61 (0.16)	0.61 (0.28)	0.70 (0.28)	0.69 (0.25)	0.65 (0.18)	0.74 (0.05)
VO_2 at metabolic threshold, mL/min/kg	15	10.1 (2.4)	10.4 (3.5)	11.6 (2.9)	11.3 (1.9)	10.5 (1.2)	12.5 (0.7)
VE/VCO_2 slope	14	38.3 (12.2)	34.3 (7.2)	37.5 (15.8)	27.9 (4.9)	27.9 (3.0)	27.5 (6.5)
VO_2/WR slope	11	5.2 (4.2)	5.8 (3.1)	6.0 (1.8)	6.4 (2.0)	6.0 (1.3)	6.9 (1.0)
Peak O_2 pulse	15	6.1 (1.6)	6.3 (1.4)	6.7 (1.5)	6.9 (1.8)	6.5 (0.6)	6.8 (1.1)

Supplementary Table 8.3.2 Primary and secondary endpoints of the Ferinject (China) clinical trial.
All continuous variables are median (interquartile range). Participants were randomized to receive either iron (Iron-Placebo) or saline (Placebo-Iron) at baseline in this crossover design. The Iron-Placebo group received intravenous iron at baseline and saline at 12 weeks. The Placebo-Iron group received saline at baseline and intravenous iron at 12 weeks. Column 'N' gives the total number of participants completing the assessment at every visit (baseline, 12 week, 24 week). Definition of abbreviations: TSAT = transferrin saturation; sTfR = soluble transferrin receptor; ECPET = endurance cardiopulmonary exercise test; 6MWD = six-minute walk distance; NTproBNP = N-terminal - pro hormone BNP; PVR = pulmonary vascular resistance; mPAP = mean pulmonary arterial pressure; mRAP = mean right atrial pressure; mPCWP = mean pulmonary capillary wedge pressure; QoL = quality of life; ICPET = incremental cardiopulmonary exercise test; VO2 = oxygen consumption; WR = work rate; VE = minute ventilation; VCO2 = carbon dioxide output.

Dependent variable (outcome)	Independent variables				
	sTfR/ferritin ratio change	Baseline sTfR/ferritin ratio	Baseline outcome	Age	Sex (base=female)
Endurance time change, secs	-66.33 (-134.22-1.57) 0.055	10.89 (-46.69-68.47) 0.699	-0.43 (-0.7--0.17) 0.003	3.47 (-0.92-7.86) 0.116	-49.19 (-168.13-69.76)
ICPET					
Peak VO2 change, L/min	-0.07 (-0.15-0) 0.063	0.004 (-0.07-0.08) 0.904	-0.15 (-0.38-0.08) 0.194	-0.005 (-0.01-0.0004) 0.071	-0.11 (-0.29-0.07) 0.208
Peak VO2 change, mL/min/kg	-0.43 (-1.52-0.67) 0.434	-0.14 (-1-0.72) 0.738	-0.23 (-0.45-0) 0.045	-0.06 (-0.12-0) 0.061	-1.08 (-3.19-1.03) 0.304
VO2 at metabolic threshold change, L/min	-0.02 (-0.08-0.04) 0.595	-0.04 (-0.09-0.01) 0.138	-0.41 (-0.65--0.16) 0.002	-0.002 (-0.01-0.002) 0.273	0.05 (-0.07-0.17) 0.37
VO2 at metabolic threshold change, mL/min/kg	0.45 (-0.28-1.19) 0.218	-0.59 (-1.19-0) 0.051	-0.61 (-0.86--0.35) <0.001	-0.03 (-0.07-0.02) 0.21	0.87 (-0.49-2.23) 0.201
VE/VCO2 slope change	1.4 (-1.07-3.87) 0.257	1.56 (-0.52-3.64) 0.137	-0.25 (-0.41--0.08) 0.005	-0.002 (-0.16-0.15) 0.975	0.44 (-3.9-4.78) 0.839
VO2/WR slope change	-0.03 (-1.12-1.06) 0.952	0.13 (-0.71-0.96) 0.76	-0.71 (-1.04--0.38) <0.001	0.01 (-0.05-0.07) 0.711	0.01 (-1.87-1.89) 0.99
Peak O2 pulse	-0.58 (-1.01--0.14) 0.011	0.32 (-0.07-0.7) 0.1	-0.02 (-0.05-0) 0.083	-0.02 (-0.05-0) 0.083	-0.66 (-1.58-0.25) 0.148
ECPET					
VO2 at 80% peak work rate change, L/min	-0.06 (-0.12-0.01) 0.089	-0.02 (-0.07-0.04) 0.551	-0.17 (-0.39-0.04) 0.107	0.001 (-0.004-0.01) 0.732	-0.03 (-0.18-0.12) 0.673
VO2 at 80% peak work rate change, mL/min/kg	0.08 (-0.87-1.04) 0.862	-0.43 (-1.18-0.31) 0.239	-0.34 (-0.55--0.14) 0.002	0.002 (-0.05-0.06) 0.938	0.34 (-1.41-2.09) 0.692
VO2 at 3mins at 80% peak work rate change, L/min	0.6 (-0.37-1.57) 0.211	-0.49 (-1.42-0.44) 0.28	-1.65 (-4.2-0.9) 0.191	-0.02 (-0.08-0.04) 0.493	1.65 (-0.04-3.33) 0.055
VO2 at 3mins at 80% peak work rate change, mL/min/kg	0.51 (-0.89-1.91) 0.456	-0.52 (-1.88-0.83) 0.426	-0.01 (-0.32-0.3) 0.931	-0.01 (-0.09-0.07) 0.775	-0.05 (-2.59-2.5) 0.97
6MWD change, m	5.17 (-11.86-22.21) 0.54	11.78 (-2.99-26.55) 0.114	-0.25 (-0.4--0.1) 0.002	0.54 (-0.52-1.6) 0.307	-6.75 (-38.69-25.19) 0.669
Dyspnea score (Borg index) change	0.12 (-0.61-0.86) 0.735	0.11 (-0.49-0.71) 0.708	-0.23 (-0.53-0.07) 0.135	0.03 (-0.01-0.08) 0.102	-0.43 (-1.72-0.86) 0.503
NTproBNP change, ng/L	-14.87 (-121.6-91.82) 0.778	-36.9 (-129.7-55.9) 0.424	-0.31 (-0.47--0.15) <0.001	1.16 (-5.15-7.48) 0.709	37.08 (-147.04-221.19)
QoL score change	0.1 (-1.42-1.62) 0.893	0.24 (-1.01-1.49) 0.695	-0.02 (-0.19-0.14) 0.775	0.06 (-0.03-0.15) 0.194	1.37 (-1.27-4.02) 0.294

Supplementary Table 8.3.3 Results of the treatment-only analyses (whole cohort).
Results of all coefficients (independent variables) in sixteen regression models for predicting change in outcome from iron infusion (baseline) to week 12. Per-coefficient statistics are given as follows: beta (lower 95% CI – upper 95% CI) P-value. Definition of abbreviations: sTfR = soluble transferrin receptor; 6MWD = six-minute walk distance; NTproBNP = N-terminal -pro hormone BNP; QoL = quality of life; VO2 = oxygen consumption; WR = work rate; VE = minute ventilation; VCO2 = carbon dioxide output.

Dependent variable (outcome)	Independent variables				
	sTfR/ferritin ratio change	Baseline sTfR/ferritin ratio	Baseline outcome	Age	Sex (base=female)
Endurance time change, secs	-36.2 (-86.95-14.55) 0.146	-3.48 (-68.71-61.75) 0.909	-0.48 (-0.89--0.06) 0.029	1.34 (-2.94-5.62) 0.509	-54.29 (-148.51-39.93)
ICPET					
Peak VO2 change, L/min	-0.11 (-0.19--0.02) 0.015	-0.07 (-0.17-0.03) 0.133	0.11 (-0.18-0.4) 0.42	-0.002 (-0.01-0.003) 0.325	-0.29 (-0.5--0.08) 0.008
Peak VO2 change, mL/min/kg	-0.36 (-1.57-0.85) 0.54	-1.53 (-2.91--0.16) 0.031	-0.27 (-0.55-0.02) 0.063	-0.02 (-0.09-0.05) 0.525	-1.19 (-3.54-1.17) 0.305
VO2 at metabolic threshold change, L/min	0 (-0.07-0.07) 0.978	-0.09 (-0.18-0.01) 0.069	-0.36 (-0.64--0.08) 0.016	-0.002 (-0.01-0.008) 0.42	0.03 (-0.12-0.17) 0.707
VO2 at metabolic threshold change, mL/min/kg	0.45 (-0.35-1.25) 0.252	-1.43 (-2.46--0.4) 0.009	-0.63 (-0.91--0.34) <0.001	-0.003 (-0.06-0.05) 0.909	0.63 (-0.93-2.2) 0.406
VE/VCO2 slope change	1.18 (-2.36-4.71) 0.494	1.19 (-3.43-5.81) 0.595	-0.18 (-0.51-0.15) 0.26	0.01 (-0.23-0.25) 0.961	0.79 (-5.83-7.4) 0.805
VO2/WR slope change	-0.34 (-1.76-1.09) 0.617	0.27 (-1.42-1.96) 0.735	-0.39 (-0.87-0.09) 0.1	-0.01 (-0.09-0.07) 0.809	-1.75 (-4.42-0.91) 0.177
Peak O2 pulse	-0.88 (-1.4--0.36) 0.002	-0.14 (-0.81-0.53) 0.66	-0.02 (-0.27-0.24) 0.894	-0.01 (-0.04-0.03) 0.73	-1.55 (-2.85--0.24) 0.023
ECPET					
VO2 at 80% peak work rate change, L/min	-0.04 (-0.13-0.06) 0.394	-0.09 (-0.21-0.02) 0.11	-0.3 (-0.62-0.03) 0.071	0.002 (-0.004-0.01) 0.462	0.02 (-0.18-0.22) 0.825
VO2 at 80% peak work rate change, mL/min/kg	0.81 (-0.12-1.74) 0.081	-1.57 (-2.61--0.53) 0.007	-0.66 (-0.9--0.41) <0.001	0.03 (-0.02-0.08) 0.256	1.37 (-0.21-2.95) 0.083
VO2 at 3mins at 80% peak work rate change, L/min	1.12 (-0.53-2.77) 0.16	-0.92 (-3.33-1.49) 0.413	-1.39 (-5.58-2.81) 0.478	-0.02 (-0.13-0.09) 0.67	2.36 (-0.37-5.1) 0.083
VO2 at 3mins at 80% peak work rate change, mL/min/kg	14.75 (-2.49-32) 0.086	-10.56 (-35.62-14.5) 0.37	-2.85 (-7.35-1.65) 0.189	-0.39 (-1.57-0.79) 0.476	21.62 (-9.14-52.38) 0.148
6MWD change, m	9.35 (-9.55-28.24) 0.312	12.87 (-13.75-39.49) 0.323	-0.19 (-0.4-0.02) 0.068	0.04 (-1.24-1.31) 0.954	-3.33 (-39.47-32.8) 0.848
Dyspnea score (Borg index) change	0.28 (-0.75-1.32) 0.573	0.46 (-0.75-1.66) 0.437	-0.12 (-0.62-0.39) 0.625	0.01 (-0.04-0.07) 0.638	-0.34 (-2.02-1.34) 0.678
NTproBNP change, ng/L	-15.1 (-182.5-152.24) 0.851	-28.9 (-244.9-187.01) 0.781	-0.32 (-0.54--0.1) 0.007	0.99 (-9.31-11.3) 0.842	104.9 (-192.6-402.33) 0.468
QoL score change	-1.1 (-3.8-1.61) 0.397	0.27 (-3.1-3.63) 0.867	-0.16 (-0.47-0.14) 0.272	0.1 (-0.06-0.25) 0.195	-0.82 (-5.14-3.49) 0.687

Supplementary Table 8.3.4 Results of the treatment-only analyses (severely iron deficient).
Results of all coefficients (independent variables) in sixteen regression models for predicting change in outcome from iron infusion (baseline) to week 12. Per-coefficient statistics are given as follows: beta (lower 95% CI – upper 95% CI) P-value. Definition of abbreviations: sTfR = soluble transferrin receptor; 6MWD = six-minute walk distance; NTproBNP = N-terminal -pro hormone BNP; QoL = quality of life; VO2 = oxygen consumption; WR = work rate; VE = minute ventilation; VCO2 = carbon dioxide output.

8.4 Expression quantitative trait locus mapping in pulmonary arterial hypertension

Supplementary Table 8.4.1 Lead variants for novel expression quantitative trait loci in the PAH Cohort ordered by variance explained in gene expression.
Column names: rs ID = dbSNP (150) variant identifier; chr:pos = chromosome and genomic position (Ensembl GRCh37 coordinates); ref = reference allele in the human genome reference GRCh37; alt = alternative allele in the human genome reference GRCh37; Ensembl gene ID = Ensembl stable gene identifiers; variant type = dbSNP version 150 variant annotation; R^2 (%) = percent variance explained in gene expression by the genetic variant.

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs2295768	6:2954602	C	T	HTATSF1P2	ENSG00000271361	cis	7.1x10-83	57.4
rs2240101	15:31115709	G	C	RP5.1086D14.5	ENSG00000259845	cis	4.5x10-79	56.2
rs2302559	7:105903904	T	C	NAMPTL	ENSG00000229644	trans	6.35x10-78	55.9
rs745666	17:72744798	G	C	RP11.452I5.2	ENSG00000266036	cis	1.9x10-70	53.3
rs7635	11:721570	C	T	RPLP2	ENSG00000177600	cis	2.05x10-70	53.3
rs471521	12:10115647	A	T	CLEC12B	ENSG00000256660	cis	1.94x10-68	52.5
rs67933574	12:11226237	A	G	TAS2R43	ENSG00000255374	cis	8.58x10-65	51.1
rs11079015	17:39975164	C	T	NT5C3B	ENSG00000141698	cis	2.28x10-64	51
rs2280370	16:89627460	T	G	RPL13P12	ENSG00000215030	trans	1.72x10-62	50.2
rs9892882	17:18501596	A	G	USP32P1	ENSG00000188933	trans	2.03x10-62	50.2
rs9892882	17:18501596	A	G	CCDC144A	ENSG00000170160	trans	3.71x10-62	50.1
rs72785089	16:70005967	A	G	CLEC18A	ENSG00000157322	cis	1.14x10-61	49.9
rs10947675	6:37479972	G	A	RP1.153P14.8	ENSG00000204110	cis	4.66x10-60	49.2
rs1455979	21:44524006	G	A	U2AF1	ENSG00000160201	cis	2.5x10-59	48.9
rs12975591	19:10627814	A	G	CTC.539A10.4	ENSG00000267100	cis	1.6x10-58	48.5
rs10790962	11:128385169	G	A	Sep-03	ENSG00000100167	trans	4.37x10-57	47.9
rs12225372	11:74604594	A	G	XRRA1	ENSG00000166435	cis	4.79x10-56	47.4
rs2732603	17:44367288	G	C	RP11.197O15.1	ENSG00000214401	cis	5.07x10-55	46.9
rs79240608	1:149585945	T	C	NBPF26	ENSG00000273136	trans	6.4x10-55	46.9
rs2336136	14:74030518	T	A	ACOT1	ENSG00000184227	cis	1.12x10-54	46.8
rs2286466	16:2014283	A	G	RPS2P5	ENSG00000240342	trans	4.46x10-54	46.5
rs2656417	17:18475630	T	C	USP32P2	ENSG00000233327	cis	7.58x10-54	46.4
rs55969088	18:9024224	G	T	NDUFV2P1	ENSG00000267809	trans	7.55x10-54	46.4
rs9264607	6:31237605	G	A	HLA.L	ENSG00000243753	trans	2.13x10-53	46.2
rs61744075	8:23288346	C	A	NKX3.1	ENSG00000167034	cis	1.92x10-52	45.7
rs17847967	12:8082367	T	C	SLC2A14	ENSG00000173262	cis	2.4x10-52	45.6
rs11677238	2:114424974	T	G	AC002055.4	ENSG00000184319	trans	4.55x10-52	45.5
rs73287153	7:12139272	A	G	VWDE	ENSG00000146530	cis	7.37x10-52	45.4
rs216257	6:24934664	A	T	LSAMP	ENSG00000185565	trans	1x10-50	44.8
rs10489762	1:183570545	G	T	CTD.2058B24.2	ENSG00000259048	trans	1.84x10-50	44.7
rs2732603	17:44367288	G	C	RP11.583F2.1	ENSG00000264057	trans	2.02x10-50	44.7
rs111839387	7:65285169	C	T	PSPH	ENSG00000146733	trans	5.95x10-50	44.4
rs12447763	16:23753594	G	T	RP11.151E14.1	ENSG00000261035	trans	8.42x10-50	44.4
rs9934565	16:87764267	A	T	RP11.278A23.2	ENSG00000260671	cis	1.73x10-49	44.2
rs2732617	17:44363632	A	C	RP11.359G22.2	ENSG00000230400	trans	4.78x10-49	44
rs9273495	6:32628273	G	A	HLA.DQB2	ENSG00000232629	cis	1.07x10-48	43.8

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs4976210	5:70307464	T	C	SERF1A	ENSG00000172058	cis	1.65x10-48	43.7
rs2732603	17:44367288	G	C	DND1P1	ENSG00000264070	cis	2.68x10-48	43.6
rs145140188	11:77627431	G	C	INTS4P1	ENSG00000164669	trans	4.36x10-48	43.5
rs2294641	1:11721854	C	T	FBXO44	ENSG00000132879	cis	1.12x10-46	42.7
rs111164082	17:44047037	T	A	RP11.798G7.3	ENSG00000214425	cis	1.8x10-46	42.6
rs6085421	20:6142259	T	A	RP11.199O14.1	ENSG00000229876	cis	2.95x10-46	42.5
rs471521	12:10115647	A	T	AC091814.2	ENSG00000225231	cis	8.12x10-46	42.2
rs10409076	19:45057224	G	T	CEACAM22P	ENSG00000230666	cis	1.68x10-45	42.1
rs12912107	15:48520590	G	A	C10orf111	ENSG00000176236	trans	2.46x10-45	42
rs150996856	7:76596954	G	A	CCDC146	ENSG00000135205	cis	2.44x10-45	42
rs2272550	14:22591988	C	A	TRAV26.1	ENSG00000211807	cis	2.02x10-45	42
rs739239	22:50050468	G	A	RP1.29C18.10	ENSG00000212939	cis	2.29x10-45	42
rs4774967	15:57855760	G	T	GCOM1	ENSG00000137878	cis	1.88x10-44	41.5
rs7207087	17:46917953	C	G	CALCOCO2	ENSG00000136436	cis	3.11x10-44	41.4
rs7952689	11:7492630	T	G	CTD.2516F10.2	ENSG00000251364	cis	7.99x10-44	41.1
rs62557076	9:69449629	T	C	ANKRD20A4	ENSG00000172014	cis	2.74x10-42	40.2
rs67851010	6:32501347	G	A	HLA.DRB5	ENSG00000198502	cis	4.55x10-42	40.1
rs3179969	14:88862529	G	A	SPATA7	ENSG00000042317	cis	1.43x10-41	39.8
rs236628	7:75023850	G	T	AC006014.6	ENSG00000205583	cis	2.03x10-41	39.7
rs330924	8:8999187	G	C	RP11.10A14.3	ENSG00000254340	cis	4.11x10-41	39.5
rs9408297	9:48559	G	T	WASHC1	ENSG00000181404	cis	2.19x10-40	39.1
rs12047480	1:198635256	A	G	WISP2	ENSG00000064205	trans	5.28x10-40	38.8
rs1141684	16:2014591	C	T	RP11.364L4.1	ENSG00000228981	trans	3.44x10-39	38.3
rs9264669	6:31239681	A	T	HLA.C	ENSG00000204525	cis	4.11x10-39	38.3
rs10931780	2:198175297	G	C	AC013264.2	ENSG00000231621	cis	9.22x10-39	38.1
rs76909629	2:68258988	T	C	C1D	ENSG00000197223	cis	1.81x10-38	37.9
rs10130430	14:37448518	T	C	SLC25A21	ENSG00000183032	cis	2.1x10-38	37.8
rs62072802	17:18490410	A	G	RP11.219A15.1	ENSG00000266302	trans	2.19x10-38	37.8
rs61809765	1:149660625	A	G	NBPF8P	ENSG00000270231	trans	1.54x10-37	37.3
rs17816376	15:90233553	T	C	PEX11A	ENSG00000166821	cis	1.88x10-37	37.2
rs11765305	7:143111112	C	G	AC092214.10	ENSG00000229153	cis	2.68x10-37	37.1
rs2714313	6:7331682	G	C	DSP	ENSG00000096696	cis	2.58x10-37	37.1
rs11772895	7:143081942	G	C	TAS2R41	ENSG00000221855	cis	3.57x10-37	37
rs5761004	22:23978677	G	A	DRICH1	ENSG00000189269	cis	5.78x10-37	36.9
rs1451659	11:4439535	T	C	OR51R1P	ENSG00000237272	cis	9.34x10-37	36.8
rs7093	17:3566232	C	T	TAX1BP3	ENSG00000213977	cis	7.11x10-37	36.8
rs62192251	2:231277825	G	T	RP11.634B22.4	ENSG00000254467	trans	1.36x10-36	36.6
rs7225024	17:57780223	C	A	TMEM148	ENSG00000179219	trans	2.14x10-36	36.5
rs7575806	2:201632512	G	C	ORC2	ENSG00000115942	cis	2.52x10-36	36.5
rs16995310	20:1550334	C	A	RP4.576H24.4	ENSG00000260861	cis	2.69x10-36	36.4
rs12415	2:232573292	T	C	PTMA	ENSG00000187514	cis	5.05x10-36	36.3
rs13099099	3:132441268	T	C	NPHP3	ENSG00000113971	cis	5.1x10-36	36.3
rs2732603	17:44367288	G	C	MAPK8IP1	ENSG00000121653	trans	4.15x10-36	36.3
rs376215580	1:146385910	A	G	NBPF12	ENSG00000268043	trans	1.2x10-35	36
rs1510881	4:144439082	T	C	FREM3	ENSG00000183090	cis	2.13x10-35	35.8

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs28408686	17:30846624	C	G	CDK5R1	ENSG00000176749	cis	3x10-35	35.7
rs12962555	18:46901634	T	A	RPL17.C18orf32	ENSG00000215472	cis	4.32x10-35	35.6
rs72843816	10:131846688	T	C	RP11.500G10.1	ENSG00000237489	cis	7.05x10-35	35.5
rs9273841	6:32629582	C	T	RP11.544M22.1	ENSG00000198358	trans	1.66x10-34	35.2
rs375407631	9:66470329	A	G	RP11.262H14.3	ENSG00000234665	cis	3.48x10-34	35
rs67651814	20:43598154	G	C	RP11.369C8.1	ENSG00000258616	trans	3.62x10-34	35
rs11709127	3:20228009	T	C	AC099057.8	ENSG00000231304	cis	4.87x10-34	34.9
rs62557076	9:69449629	T	C	RP11.146D12.2	ENSG00000240240	trans	5.42x10-34	34.9
rs56402156	7:143103481	G	A	OR2R1P	ENSG00000236853	cis	7.86x10-34	34.8
rs7142957	14:21724070	G	A	AP002387.1	ENSG00000172900	trans	6.94x10-34	34.8
rs1637340	7:74337634	G	A	GTF2IRD2	ENSG00000196275	cis	1.45x10-33	34.6
rs7146961	14:106993798	C	T	IGHV3.48	ENSG00000211964	cis	1.89x10-33	34.5
rs7952890	12:9478318	C	G	RP13.735L24.1	ENSG00000260423	cis	1.89x10-33	34.5
rs2505811	10:42872833	G	C	RP11.313J2.1	ENSG00000215146	cis	4.63x10-33	34.2
rs982085	5:137775581	C	G	KDM3B	ENSG00000120733	cis	8.47x10-33	34
rs522042	1:74666274	C	T	RP1.20N18.6	ENSG00000224308	trans	2.52x10-32	33.7
rs72642336	4:68466995	C	A	TMPRSS11D	ENSG00000153802	cis	2.79x10-32	33.6
rs2975065	7:38401522	A	G	TRGV4	ENSG00000211698	cis	3.91x10-32	33.5
rs56350361	11:64156776	A	T	AP003774.6	ENSG00000231680	cis	9.39x10-32	33.3
rs6725743	2:68622990	T	C	PLEK	ENSG00000115956	cis	9.19x10-32	33.3
rs2945641	11:93807605	A	C	HEPHL1	ENSG00000181333	cis	9.58x10-32	33.2
rs1803262	2:87068827	G	A	AC108868.4	ENSG00000254126	trans	1.92x10-31	33
rs3803798	17:7460517	G	C	TNFSF12.TNFSF13	ENSG00000248871	cis	1.94x10-31	33
rs10902267	11:1007831	T	C	AP2A2	ENSG00000183020	cis	3.43x10-31	32.8
rs7960229	12:93717719	T	C	RP11.511B23.2	ENSG00000257322	cis	3.42x10-31	32.8
rs684552	1:120917635	G	A	NBPF8P	ENSG00000270231	cis	5.66x10-31	32.7
rs1777459	1:108584999	G	C	RP11.356N1.2	ENSG00000226822	cis	6.36x10-31	32.6
rs2523590	6:31327064	T	C	HCG4P3	ENSG00000237669	trans	7.37x10-31	32.6
rs669340	17:18263957	G	C	SHMT1	ENSG00000176974	cis	6.08x10-31	32.6
rs2732617	17:44363632	A	C	KANSL1	ENSG00000120071	cis	1.04x10-30	32.5
rs17817950	17:53045325	G	A	STXBP4	ENSG00000166263	cis	2.71x10-30	32.2
rs9611170	22:39784845	G	C	SYNGR1	ENSG00000100321	cis	3.08x10-30	32.1
rs10448	19:59093239	T	C	RPL23AP79	ENSG00000213753	cis	5.39x10-30	31.9
rs12184651	13:21329029	G	A	EEF1AKMT1	ENSG00000150456	cis	6.06x10-30	31.9
rs17552066	12:27863531	T	C	RP11.1060J15.4	ENSG00000256377	cis	5.84x10-30	31.9
rs6806332	3:125763566	A	G	ROPN1B	ENSG00000114547	cis	5.27x10-30	31.9
rs3757093	6:2766779	G	A	MYLK4	ENSG00000145949	cis	7.14x10-30	31.8
rs12798684	11:5215042	C	T	AC104389.16	ENSG00000224091	cis	1.42x10-29	31.6
rs3918010	1:36944462	A	G	CBLN2	ENSG00000141668	trans	2.79x10-29	31.4
rs1808209	17:46140291	C	T	HOXB2	ENSG00000173917	cis	5.32x10-29	31.2
rs143043168	19:22857004	A	G	ZNF492	ENSG00000229676	cis	7.19x10-29	31.1
rs75045569	7:143109208	T	G	OR10AC1P	ENSG00000176510	cis	5.69x10-29	31.1
rs17180100	6:29936065	C	T	HLA.G	ENSG00000204632	cis	1.59x10-28	30.8
rs10081693	9:94966486	T	G	OGN	ENSG00000106809	cis	2.31x10-28	30.7
rs12124948	1:150527294	C	T	HORMAD1	ENSG00000143452	cis	1.76x10-28	30.7

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs10772397	12:11138683	C	T	TAS2R50	ENSG00000212126	cis	2.98x10-28	30.6
rs763271	22:27043233	C	G	CTA.211A9.5	ENSG00000244625	cis	3.05x10-28	30.6
rs4785093	16:60083019	A	G	RP11.430C1.1	ENSG00000261807	cis	3.98x10-28	30.5
rs3792988	6:24876260	T	C	RP5.1112D6.5	ENSG00000229276	trans	5.41x10-28	30.4
rs73046589	12:10944122	T	C	RP11.711K1.7	ENSG00000255790	cis	4.79x10-28	30.4
rs112534607	17:44344376	G	A	AC126544.4	ENSG00000204650	cis	6.6x10-28	30.3
rs62282782	3:195687736	G	C	AC069513.3	ENSG00000242086	cis	6.97x10-28	30.3
rs2844827	6:29811237	T	C	HCG4P5	ENSG00000227766	cis	9.03x10-28	30.2
rs67933574	12:11226237	A	G	TAS2R64P	ENSG00000256274	cis	7.7x10-28	30.2
rs10794309	11:325948	C	T	RP11.326C3.15	ENSG00000270972	cis	1.06x10-27	30.1
rs4658519	1:243110266	C	T	CEP170P1	ENSG00000154608	trans	1.69x10-27	30
rs1133222	10:52516665	C	T	ASAH2	ENSG00000188611	cis	9.34x10-27	29.4
rs9894225	17:4546559	A	G	ALOX15	ENSG00000161905	cis	9.14x10-27	29.4
rs1043314	11:118915944	A	G	HYOU1	ENSG00000149428	cis	1.75x10-26	29.1
rs11914110	22:23660328	G	A	AP000550.1	ENSG00000169668	trans	1.74x10-26	29.1
rs2316770	17:43934256	G	A	LRRC37A3	ENSG00000176809	trans	2.1x10-26	29.1
rs9657894	7:150437927	C	T	RP5.1051J4.6	ENSG00000177590	cis	2.82x10-26	29
rs146432789	5:178453100	A	G	ZNF879	ENSG00000234284	cis	4.33x10-26	28.8
rs6465906	7:102787381	A	C	DPY19L2P2	ENSG00000170629	cis	4.21x10-26	28.8
rs1637340	7:74337634	G	A	AC004166.5	ENSG00000123965	cis	5.47x10-26	28.7
rs57890152	5:149821472	T	C	RP11.343H5.4	ENSG00000224114	trans	6.4x10-26	28.7
rs72642336	4:68466995	C	A	GNRHR	ENSG00000109163	cis	6.24x10-26	28.7
rs13081249	3:152092101	T	C	RP11.513O17.2	ENSG00000253301	trans	9.68x10-26	28.5
rs373427274	5:177218466	C	T	FAM153C	ENSG00000204677	cis	1.23x10-25	28.4
rs10907940	9:42191070	C	A	RP11.204M4.2	ENSG00000204837	cis	1.5x10-25	28.3
rs10907940	9:42191070	C	A	RP11.111F5.5	ENSG00000227449	trans	1.86x10-25	28.3
rs143820282	1:1676091	C	T	RP1.283E3.8	ENSG00000268575	cis	1.57x10-25	28.3
rs2008345	14:23401442	T	A	RP11.298I3.1	ENSG00000257285	cis	1.86x10-25	28.3
rs62027001	15:102317025	T	C	CTD.2611K5.6	ENSG00000259660	cis	1.47x10-25	28.3
rs905707	17:72563644	A	G	CTD.2006K23.1	ENSG00000261222	cis	1.57x10-25	28.3
rs2685190	3:130564579	G	A	NEK11	ENSG00000114670	cis	3.05x10-25	28.1
rs72929707	2:196439583	C	T	AC064834.2	ENSG00000223466	cis	4.24x10-25	28
rs78821460	5:127874149	G	C	FBN2	ENSG00000138829	cis	3.77x10-25	28
rs13255886	8:71565954	G	T	XKR9	ENSG00000221947	cis	4.93x10-25	27.9
rs2836200	21:39593581	G	A	AP001434.2	ENSG00000226012	cis	5.35x10-25	27.9
rs6933976	6:28969522	G	A	ZNF311	ENSG00000197935	cis	5.62x10-25	27.9
rs10794309	11:325948	C	T	RP11.326C3.12	ENSG00000255328	cis	6.15x10-25	27.8
rs988067	8:133791175	C	G	TG	ENSG00000042832	cis	5.79x10-25	27.8
rs79379134	16:30292500	A	T	RP11.345J4.5	ENSG00000261740	cis	9.43x10-25	27.7
rs11040487	11:6086071	C	T	CTD.2008A1.2	ENSG00000259479	trans	2.27x10-24	27.3
rs114586219	15:82537939	C	A	FAM154B	ENSG00000188659	cis	2.14x10-24	27.3
rs2982122	1:148002289	C	T	NBPF26	ENSG00000273136	trans	2.72x10-24	27.3
rs324122	19:52893957	C	T	ZNF880	ENSG00000221923	cis	2.43x10-24	27.3
rs3366	21:44834825	G	A	SIK1	ENSG00000142178	cis	2.2x10-24	27.3
rs1985275	11:9000993	A	G	TMEM9B.AS1	ENSG00000254860	cis	3.41x10-24	27.2

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs75364523	17:4723121	G	A	CHRNE	ENSG00000108556	cis	3.22x10-24	27.2
rs1056739	10:104236542	T	C	MFSD13A	ENSG00000138111	cis	3.87x10-24	27.1
rs4267298	16:70004028	C	T	CLEC18B	ENSG00000140839	trans	4.39x10-24	27.1
rs7791024	7:66766024	T	C	PMS2L4	ENSG00000067601	cis	3.75x10-24	27.1
rs11099650	4:85483950	A	C	AC005022.1	ENSG00000224017	trans	5.76x10-24	27
rs12269882	11:79092486	A	G	ODZ4	ENSG00000149256	cis	4.87x10-24	27
rs7952790	12:68406452	C	T	GS1.410F4.3	ENSG00000236946	cis	5.03x10-24	27
rs1892568	21:39563599	T	C	AP001434.1	ENSG00000231123	cis	6.26x10-24	26.9
rs4050880	22:22624189	C	A	IGLV5.52	ENSG00000211643	cis	6.56x10-24	26.9
rs56006248	17:43815141	A	G	LRRC37A	ENSG00000176681	cis	7.47x10-24	26.9
rs639455	1:20978221	C	T	RP11.401M16.7	ENSG00000117242	cis	7.37x10-24	26.9
rs7872739	9:95217930	C	T	ECM2	ENSG00000106823	cis	6.58x10-24	26.9
rs3894776	19:1457370	C	T	CTB.25B13.12	ENSG00000267317	cis	1.1x10-23	26.7
rs10221200	17:15589734	G	A	TRIM16	ENSG00000221926	cis	1.67x10-23	26.6
rs1501371	15:82460440	A	C	UBE2Q2P2	ENSG00000259429	cis	1.49x10-23	26.6
rs375407631	9:66470329	A	G	RP11.262H14.7	ENSG00000227582	cis	1.5x10-23	26.6
rs4006668	2:219595001	C	T	RP11.459I19.1	ENSG00000272555	cis	1.87x10-23	26.5
rs77727254	17:30986984	C	T	TMEM98	ENSG00000006042	cis	2.08x10-23	26.5
rs9926	4:10076860	G	A	WDR1	ENSG00000071127	cis	2.02x10-23	26.5
rs10470179	21:39570684	C	T	AC079753.4	ENSG00000180152	trans	2.27x10-23	26.4
rs13340529	7:66466119	A	C	TYW1	ENSG00000198874	cis	2.31x10-23	26.4
rs2641419	11:4555193	A	C	OR52I2	ENSG00000226288	cis	2.25x10-23	26.4
rs7898547	10:6659210	A	C	RP11.554I8.1	ENSG00000225948	cis	2.58x10-23	26.4
rs3917945	1:36942233	T	G	CA5A	ENSG00000174990	trans	3.21x10-23	26.3
rs9264682	6:31240872	T	G	HLA.K	ENSG00000230795	trans	3.48x10-23	26.3
rs140950438	5:59983635	G	C	DEPDC1B	ENSG00000035499	cis	4.59x10-23	26.2
rs35117338	7:66072781	A	T	GS1.124K5.12	ENSG00000229180	cis	4.15x10-23	26.2
rs3930740	17:8294959	C	T	KRBA2	ENSG00000184619	cis	4.42x10-23	26.2
rs4685744	3:4403537	C	T	AC023480.1	ENSG00000229241	cis	4.24x10-23	26.2
rs10179864	2:238675892	A	G	RP11.255C15.1	ENSG00000240429	trans	5.65x10-23	26.1
rs112161250	15:58440089	G	GCACA	RP1	ENSG00000104237	trans	5.3x10-23	26.1
rs2596496	6:31322782	C	G	HLA.B	ENSG00000234745	cis	6.12x10-23	26.1
rs113590651	8:82707315	A	G	SNX16	ENSG00000104497	cis	6.42x10-23	26
rs1749596	1:236380913	T	C	ERO1B	ENSG00000086619	cis	7.55x10-23	26
rs2596503	6:31320810	A	G	DDX39B	ENSG00000198563	cis	6.22x10-23	26
rs12366	2:75185856	G	T	AC104135.2	ENSG00000236209	cis	1.02x10-22	25.9
rs3026243	12:9183400	T	C	RP11.118B22.1	ENSG00000256817	cis	8.09x10-23	25.9
rs2285170	1:172425529	C	G	SUCO	ENSG00000094975	cis	1.26x10-22	25.8
rs7400769	14:94613418	A	T	PPP4R4	ENSG00000119698	cis	1.09x10-22	25.8
rs77972028	1:149674716	G	A	CH17.118O6.2	ENSG00000226067	cis	1.15x10-22	25.8
rs7295439	12:89825122	G	T	RP11.734K2.4	ENSG00000270344	cis	1.4x10-22	25.7
rs2435201	17:44078982	T	A	RP11.995C19.3	ENSG00000232300	cis	1.89x10-22	25.6
rs6502892	17:1617608	T	C	TLCD2	ENSG00000185561	cis	2.17x10-22	25.6
rs7096229	10:91394900	T	G	RP11.248C1.2	ENSG00000232229	cis	1.75x10-22	25.6
rs72697606	8:145611531	A	T	CPSF1	ENSG00000071894	cis	1.75x10-22	25.6

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs11657128	17:56048840	G	A	RP11.415I12.3	ENSG00000249753	trans	2.59x10-22	25.5
rs4013507	10:48303741	G	A	RP11.144G6.12	ENSG00000254929	trans	3.35x10-22	25.4
rs76549974	9:43793814	C	T	RP11.146D12.2	ENSG00000240240	trans	3.21x10-22	25.4
rs17442763	12:40538796	T	A	RP11.476D10.1	ENSG00000260943	cis	4.33x10-22	25.3
rs3117640	19:55338833	G	A	KIR3DL1	ENSG00000167633	cis	4.03x10-22	25.3
rs368785114	9:45729667	C	A	RP11.146D12.2	ENSG00000240240	trans	3.94x10-22	25.3
rs7838905	8:2954809	C	G	CSMD1	ENSG00000183117	cis	4.47x10-22	25.3
rs2523815	6:29840243	A	G	HLA.K	ENSG00000230795	cis	5.44x10-22	25.2
rs3087563	11:117707937	T	C	FXYD6	ENSG00000137726	cis	5.38x10-22	25.2
rs6488334	12:11140444	T	C	TAS2R46	ENSG00000226761	cis	5.15x10-22	25.2
rs8073963	17:46609918	G	A	HOXB.AS1	ENSG00000230148	cis	5.01x10-22	25.2
rs9368476	6:26906789	A	G	RP11.457M11.5	ENSG00000261584	cis	5.45x10-22	25.2
rs995059	9:114439510	C	T	RP11.18M11.2	ENSG00000224812	trans	5.88x10-22	25.2
rs1852270	17:43655280	G	T	LRRC37A	ENSG00000176681	cis	7.65x10-22	25.1
rs2523815	6:29840243	A	G	HCG4B	ENSG00000227262	cis	6.32x10-22	25.1
rs62050500	17:21446432	G	A	C17orf51	ENSG00000212719	cis	6.93x10-22	25.1
rs6720375	2:132440073	G	C	AP000525.9	ENSG00000206195	trans	7.08x10-22	25.1
rs7480474	11:307912	T	C	IFITM3	ENSG00000142089	cis	6.29x10-22	25.1
rs12489198	3:193294307	G	T	RP11.175P19.2	ENSG00000236297	cis	9.45x10-22	25
rs8068200	17:45390562	A	G	RP11.66B24.7	ENSG00000272808	trans	9.41x10-22	25
rs8077275	17:18705412	G	A	TVP23B	ENSG00000171928	cis	8.76x10-22	25
rs12763558	10:70872970	G	A	VPS26A	ENSG00000122958	cis	1.36x10-21	24.8
rs879985390	7:101977597	G	A	LRWD1	ENSG00000161036	cis	1.36x10-21	24.8
rs13401870	2:68340446	A	G	CNRIP1	ENSG00000119865	cis	1.82x10-21	24.7
rs1678185	2:231409854	C	T	UNC13C	ENSG00000137766	trans	2.25x10-21	24.6
rs1834274	2:131951961	G	A	NF1L2	ENSG00000236956	cis	2.38x10-21	24.6
rs2388419	15:94400712	G	A	RP11.76E17.4	ENSG00000258831	cis	2.38x10-21	24.6
rs10995796	10:51592744	T	G	PARGP1	ENSG00000239883	trans	3.23x10-21	24.5
rs58304714	1:150831561	T	A	RP11.732A21.2	ENSG00000260008	trans	3.31x10-21	24.5
rs72836318	17:44121579	T	C	RP11.798G7.8	ENSG00000266918	cis	2.91x10-21	24.5
rs7897787	10:18921314	C	A	RP11.499P20.2	ENSG00000240291	cis	2.97x10-21	24.5
rs11648223	16:29585333	C	A	RP11.231C14.3	ENSG00000254634	cis	3.56x10-21	24.4
rs12475433	2:173293419	G	T	AC078883.4	ENSG00000226963	cis	5.69x10-21	24.2
rs2490849	9:15388507	C	T	CCDC171	ENSG00000164989	cis	5.6x10-21	24.2
rs2753747	10:81512832	A	G	NUTM2B	ENSG00000188199	cis	5.76x10-21	24.2
rs2908776	16:53407506	C	G	RP11.44F14.2	ENSG00000261804	cis	5.58x10-21	24.2
rs72744331	9:76788027	T	C	RORB	ENSG00000198963	cis	6.04x10-21	24.2
rs7947900	11:316299	T	G	IFITM2	ENSG00000185201	cis	6.28x10-21	24.2
rs430748	22:18896529	C	T	DGCR6	ENSG00000183628	cis	8.04x10-21	24.1
rs9761856	4:119579892	G	A	SYNPO2	ENSG00000172403	cis	7.48x10-21	24.1
rs12653907	5:175485237	C	A	RP11.844P9.5	ENSG00000248596	cis	1.07x10-20	24
rs190206	17:15464679	G	A	TVP23C.CDRT4	ENSG00000259024	cis	1.03x10-20	24
rs6925979	6:142651052	G	A	ADGRG6	ENSG00000112414	cis	9.91x10-21	24
rs876554	12:89885473	T	C	POC1B	ENSG00000139323	cis	9.87x10-21	24
rs2377220	1:1660635	T	G	CDK11A	ENSG00000008128	cis	1.3x10-20	23.9

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs3118000	1:16909820	C	G	MST1L	ENSG00000186715	cis	1.28x10-20	23.9
rs7273242	20:62594160	T	C	SAMD10	ENSG00000130590	cis	1.41x10-20	23.9
rs72836318	17:44121579	T	C	ARL17B	ENSG00000228696	cis	1.18x10-20	23.9
rs11619622	13:42570201	A	G	VWA8	ENSG00000102763	cis	1.81x10-20	23.8
rs3095174	17:16254494	C	G	CENPV	ENSG00000166582	cis	1.62x10-20	23.8
rs3098394	19:12397460	G	A	CTD.2666L21.2	ENSG00000234750	cis	1.45x10-20	23.8
rs4236478	7:142411527	T	C	U66061.35	ENSG00000244273	cis	1.56x10-20	23.8
rs4484184	3:53981048	T	C	ACTR8	ENSG00000113812	cis	1.62x10-20	23.8
rs12975429	19:33199227	G	A	NUDT19	ENSG00000213965	cis	1.97x10-20	23.7
rs3102342	16:89451575	T	C	RP1.168P16.1	ENSG00000261692	cis	1.96x10-20	23.7
rs6960171	7:64910100	G	A	RP11.792A8.4	ENSG00000272831	cis	1.95x10-20	23.7
rs1042229	19:52249672	A	C	FPR1	ENSG00000171051	cis	2.34x10-20	23.6
rs7829886	8:95497388	T	A	KIAA1429	ENSG00000164944	cis	2.53x10-20	23.6
rs2373340	7:148001246	A	G	CNTNAP2	ENSG00000174469	trans	3.68x10-20	23.5
rs1375862	9:125617037	G	A	ZBTB26	ENSG00000171448	cis	4.57x10-20	23.4
rs72689277	4:185593699	G	A	PRIMPOL	ENSG00000164306	cis	4.62x10-20	23.4
rs1051664	1:45116985	C	T	RNF220	ENSG00000187147	cis	5.82x10-20	23.3
rs2003490	14:75377555	C	A	DLST	ENSG00000119689	cis	5.97x10-20	23.3
rs7350611	12:11331094	T	C	TAS2R13	ENSG00000212128	cis	5.99x10-20	23.3
rs112413396	4:70278401	T	C	UGT2B28	ENSG00000135226	cis	7.37x10-20	23.2
rs28652789	16:83807	G	C	DDX11L10	ENSG00000233614	cis	6.36x10-20	23.2
rs56104184	19:49370310	C	T	PPP1R15A	ENSG00000087074	cis	6.84x10-20	23.2
rs12704823	7:76822127	C	T	AC073635.5	ENSG00000250990	cis	8.66x10-20	23.1
rs867663893	3:122443492	A	T	DIRC2	ENSG00000138463	cis	8.2x10-20	23.1
rs13333072	16:8823848	A	G	ABAT	ENSG00000183044	cis	1.12x10-19	23
rs142929928	17:4783340	C	G	C17orf107	ENSG00000205710	cis	1.08x10-19	23
rs2286466	16:2014283	A	G	RPS2P46	ENSG00000189343	trans	1.21x10-19	23
rs6085271	20:5748056	T	C	C20orf196	ENSG00000171984	cis	1.29x10-19	22.9
rs987534	8:120713652	G	A	DSCC1	ENSG00000136982	cis	1.48x10-19	22.9
rs9796412	14:77211737	C	T	RP11.488C13.5	ENSG00000258301	cis	1.64x10-19	22.8
rs11076704	16:88777995	C	T	CTU2	ENSG00000174177	cis	2.33x10-19	22.7
rs1355775	6:52529358	G	A	RP1.152L7.5	ENSG00000216775	cis	2.05x10-19	22.7
rs76312359	15:55829689	A	T	PYGO1	ENSG00000171016	cis	2.14x10-19	22.7
rs9514837	13:108995148	G	C	MYO16	ENSG00000041515	cis	1.99x10-19	22.7
rs12167058	22:29907401	C	T	RP11.638L3.3	ENSG00000265555	trans	2.65x10-19	22.6
rs2769345	6:167366394	T	C	RP11.514O12.4	ENSG00000249141	cis	2.7x10-19	22.6
rs35940311	1:93663604	G	A	DR1	ENSG00000117505	cis	2.7x10-19	22.6
rs9272779	6:32610314	T	C	HLA.DQA1	ENSG00000196735	cis	2.64x10-19	22.6
rs162298	6:167455629	G	A	NPTXR	ENSG00000221890	trans	3.28x10-19	22.5
rs58581511	17:57950621	T	G	RNFT1	ENSG00000189050	cis	3.39x10-19	22.5
rs2588347	12:10944783	A	G	TAS2R31	ENSG00000256436	cis	4.88x10-19	22.4
rs7718927	5:109907035	A	T	TMEM232	ENSG00000186952	cis	3.99x10-19	22.4
rs12216750	8:82066711	T	A	RP11.1149M10.2	ENSG00000253214	cis	7.12x10-19	22.2
rs12464694	2:196908927	A	G	DNAH7	ENSG00000118997	cis	6.41x10-19	22.2
rs2924251	17:46338677	A	G	HOXB3	ENSG00000120093	cis	6.45x10-19	22.2

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs4674336	2:219632820	A	G	PRKAG3	ENSG00000115592	cis	6.31x10-19	22.2
rs586746	19:53080928	A	G	ZNF701	ENSG00000167562	cis	9.73x10-19	22.1
rs1873379	15:75498744	G	C	RP11.817O13.8	ENSG00000260274	cis	1.22x10-18	22
rs2302058	19:57703077	T	G	ZNF264	ENSG00000083844	cis	1.15x10-18	22
rs2832328	21:30821767	C	G	RP11.313P22.1	ENSG00000273464	cis	9.97x10-19	22
rs6589587	11:116875338	T	G	AP000936.1	ENSG00000234268	cis	1.16x10-18	22
rs10400906	15:83297138	C	T	CPEB1	ENSG00000214575	cis	1.43x10-18	21.9
rs2780681	10:27512767	A	G	RP11.85G18.4	ENSG00000230445	cis	1.29x10-18	21.9
rs61148413	15:84828326	T	A	GOLGA6L4	ENSG00000184206	cis	1.48x10-18	21.9
rs7223043	17:33826654	A	G	RP11.1094M14.8	ENSG00000267369	cis	1.37x10-18	21.9
rs8058597	16:70082214	G	C	NPIPB15	ENSG00000196436	trans	1.44x10-18	21.9
rs145475856	1:146367867	G	T	PFN1P8	ENSG00000244371	cis	1.66x10-18	21.8
rs62371990	5:42604782	G	A	CCDC152	ENSG00000198865	cis	1.76x10-18	21.8
rs2588347	12:10944783	A	G	TAS2R12	ENSG00000256682	cis	2.05x10-18	21.7
rs964611	15:48597514	C	A	RP5.898J17.1	ENSG00000231987	trans	2.43x10-18	21.7
rs7223732	17:45881197	G	C	SCRN2	ENSG00000141295	cis	3.08x10-18	21.6
rs9893390	17:33751315	T	A	RP11.686D22.3	ENSG00000267648	cis	3.09x10-18	21.6
rs10079553	5:178959047	C	T	RUFY1	ENSG00000176783	cis	3.54x10-18	21.5
rs12728538	1:63959125	A	C	ITGB3BP	ENSG00000142856	cis	3.25x10-18	21.5
rs12939821	17:62079286	G	A	PRR29	ENSG00000224383	cis	3.75x10-18	21.5
rs200276266	5:175530214	C	T	FAM153A	ENSG00000170074	trans	3.13x10-18	21.5
rs4744150	9:95630363	C	T	RP11.526D8.3	ENSG00000226721	cis	3.44x10-18	21.5
rs78042407	10:49333147	G	C	AGAP9	ENSG00000204172	trans	3.35x10-18	21.5
rs10125553	9:37841380	A	G	DCAF10	ENSG00000122741	cis	4.48x10-18	21.4
rs144921768	16:32973407	G	A	DUSP22	ENSG00000112679	trans	5.91x10-18	21.3
rs2450556	8:95330332	A	C	FSBP	ENSG00000265817	cis	7.31x10-18	21.2
rs4083651	9:19211852	A	G	RP11.513M16.8	ENSG00000273226	cis	7.64x10-18	21.2
rs11659948	18:9197372	A	G	RP11.63E9.1	ENSG00000259828	trans	9.25x10-18	21.1
rs12718958	7:55431162	A	G	LANCL2	ENSG00000132434	cis	8.91x10-18	21.1
rs2873872	17:49508445	A	T	RP11.1018N14.5	ENSG00000267452	cis	8.46x10-18	21.1
rs6558530	8:1706207	A	G	AC093375.1	ENSG00000226383	trans	8.42x10-18	21.1
rs116918202	16:89587080	C	T	CDK10	ENSG00000185324	cis	1.17x10-17	21
rs1817303	10:51450060	A	C	RP11.546J1.1	ENSG00000270557	trans	1.05x10-17	21
rs9412607	9:42901002	T	G	RP11.745L13.2	ENSG00000251555	trans	1.16x10-17	21
rs10405910	19:57961191	A	T	AC004076.9	ENSG00000268163	cis	1.43x10-17	20.9
rs179468	22:24109774	T	G	CHCHD10	ENSG00000250479	cis	1.49x10-17	20.9
rs2671827	17:33732330	C	T	RP11.686D22.8	ENSG00000267745	cis	1.35x10-17	20.9
rs73153794	7:134853085	A	C	RP11.90K6.1	ENSG00000239482	trans	1.47x10-17	20.9
rs9353446	6:87780724	T	C	C6orf163	ENSG00000203872	cis	1.36x10-17	20.9
rs111277130	5:1634070	C	T	CTD.2083E4.4	ENSG00000260774	trans	1.62x10-17	20.8
rs1128894	1:153604422	T	C	S100A1	ENSG00000160678	cis	1.66x10-17	20.8
rs1367224	2:175391901	T	C	AC018890.6	ENSG00000236449	cis	1.54x10-17	20.8
rs67425176	8:66543064	T	C	CTD.3025N20.3	ENSG00000272010	cis	1.76x10-17	20.8
rs6816018	4:88735959	G	C	SPP1	ENSG00000118785	cis	1.76x10-17	20.8
rs73213579	4:198953	C	A	SOHLH2	ENSG00000120669	trans	1.55x10-17	20.8

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs1007591	11:35731800	T	G	TRIM44	ENSG00000166326	cis	2.05x10-17	20.7
rs11009104	10:33116039	A	G	CCDC7	ENSG00000216937	cis	1.89x10-17	20.7
rs13081249	3:152092101	T	C	ANO7	ENSG00000146205	trans	2.22x10-17	20.7
rs529150802	16:89268323	G	T	FANCA	ENSG00000187741	cis	2.14x10-17	20.7
rs11150958	18:74248237	A	G	RP11.17M16.2	ENSG00000265778	cis	2.56x10-17	20.6
rs2075458	7:130528147	A	G	AC058791.1	ENSG00000226380	cis	2.68x10-17	20.6
rs2732617	17:44363632	A	C	MKRN9P	ENSG00000258128	trans	2.57x10-17	20.6
rs591411	9:79313529	C	T	PCA3	ENSG00000225937	cis	2.84x10-17	20.6
rs62469316	7:64674302	T	G	RP11.792A8.4	ENSG00000272831	trans	2.47x10-17	20.6
rs2416548	12:11324176	C	A	SMIM10L1	ENSG00000256537	cis	3.19x10-17	20.5
rs4472162	4:79626160	A	T	RP11.109G23.3	ENSG00000260278	cis	3.41x10-17	20.5
rs1541533	11:17096211	A	G	SNORD14A	ENSG00000272034	cis	4.41x10-17	20.4
rs1967415	17:57113124	C	A	TRIM37	ENSG00000108395	cis	4.04x10-17	20.4
rs1982074	19:10668673	A	G	KRI1	ENSG00000129347	cis	4.16x10-17	20.4
rs373769561	9:41922713	G	C	RP11.262H14.3	ENSG00000234665	trans	3.89x10-17	20.4
rs7105835	11:78075725	T	C	RP11.452H21.1	ENSG00000254420	cis	3.74x10-17	20.4
rs7837964	8:22264417	G	C	SLC39A14	ENSG00000104635	cis	3.8x10-17	20.4
rs11078300	17:2571489	G	A	PAFAH1B1	ENSG00000007168	cis	5.31x10-17	20.3
rs11249686	5:177181014	A	G	FAM153B	ENSG00000182230	trans	5.13x10-17	20.3
rs2730272	7:158798622	G	C	LINC00689	ENSG00000231419	cis	4.96x10-17	20.3
rs492095	17:72869078	A	G	FDXR	ENSG00000161513	cis	4.85x10-17	20.3
rs11186721	10:93641908	C	A	AC005307.1	ENSG00000260725	trans	6.82x10-17	20.2
rs376869011	9:66530411	A	G	FGF7	ENSG00000140285	trans	5.72x10-17	20.2
rs9893390	17:33751315	T	A	RP11.686D22.4	ENSG00000267547	cis	6.18x10-17	20.2
rs12514888	5:34221989	G	A	RP11.1023L17.2	ENSG00000215156	cis	7.93x10-17	20.1
rs3127726	10:46932364	T	C	RP11.144G6.12	ENSG00000254929	cis	7.76x10-17	20.1
rs35858034	19:41299864	A	C	CYP2T2P	ENSG00000233622	cis	7.39x10-17	20.1
rs11209243	1:68738861	G	C	RP11.518D3.3	ENSG00000234383	cis	9.9x10-17	20
rs11799352	1:118576627	A	T	SPAG17	ENSG00000155761	cis	1.06x10-16	20
rs150134394	10:81455053	T	C	RP11.119F19.4	ENSG00000224886	cis	1.09x10-16	20
rs4968290	17:45047384	C	T	RP11.156P1.3	ENSG00000262879	cis	1.02x10-16	20
rs61920361	12:27165921	A	T	RP11.582E3.4	ENSG00000256625	cis	9x10-17	20
rs10752555	9:70484945	A	G	RP11.262H14.3	ENSG00000234665	trans	1.23x10-16	19.9
rs4624506	3:39444901	T	C	SLC25A38	ENSG00000144659	cis	1.29x10-16	19.9
rs55962832	3:40244451	T	C	ENTPD3	ENSG00000168032	cis	1.34x10-16	19.9
rs10772397	12:11138683	C	T	TAS2R19	ENSG00000212124	cis	1.43x10-16	19.8
rs34511016	5:79881909	A	G	FAM151B	ENSG00000152380	cis	1.59x10-16	19.8
rs677078	11:121432287	C	A	RP11.86H7.1	ENSG00000224081	trans	1.58x10-16	19.8
rs72642346	4:68479934	C	T	AC079880.1	ENSG00000232398	cis	1.6x10-16	19.8
rs7839535	8:134586637	A	G	RP11.629O1.2	ENSG00000261220	cis	1.54x10-16	19.8
rs9353446	6:87780724	T	C	SMIM8	ENSG00000111850	cis	1.61x10-16	19.8
NA	17:80735913	C	CCTCA	ARL16	ENSG00000214087	trans	1.69x10-16	19.8
rs132650	22:36554588	C	T	CADPS	ENSG00000163618	trans	1.72x10-16	19.7
rs2388419	15:94400712	G	A	RP11.76E17.3	ENSG00000258476	cis	2.05x10-16	19.7
rs4585080	2:99557327	T	C	AC019097.7	ENSG00000231822	cis	1.74x10-16	19.7

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs71409258	15:80159117	A	G	ST20.MTHFS	ENSG00000259332	cis	1.85x10-16	19.7
rs10771996	12:9811352	A	G	AC003005.2	ENSG00000268266	trans	2.16x10-16	19.6
rs2008758	17:19893590	C	T	AKAP10	ENSG00000108599	cis	2.15x10-16	19.6
rs2531805	6:28412326	C	T	ZSCAN23	ENSG00000187987	cis	2.16x10-16	19.6
rs28549275	7:138886066	C	T	LUC7L2	ENSG00000146963	cis	2.25x10-16	19.6
rs3785884	17:44057595	G	A	LRRC37A2	ENSG00000238083	cis	2.33x10-16	19.6
rs7962354	12:62823325	T	C	RP11.769N19.2	ENSG00000257880	cis	2.17x10-16	19.6
rs2309426	9:71830308	T	G	TJP2	ENSG00000119139	cis	3.17x10-16	19.5
rs2945247	8:8096004	G	A	ABC11.48400900C8.1	ENSG00000254527	trans	2.78x10-16	19.5
rs374559823	2:87366345	C	T	RGPD1	ENSG00000187627	cis	3.17x10-16	19.5
rs61811182	1:183585327	A	G	SCGB3A2	ENSG00000164265	trans	2.9x10-16	19.5
rs74651789	4:165898682	C	T	TRIM61	ENSG00000183439	cis	2.78x10-16	19.5
rs116044941	17:73874684	G	C	RP11.552F3.9	ENSG00000267801	cis	3.44x10-16	19.4
rs3106150	17:62919344	G	A	DND1P1	ENSG00000264070	trans	3.86x10-16	19.4
rs743810	22:35742925	T	G	TOM1	ENSG00000100284	cis	3.68x10-16	19.4
rs9895898	17:18644142	T	C	TRIM16L	ENSG00000108448	cis	3.64x10-16	19.4
rs11052771	12:9891371	C	T	RP11.75L1.2	ENSG00000213443	cis	4.3x10-16	19.3
rs2169764	19:14193735	G	A	CTB.55O6.8	ENSG00000141854	cis	4.49x10-16	19.3
rs2540431	2:202367621	G	T	AC007256.5	ENSG00000213090	cis	4.58x10-16	19.3
rs2949585	1:148002266	A	C	NBPF8P	ENSG00000270231	trans	4.35x10-16	19.3
rs9289971	3:156389947	T	C	RP11.791I24.4	ENSG00000243926	cis	4.07x10-16	19.3
rs7978962	12:10524824	T	C	RP11.277P12.6	ENSG00000255641	cis	5.07x10-16	19.2
rs111527470	2:239149377	G	C	AC016757.3	ENSG00000186235	cis	6.53x10-16	19.1
rs11912319	22:50937900	A	G	CTA.384D8.34	ENSG00000273272	cis	6.52x10-16	19.1
rs148853830	7:99912929	A	T	PMS2L1	ENSG00000078319	cis	6.27x10-16	19.1
rs2569102	5:149800881	G	C	RPS14	ENSG00000164587	cis	6.29x10-16	19.1
rs2611491	10:51489393	T	G	CTGLF10P	ENSG00000230869	trans	7.32x10-16	19.1
rs28625723	7:135715783	T	C	AC024084.1	ENSG00000230649	cis	6.27x10-16	19.1
rs7216865	17:6891851	A	T	AC027763.2	ENSG00000215067	cis	7.26x10-16	19.1
rs9834	5:176733616	T	C	RP11.753G20.1	ENSG00000217325	trans	7.16x10-16	19.1
rs509421	1:24035696	G	A	CNR2	ENSG00000188822	cis	8.01x10-16	19
rs11048496	12:9172421	G	A	AP000688.11	ENSG00000236677	trans	1.15x10-15	18.9
rs11640445	16:60195615	T	A	RP11.430C1.1	ENSG00000261807	cis	1.07x10-15	18.9
rs218663	17:6567190	A	G	CTC.281F24.1	ENSG00000261996	cis	9.55x10-16	18.9
rs2530839	17:28964727	C	T	RP11.271K11.2	ENSG00000264943	cis	1.12x10-15	18.9
rs4657614	1:166899261	A	G	ILDR2	ENSG00000143195	cis	9.94x10-16	18.9
rs12441089	15:93277514	T	C	RP11.386M24.4	ENSG00000258741	cis	1.32x10-15	18.8
rs12482295	21:15263797	C	T	RP11.757O6.2	ENSG00000265787	trans	1.42x10-15	18.8
rs12917526	15:81025457	G	C	FAM108C1	ENSG00000136379	cis	1.45x10-15	18.8
rs115654156	6:31395040	G	C	MICA	ENSG00000204520	cis	1.55x10-15	18.7
rs4649586	1:144611851	G	A	CH17.118O6.2	ENSG00000226067	trans	1.47x10-15	18.7
rs60545295	15:48521101	G	A	CTB.43E15.3	ENSG00000253686	trans	1.59x10-15	18.7
rs10136164	14:37483309	T	C	PAX9	ENSG00000198807	cis	1.89x10-15	18.6
rs10752555	9:70484945	A	G	FGF7	ENSG00000140285	trans	1.9x10-15	18.6
rs11655481	17:30622552	C	T	CYP4F11	ENSG00000171903	trans	1.85x10-15	18.6

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs1808199	6:29924542	G	T	SPINK1	ENSG00000164266	trans	1.9x10-15	18.6
rs2005594	22:19021819	A	G	AC000095.8	ENSG00000241527	cis	2.12x10-15	18.6
rs489552	9:27196617	C	T	RP11.56F10.3	ENSG00000254396	cis	2.13x10-15	18.6
rs79710144	10:17723145	T	A	FAT3	ENSG00000165323	trans	1.94x10-15	18.6
rs2353408	3:142791030	A	C	U2SURP	ENSG00000163714	cis	2.48x10-15	18.5
rs34499924	11:60970328	A	G	PGA3	ENSG00000229859	cis	2.58x10-15	18.5
rs375809946	5:175499074	G	A	FAM153C	ENSG00000204677	trans	2.57x10-15	18.5
rs7531159	1:74653039	C	T	FPGT.TNNI3K	ENSG00000259030	cis	2.33x10-15	18.5
rs7835354	8:19564149	A	G	RP11.1105O14.1	ENSG00000253270	cis	2.35x10-15	18.5
rs3904094	13:37683479	T	C	VMAC	ENSG00000187650	trans	2.86x10-15	18.4
rs9949547	18:37381312	T	C	RP11.244M2.1	ENSG00000267374	cis	3.34x10-15	18.4
rs201958810	10:47097863	G	GTCTC	AGAP9	ENSG00000204172	cis	3.92x10-15	18.3
rs3865258	17:18709492	T	C	TVP23C	ENSG00000175106	trans	4x10-15	18.3
rs3915878	4:144915780	C	T	RP11.248J18.2	ENSG00000269906	trans	3.92x10-15	18.3
rs4579515	8:59476807	A	G	AC068522.4	ENSG00000242970	cis	3.73x10-15	18.3
rs72805116	16:85792967	A	G	C16orf74	ENSG00000154102	cis	3.57x10-15	18.3
rs112697442	14:35580186	C	T	SEPT7P1	ENSG00000259090	cis	4.51x10-15	18.2
rs1411476	1:104569468	C	T	RP11.364B6.2	ENSG00000215869	cis	5.64x10-15	18.1
rs142863425	6:27885980	C	T	ZNF389	ENSG00000226314	cis	6.07x10-15	18.1
rs17216887	4:100010273	C	G	RP11.696N14.1	ENSG00000246090	cis	5.73x10-15	18.1
rs2269984	7:141362693	A	T	RP5.894A10.6	ENSG00000270157	cis	5.46x10-15	18.1
rs2314645	4:15538922	G	T	ADRB3	ENSG00000188778	trans	5.99x10-15	18.1
rs2315601	17:39086932	C	T	KRT23	ENSG00000108244	cis	6.04x10-15	18.1
rs2739337	22:24296483	C	T	DDT	ENSG00000099977	cis	6.23x10-15	18.1
rs2831947	21:30022213	A	G	CCT8	ENSG00000156261	cis	5.57x10-15	18.1
rs28565512	4:57397196	C	G	THEGL	ENSG00000249693	cis	6.34x10-15	18.1
rs797515	13:51227415	G	A	DLEU7	ENSG00000186047	cis	5.67x10-15	18.1
rs113734006	8:59345027	C	T	UBXN2B	ENSG00000215114	cis	7.6x10-15	18
rs1991401	17:62502435	A	G	DDX5	ENSG00000108654	cis	6.94x10-15	18
rs4273875	8:2585105	T	C	GS1.57L11.1	ENSG00000253853	cis	6.5x10-15	18
rs77245739	16:4801939	G	T	ZNF500	ENSG00000103199	cis	7.53x10-15	18
rs78095701	17:66443609	A	C	WIPI1	ENSG00000070540	cis	6.42x10-15	18
rs112521249	14:35244362	C	T	RP11.73E17.2	ENSG00000258738	cis	8.3x10-15	17.9
rs11631833	15:75588262	G	A	CTD.2323K18.1	ENSG00000260269	cis	9.14x10-15	17.9
rs11668526	19:54749060	C	T	LILRA6	ENSG00000244482	cis	8.6x10-15	17.9
rs4141794	7:130667258	G	A	AMER2	ENSG00000165566	trans	7.86x10-15	17.9
rs4687658	3:52860366	C	G	RP5.966M1.6	ENSG00000243696	cis	8.27x10-15	17.9
rs4804785	19:7780643	A	C	CLEC4G	ENSG00000182566	cis	7.95x10-15	17.9
rs56053835	15:85093046	G	A	UBE2Q2P1	ENSG00000189136	cis	8.47x10-15	17.9
rs708277	16:54147935	G	A	FTO	ENSG00000140718	cis	9.17x10-15	17.9
rs78622635	1:16538094	A	G	ARHGEF19	ENSG00000142632	cis	8.34x10-15	17.9
rs9842198	3:156539204	G	C	PA2G4P4	ENSG00000230457	cis	8.35x10-15	17.9
rs10151105	14:62549004	G	A	RP11.355I22.3	ENSG00000186369	cis	1.14x10-14	17.8
rs11653833	17:15592143	C	G	ZNF286B	ENSG00000249459	trans	1.07x10-14	17.8
rs12927623	16:34021783	A	G	DUSP22	ENSG00000112679	trans	1.05x10-14	17.8

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs2712204	7:107148544	A	T	GPR22	ENSG00000172209	cis	1.14x10-14	17.8
rs2985424	1:145386391	A	G	RP11.458D21.5	ENSG00000213240	cis	1.04x10-14	17.8
rs34352889	10:88974706	G	A	RP11.506M13.3	ENSG00000244733	trans	1.11x10-14	17.8
rs11118912	1:222149130	C	T	RP11.815M8.1	ENSG00000238042	cis	1.42x10-14	17.7
rs11743541	5:75738858	G	A	CTD.2236F14.1	ENSG00000249713	cis	1.39x10-14	17.7
rs2454705	12:123192848	G	A	RP11.324E6.6	ENSG00000256249	cis	1.28x10-14	17.7
rs2961559	10:4864809	T	C	AKR1E2	ENSG00000165568	cis	1.39x10-14	17.7
rs35775224	7:7015632	A	C	RP11.266K4.12	ENSG00000256589	trans	1.27x10-14	17.7
rs56288025	19:58963717	G	T	ZNF324B	ENSG00000249471	cis	1.19x10-14	17.7
rs60163605	17:43334649	C	A	SPATA32	ENSG00000184361	cis	1.3x10-14	17.7
rs1227809	19:37269981	G	C	ZNF790	ENSG00000197863	cis	1.51x10-14	17.6
rs12814509	12:21613367	A	T	PYROXD1	ENSG00000121350	cis	1.47x10-14	17.6
rs200210467	14:59983185	T	G	CCDC175	ENSG00000151838	cis	1.79x10-14	17.6
rs2904758	6:29923351	G	A	HLA.W	ENSG00000235290	cis	1.46x10-14	17.6
rs62147427	2:87126664	C	T	RP1.81D8.3	ENSG00000224477	trans	1.46x10-14	17.6
rs1803005	6:24804860	G	T	C6orf229	ENSG00000260286	cis	1.82x10-14	17.5
rs2271089	17:79210707	G	A	C17orf89	ENSG00000224877	cis	2.47x10-14	17.4
rs2942202	8:23418444	A	C	PRDM6	ENSG00000061455	trans	2.31x10-14	17.4
rs76021834	7:75607155	A	G	MDH2	ENSG00000146701	cis	2.26x10-14	17.4
rs10081693	9:94966486	T	G	ASPN	ENSG00000106819	cis	2.89x10-14	17.3
rs10255707	7:45954692	T	C	AC073115.7	ENSG00000229628	cis	2.8x10-14	17.3
rs12168582	22:28316227	A	G	TTC28	ENSG00000100154	cis	2.71x10-14	17.3
rs1883595	6:40935728	G	A	APOBEC2	ENSG00000124701	cis	2.86x10-14	17.3
rs251732	16:4310468	C	G	TFAP4	ENSG00000090447	cis	2.81x10-14	17.3
rs3807891	7:95226709	G	C	AC002451.3	ENSG00000231170	cis	3.19x10-14	17.3
rs62065454	17:43573592	A	G	RP11.156P1.1	ENSG00000263142	trans	2.86x10-14	17.3
rs12453507	17:38053207	C	G	GSDMB	ENSG00000073605	cis	3.96x10-14	17.2
rs147751661	17:62927969	G	A	RP11.197O15.1	ENSG00000214401	trans	3.73x10-14	17.2
rs4754708	11:100759034	A	G	ARHGAP42	ENSG00000165895	cis	3.34x10-14	17.2
rs12539729	7:128142023	A	T	METTL2B	ENSG00000165055	cis	4.3x10-14	17.1
rs4647870	17:74516655	A	C	SNHG16	ENSG00000163597	cis	4.25x10-14	17.1
rs4736693	8:134553603	G	C	ERICH3	ENSG00000178965	trans	4.79x10-14	17.1
rs58173607	18:74277779	G	T	RP11.111H3.2	ENSG00000266256	cis	4.99x10-14	17.1
rs60659886	1:26994245	A	T	PIGV	ENSG00000060642	cis	4.33x10-14	17.1
rs7082545	10:52478100	T	G	ASAH2B	ENSG00000204147	cis	4.11x10-14	17.1
rs73359005	8:135795757	G	A	AC083843.1	ENSG00000259820	cis	4.85x10-14	17.1
rs11042607	11:2121817	A	G	IGF2	ENSG00000167244	cis	5.15x10-14	17
rs117491574	17:58325048	T	G	RP11.3K24.1	ENSG00000241157	cis	5.02x10-14	17
rs11951414	5:68431872	C	T	RP11.309M7.1	ENSG00000261298	trans	5.85x10-14	17
rs17259058	7:87921277	C	G	AC003991.3	ENSG00000228113	cis	5.72x10-14	17
rs2230707	1:19202896	G	A	ALDH4A1	ENSG00000159423	cis	5.13x10-14	17
rs34630775	8:79417693	G	T	RP11.578O24.2	ENSG00000254352	cis	5.17x10-14	17
rs71473137	15:30948494	C	T	RP11.932O9.9	ENSG00000269930	cis	5.45x10-14	17
rs72751993	1:233468388	T	C	RP5.862P8.2	ENSG00000143674	cis	5.21x10-14	17
rs867845230	17:62920612	A	G	RP11.583F2.1	ENSG00000264057	cis	6.09x10-14	17

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs9303577	17:49249186	C	A	NME1.NME2	ENSG00000011052	cis	6.08x10-14	17
rs1665294	2:114369935	T	G	AC016629.3	ENSG00000269600	trans	6.24x10-14	16.9
rs61978213	14:70653758	G	A	SLC8A3	ENSG00000100678	cis	6.89x10-14	16.9
rs12149595	16:35259917	G	A	DUSP22	ENSG00000112679	trans	9.01x10-14	16.8
rs3115858	1:755890	A	T	FAM87B	ENSG00000177757	cis	7.68x10-14	16.8
rs6503453	17:44062603	A	G	AC007679.3	ENSG00000225610	trans	9.16x10-14	16.8
rs6866645	5:137988944	A	G	LRRTM2	ENSG00000146006	cis	7.63x10-14	16.8
rs1852270	17:43655280	G	T	LRRC37A3	ENSG00000176809	trans	1.02x10-13	16.7
rs28709161	22:22646562	T	C	RP11.23E10.3	ENSG00000260626	trans	9.88x10-14	16.7
rs7446704	5:70662060	C	A	RP11.1280N14.4	ENSG00000215630	cis	9.27x10-14	16.7
rs10907576	9:46918872	A	G	RP11.262H14.3	ENSG00000234665	trans	1.14x10-13	16.6
rs12603589	17:65825248	T	C	AC005332.1	ENSG00000237854	cis	1.15x10-13	16.6
rs13122	14:31359024	A	T	RP11.829H16.3	ENSG00000258525	cis	1.17x10-13	16.6
rs3016125	22:20475050	T	C	TMEM191C	ENSG00000206140	trans	1.15x10-13	16.6
rs3106149	17:62919455	G	A	RP11.798G7.3	ENSG00000214425	trans	1.3x10-13	16.6
rs62307246	4:70295728	A	C	UGT2B11	ENSG00000213759	cis	1.19x10-13	16.6
rs62542381	9:27488566	C	T	RP11.120I21.2	ENSG00000254202	trans	1.13x10-13	16.6
rs75655792	7:99813285	C	A	PVRIG	ENSG00000213413	cis	1.14x10-13	16.6
rs184608345	9:66584918	C	T	RP11.745L13.2	ENSG00000251555	trans	1.42x10-13	16.5
rs185073	5:179334857	A	G	TBC1D9B	ENSG00000197226	cis	1.61x10-13	16.5
rs2562240	16:83458	A	G	AP006222.1	ENSG00000228463	trans	1.63x10-13	16.5
rs28551168	1:229456185	C	T	CCSAP	ENSG00000154429	cis	1.5x10-13	16.5
rs4356723	20:24899051	T	C	ACSS1	ENSG00000154930	cis	1.63x10-13	16.5
rs56248825	21:39465015	C	T	SCAMP4	ENSG00000227500	trans	1.54x10-13	16.5
rs59812768	1:198599981	T	C	RP11.55L3.1	ENSG00000250791	trans	1.59x10-13	16.5
rs77913677	3:122038460	T	C	KPNA1	ENSG00000114030	cis	1.52x10-13	16.5
rs113352051	2:89283617	T	C	IGKV1.9	ENSG00000241755	cis	1.69x10-13	16.4
rs148240181	12:93080	C	G	ABC7.42389800N19.1	ENSG00000226210	cis	1.9x10-13	16.4
rs1641716	12:12224718	G	A	BCL2L14	ENSG00000121380	cis	1.84x10-13	16.4
rs4798075	18:3247256	A	C	RP13.270P17.3	ENSG00000272688	cis	2x10-13	16.4
rs62253656	3:53023428	G	C	RP11.894J14.5	ENSG00000272305	cis	1.84x10-13	16.4
rs1035559	16:72031860	A	G	DHX38	ENSG00000140829	cis	2.43x10-13	16.3
rs1046381	1:153604112	G	A	RP1.178F15.5	ENSG00000271853	cis	2.34x10-13	16.3
rs13063033	3:46082481	G	A	RP11.793E15.1	ENSG00000229515	cis	2.47x10-13	16.3
rs2396690	6:46618289	C	G	TDRD6	ENSG00000180113	cis	2.26x10-13	16.3
rs2644720	11:10859439	C	T	SNORD97	ENSG00000238622	cis	2.12x10-13	16.3
rs372107064	5:21303103	G	C	RP11.457M11.5	ENSG00000261584	trans	2.15x10-13	16.3
rs55699816	3:196059377	A	T	TCTEX1D2	ENSG00000213123	cis	2.36x10-13	16.3
rs6718328	2:198038562	G	A	PPP2R2C	ENSG00000074211	trans	2.28x10-13	16.3
rs1784304	11:118962816	C	A	HMBS	ENSG00000256269	cis	2.69x10-13	16.2
rs4542712	17:7469327	T	C	AC113189.5	ENSG00000233223	cis	2.67x10-13	16.2
rs67654984	8:135707982	G	A	CTD.2342N23.3	ENSG00000272456	cis	2.54x10-13	16.2
rs7145650	14:35947295	G	A	RP11.287A8.2	ENSG00000229419	trans	2.69x10-13	16.2
rs7537908	1:78518939	A	G	GIPC2	ENSG00000137960	cis	2.81x10-13	16.2
rs80284522	3:139035188	T	C	RP11.319G6.1	ENSG00000248932	cis	2.74x10-13	16.2

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs869738	2:64493422	C	G	AC114752.2	ENSG00000236148	cis	2.8x10-13	16.2
rs10025792	4:164504947	T	G	RP11.557L19.1	ENSG00000272002	trans	3.15x10-13	16.1
rs11027831	11:3550720	C	T	CTC.342M10.2	ENSG00000255367	cis	3.15x10-13	16.1
rs11186897	10:94048558	C	G	Mar-05	ENSG00000198060	cis	3.08x10-13	16.1
rs1817104	12:11338589	G	T	TAS2R63P	ENSG00000256019	cis	3.3x10-13	16.1
rs200601667	2:107117022	A	AATTT	AC108868.4	ENSG00000254126	cis	3.5x10-13	16.1
rs4782394	16:88726726	T	G	RP5.1142A6.3	ENSG00000260630	cis	3.46x10-13	16.1
rs5744709	5:74891002	G	A	ANKDD1B	ENSG00000189045	cis	3.64x10-13	16.1
rs62461484	7:63911740	T	C	ERV3.1	ENSG00000213462	cis	3.14x10-13	16.1
rs6737901	2:169936393	A	G	DHRS9	ENSG00000073737	cis	3.71x10-13	16.1
rs76330394	7:102791112	A	C	RP11.401L13.4	ENSG00000224415	cis	3.21x10-13	16.1
rs796704680	15:32838699	T	C	GOLGA8R	ENSG00000186399	trans	3.19x10-13	16.1
rs13238996	7:74069645	A	G	GTF2IRD2B	ENSG00000174428	cis	4.44x10-13	16
rs17692129	17:44793283	C	T	RP11.927P21.12	ENSG00000271974	trans	3.87x10-13	16
rs530569	11:94169708	C	T	MIR548A1	ENSG00000207775	trans	3.88x10-13	16
rs142893200	15:84897901	A	G	GOLGA6L9	ENSG00000197978	trans	5.41x10-13	15.9
rs61842918	10:47746978	C	A	ANXA8	ENSG00000265190	cis	5.59x10-13	15.9
rs62118921	19:23798289	C	T	ZNF681	ENSG00000196172	cis	5.24x10-13	15.9
rs184608345	9:66584918	C	T	RP11.204M4.2	ENSG00000204837	trans	6.15x10-13	15.8
rs188871467	16:8928935	T	G	RP11.77H9.2	ENSG00000260276	cis	5.68x10-13	15.8
rs6770645	3:40227789	G	A	MYRIP	ENSG00000170011	cis	6.02x10-13	15.8
rs7399685	13:114516947	A	G	TMEM255B	ENSG00000184497	cis	6.17x10-13	15.8
rs11123747	2:99328763	T	C	GABRB2	ENSG00000145864	trans	7.04x10-13	15.7
rs142893200	15:84897901	A	G	RP11.152F13.3	ENSG00000255769	trans	8.07x10-13	15.7
rs146957881	5:177233117	C	A	RP11.1277A3.2	ENSG00000246596	cis	7.75x10-13	15.7
rs200690889	9:44544142	G	A	RP11.111F5.5	ENSG00000227449	trans	8.27x10-13	15.7
rs202086943	20:29602222	T	G	MLLT10P1	ENSG00000238151	cis	7.66x10-13	15.7
rs4476463	3:36880858	A	G	AC011816.1	ENSG00000234073	cis	7.43x10-13	15.7
rs57947330	17:16728955	T	C	CCDC144A	ENSG00000170160	cis	7.87x10-13	15.7
rs57947330	17:16728955	T	C	USP32P1	ENSG00000188933	cis	7.53x10-13	15.7
rs61998439	15:31111123	G	T	RP11.540B6.6	ENSG00000270015	cis	7.72x10-13	15.7
rs74651789	4:165898682	C	T	RP11.366M4.11	ENSG00000248632	cis	7.4x10-13	15.7
rs11052241	12:9811948	A	G	RP3.337H4.8	ENSG00000203362	trans	8.33x10-13	15.6
rs17216887	4:100010273	C	G	ADH6	ENSG00000172955	cis	8.98x10-13	15.6
rs3783164	13:41380363	C	T	RP11.346L13.3	ENSG00000168852	cis	8.68x10-13	15.6
rs4242389	8:22972321	G	A	POU5F1B	ENSG00000212993	trans	9.81x10-13	15.6
rs56065326	3:112008771	C	T	SLC9C1	ENSG00000172139	cis	8.44x10-13	15.6
rs56185013	4:106160133	G	A	TET2	ENSG00000168769	cis	9.51x10-13	15.6
rs61809245	1:148203960	G	A	NBPF20	ENSG00000162825	trans	8.88x10-13	15.6
rs62364993	5:80514325	G	C	CKMT2	ENSG00000131730	cis	8.58x10-13	15.6
rs6488334	12:11140444	T	C	TAS2R30	ENSG00000256188	cis	9.02x10-13	15.6
rs72905231	2:191303525	A	G	NEMP2	ENSG00000189362	cis	9.05x10-13	15.6
rs77391378	21:29947914	T	A	AF131217.1	ENSG00000232855	cis	9.87x10-13	15.6
rs776296	7:63804337	T	A	ZNF736	ENSG00000234444	cis	8.84x10-13	15.6
rs1233578	6:28712247	A	G	BTN3A2	ENSG00000186470	trans	1.14x10-12	15.5

rs ID	chr:pos	ref	alt	Gene symbol	Ensembl gene ID	cis/trans	P-value	R^2 (%)
rs2697291	2:198111312	A	G	RP11.794P6.1	ENSG00000255428	trans	1.12x10-12	15.5
rs61979250	14:76668905	C	G	GPATCH2L	ENSG00000089916	cis	1.09x10-12	15.5
rs34019730	16:89812930	G	A	FANCA	ENSG00000187741	cis	1.24x10-12	15.4
rs35084227	2:87277202	T	C	PLG	ENSG00000122194	trans	1.25x10-12	15.4
rs35785032	4:175697250	G	A	ADAM29	ENSG00000168594	cis	1.49x10-12	15.4
rs372161775	5:175398456	A	G	THOC3	ENSG00000051596	cis	1.29x10-12	15.4
rs41265179	1:153601828	G	T	NUP210L	ENSG00000143552	cis	1.4x10-12	15.4
rs10584	15:90328551	G	C	ANPEP	ENSG00000166825	cis	1.72x10-12	15.3
rs11246329	11:843713	C	T	TSPAN4	ENSG00000214063	cis	1.72x10-12	15.3
rs12812329	12:9167788	T	C	RP11.436I9.3	ENSG00000255776	cis	1.54x10-12	15.3
rs2684556	17:44580533	A	G	RP11.156P1.1	ENSG00000263142	cis	1.56x10-12	15.3
rs3807020	6:41747048	G	C	OR56A1	ENSG00000180934	trans	1.62x10-12	15.3
rs56175172	11:77245728	A	G	AP000580.1	ENSG00000219529	cis	1.51x10-12	15.3
rs61945858	12:120879461	A	G	DYNLL1	ENSG00000088986	cis	1.62x10-12	15.3
rs62414240	6:76274825	C	T	RNU6.1016P	ENSG00000252498	cis	1.72x10-12	15.3
rs7448600	5:21365623	C	A	RP11.823P9.3	ENSG00000233974	cis	1.82x10-12	15.3
rs10435105	7:3330994	C	T	SDK1	ENSG00000146555	cis	1.97x10-12	15.2
rs112161250	15:58440089	G	GCA	RP1	ENSG00000104237	trans	1.94x10-12	15.2
rs9567377	13:44720217	G	T	SMIM2.AS1	ENSG00000227258	cis	2x10-12	15.2

Supplementary Table 8.4.2 Significant KEGG pathways from the enrichment analysis of novel e-Genes in the PAH Cohort.
Column names: eQTL chr:pos (REF/ALT) = eQTL chromosome: base pair position (reference/alternative) alleles in the Ensembl human genome reference GRCh37; eQTL rs ID = dbSNP (150) variant identifier; eQTL type = variant annotation from ANNOVAR.

Gene symbol	Gene name	eQTL chr:pos	eQTL rs ID	cis/trans	eQTL type	eQTL P-value
Taste transduction						
TAS2R13	taste 2 receptor member 13	12:11331094 (T/C)	rs7350611	cis	intergenic	$5.99x10^{-20}$
TAS2R19	taste 2 receptor member 19	12:11138683 (C/T)	rs10772397	cis	exonic	$1.43x10^{-16}$
TAS2R30	taste 2 receptor member 30	12:11140444 (T/C)	rs6488334	cis	ncRNA intronic	$9.02x10^{-13}$
TAS2R31	taste 2 receptor member 31	12:10944783 (A/G)	rs2588347	cis	ncRNA intronic	$4.88x10^{-19}$
TAS2R41	taste 2 receptor member 41	7:143081942 (G/C)	rs11772895	cis	ncRNA intronic	$3.57x10^{-37}$
TAS2R43	taste 2 receptor member 43	12:11226237 (A/G)	rs67933574	cis	ncRNA intronic	$8.58x10^{-65}$
TAS2R46	taste 2 receptor member 46	12:11140444 (T/C)	rs6488334	cis	ncRNA intronic	$5.15x10^{-22}$
TAS2R50	taste 2 receptor member 50	12:11138683 (C/T)	rs10772397	cis	exonic	$2.98x10^{-28}$
Graft-versus-host disease						
HLA-B	major histocompatibility complex, class I, B	6:31322782 (C/G)	rs2596496	cis	intronic	$6.12x10^{-23}$
HLA-C	major histocompatibility complex, class I, C	6:31239681 (A/T)	rs9264669	cis	intronic	$4.11x10^{-39}$
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	6:32610314 (T/C)	rs9272779	cis	intronic	$2.64x10^{-19}$
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	6:32501347 (G/A)	rs67851010	cis	intergenic	$4.55x10^{-42}$
HLA-G	major histocompatibility complex, class I, G	6:31242089 (A/T)	rs2524080	trans	intergenic	$8.51x10^{-16}$
HLA-G	major histocompatibility complex, class I, G	6:29936065 (C/T)	rs17180100	cis	intergenic	$1.59x10^{-28}$
KIR3DL1	killer cell immunoglobulin like receptor, three Ig domains and cytoplasmic tail 1	19:55338833 (G/A)	rs3117640	cis	intronic	$4.03x10^{-22}$
Autoimmune thyroid disease						
HLA-B	major histocompatibility complex, class I, B	6:31322782 (C/G)	rs2596496	cis	intronic	$6.12x10^{-23}$
HLA-C	major histocompatibility complex, class I, C	6:31239681 (A/T)	rs9264669	cis	intronic	$4.11x10^{-39}$
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	6:32610314 (T/C)	rs9272779	cis	intronic	$2.64x10^{-19}$
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	6:32501347 (G/A)	rs67851010	cis	intergenic	$4.55x10^{-42}$
HLA-G	major histocompatibility complex, class I, G	6:31242089 (A/T)	rs2524080	trans	intergenic	$8.51x10^{-16}$
HLA-G	major histocompatibility complex, class I, G	6:29936065 (C/T)	rs17180100	cis	intergenic	$1.59x10^{-28}$
TG	thyroglobulin	8:133791175 (C/G)	rs988067	cis	intronic	$5.79x10^{-25}$
Allograft rejection						
HLA-B	major histocompatibility complex, class I, B	6:31322782 (C/G)	rs2596496	cis	intronic	$6.12x10^{-23}$
HLA-C	major histocompatibility complex, class I, C	6:31239681 (A/T)	rs9264669	cis	intronic	$4.11x10^{-39}$
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	6:32610314 (T/C)	rs9272779	cis	intronic	$2.64x10^{-19}$
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	6:32501347 (G/A)	rs67851010	cis	intergenic	$4.55x10^{-42}$
HLA-G	major histocompatibility complex, class I, G	6:31242089 (A/T)	rs2524080	trans	intergenic	$8.51x10^{-16}$
HLA-G	major histocompatibility complex, class I, G	6:29936065 (C/T)	rs17180100	cis	intergenic	$1.59x10^{-28}$
Type I diabetes						
HLA-B	major histocompatibility complex, class I, B	6:31322782 (C/G)	rs2596496	cis	intronic	$6.12x10^{-23}$
HLA-C	major histocompatibility complex, class I, C	6:31239681 (A/T)	rs9264669	cis	intronic	$4.11x10^{-39}$
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	6:32610314 (T/C)	rs9272779	cis	intronic	$2.64x10^{-19}$
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	6:32501347 (G/A)	rs67851010	cis	intergenic	$4.55x10^{-42}$
HLA-G	major histocompatibility complex, class I, G	6:31242089 (A/T)	rs2524080	trans	intergenic	$8.51x10^{-16}$
HLA-G	major histocompatibility complex, class I, G	6:29936065 (C/T)	rs17180100	cis	intergenic	$1.59x10^{-28}$

National Cohort Study of Idiopathic and Heritable PAH [Online]. Available: http://www.ipahcohort.com [Accessed].

Vanderbilt University Medical Center: Institute for Clinical and Translational Research. [Online]. Available: https://victr.vanderbilt.edu/pub/biovu/ [Accessed 16/01 2019].

AGUET, F., BARBEIRA, A. N., BONAZZOLA, R., BROWN, A., CASTEL, S. E., JO, B., KASELA, S., KIM-HELLMUTH, S., LIANG, Y., OLIVA, M., PARSANA, P. E., FLYNN, E., FRESARD, L., GAAMZON, E. R., HAMEL, A. R., HE, Y., HORMOZDIARI, F., MOHAMMADI, P., MUÑOZ-AGUIRRE, M., PARK, Y., SAHA, A., SEGRĆ, A. V., STROBER, B. J., WEN, X., WUCHER, V., DAS, S., GARRIDO-MARTÍN, D., GAY, N. R., HANDSAKER, R. E., HOFFMAN, P. J., KASHIN, S., KWONG, A., LI, X., MACARTHUR, D., ROUHANA, J. M., STEPHENS, M., TODRES, E., VIÑUELA, A., WANG, G., ZOU, Y., BROWN, C. D., COX, N., DERMITZAKIS, E., ENGELHARDT, B. E., GETZ, G., GUIGO, R., MONTGOMERY, S. B., STRANGER, B. E., IM, H. K., BATTLE, A., ARDLIE, K. G. & LAPPALAINEN, T. 2019. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*, 787903.

ALBOUAINI, K., EGRED, M., ALAHMAR, A. & WRIGHT, D. J. 2007. Cardiopulmonary exercise testing and its application. *Postgraduate medical journal,* 83**,** 675-682.

ALTO, M. & MAHADEVAN, V. S. 2012. Pulmonary arterial hypertension associated with congenital heart disease. *European Respiratory Review,* 21**,** 328.

ALTSHULER, D., DONNELLY, P. & THE INTERNATIONAL HAPMAP, C. 2005. A haplotype map of the human genome. *Nature,* 437**,** 1299-1320.

ALVES, J. L., GAVILANES, F., JARDIM, C., FERNANDES, C. J. C. D. S., MORINAGA, L. T. K., DIAS, B., HOETTE, S., HUMBERT, M. & SOUZA, R. 2015. Pulmonary Arterial Hypertension in the Southern Hemisphere. *Chest,* 147**,** 495-501.

ANDIAPPAN, A. K., MELCHIOTTI, R., POH, T. Y., NAH, M., PUAN, K. J., VIGANO, E., HAASE, D., YUSOF, N., SAN LUIS, B., LUM, J., KUMAR, D., FOO, S., ZHUANG, L., VASUDEV, A., IRWANTO, A., LEE, B., NARDIN, A., LIU, H., ZHANG, F., CONNOLLY, J., LIU, J., MORTELLARO, A., WANG, D. Y., POIDINGER, M., LARBI, A., ZOLEZZI, F. & ROTZSCHKE, O. 2015. Genome-wide analysis of the genetic regulation of gene expression in human neutrophils. *Nature Communications,* 6**,** 7971.

ANGRIST, J. D. & KRUEGER, A. B. 2001. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives,* 15.

ANKER, S. D., COMIN COLET, J., FILIPPATOS, G., WILLENHEIMER, R., DICKSTEIN, K., DREXLER, H., LÜSCHER, T. F., BART, B., BANASIAK, W., NIEGOWSKA, J., KIRWAN, B.-A., MORI, C., VON EISENHART ROTHE, B., POCOCK, S. J., POOLE-WILSON, P. A. & PONIKOWSKI, P. 2009. Ferric Carboxymaltose in Patients with Heart Failure and Iron Deficiency. *New England Journal of Medicine,* 361**,** 2436-2448.

ASTLE, W. J., ELDING, H., JIANG, T., ALLEN, D., RUKLISA, D., MANN, A. L., MEAD, D., BOUMAN, H., RIVEROS-MCKAY, F., KOSTADIMA, M. A., LAMBOURNE, J. J., SIVAPALARATNAM, S., DOWNES, K., KUNDU, K., BOMBA, L., BERENTSEN, K., BRADLEY, J. R., DAUGHERTY, L. C., DELANEAU, O., FRESON, K., GARNER, S. F., GRASSI, L., GUERRERO, J., HAIMEL, M., JANSSEN-MEGENS, E. M., KAAN, A., KAMAT, M., KIM, B., MANDOLI, A., MARCHINI, J., MARTENS, J. H., MEACHAM, S., MEGY, K., O'CONNELL, J., PETERSEN, R., SHARIFI, N., SHEARD, S. M., STALEY, J. R., TUNA, S., VAN DER ENT, M., WALTER, K., WANG, S. Y., WHEELER, E., WILDER, S. P., IOTCHKOVA, V., MOORE, C., SAMBROOK, J., STUNNENBERG, H. G., DI ANGELANTONIO, E., KAPTOGE, S., KUIJPERS, T. W., CARRILLO-DE-SANTA-PAU, E., JUAN, D., RICO, D., VALENCIA, A., CHEN, L., GE, B., VASQUEZ, L., KWAN, T., GARRIDO-MARTIN, D., WATT, S., YANG, Y., GUIGO, R., BECK, S., PAUL, D. S., PASTINEN, T., BUJOLD, D., BOURQUE, G., FRONTINI, M., DANESH, J., ROBERTS, D. J., OUWEHAND, W. H., BUTTERWORTH, A. S. & SORANZO, N. 2016. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell,* 167**,** 1415-1429 e19.

ATKINSON, C., STEWART, S., UPTON PAUL, D., MACHADO, R., THOMSON JENNIFER, R., TREMBATH RICHARD, C. & MORRELL NICHOLAS, W. 2002. Primary Pulmonary Hypertension Is Associated With Reduced Pulmonary Vascular Expression of Type II Bone Morphogenetic Protein Receptor. *Circulation,* 105**,** 1672-1678.

AUSTIN, E. D., MA, L., LEDUC, C., BERMAN ROSENZWEIG, E., BORCZUK, A., PHILLIPS JOHN, A., PALOMERO, T., SUMAZIN, P., KIM HYUNJAE, R., TALATI MEGHA, H., WEST, J., LOYD JAMES, E. & CHUNG WENDY, K. 2012. Whole Exome Sequencing to Identify a Novel Gene (Caveolin-1) Associated With Human Pulmonary Arterial Hypertension. *Circulation: Cardiovascular Genetics,* 5**,** 336-343.

AUTON, A., ABECASIS, G. R., ALTSHULER, D. M., DURBIN, R. M., ABECASIS, G. R., BENTLEY, D. R., CHAKRAVARTI, A., CLARK, A. G., DONNELLY, P., EICHLER, E. E., FLICEK, P., GABRIEL, S. B., GIBBS, R. A., GREEN, E. D., HURLES, M. E., KNOPPERS, B. M., KORBEL, J. O., LANDER, E. S., LEE, C., LEHRACH, H., MARDIS, E. R., MARTH, G. T., MCVEAN, G. A., NICKERSON, D. A., SCHMIDT, J. P., SHERRY, S. T., WANG, J., WILSON, R. K., GIBBS, R. A., BOERWINKLE, E., DODDAPANENI, H., HAN, Y., KORCHINA, V., KOVAR, C., LEE, S., MUZNY, D., REID, J. G., ZHU, Y., WANG, J., CHANG, Y., FENG, Q., FANG, X., GUO, X., JIAN, M., JIANG, H., JIN, X., LAN, T., LI, G., LI, J., LI, Y., LIU, S., LIU, X., LU, Y., MA, X., TANG, M., WANG, B., WANG, G., WU, H., WU, R., XU, X., YIN, Y., ZHANG, D., ZHANG, W., ZHAO, J., ZHAO, M., ZHENG, X., LANDER, E. S., ALTSHULER, D. M., GABRIEL, S. B., GUPTA, N., GHARANI, N., TOJI, L. H., GERRY, N. P., RESCH, A. M., FLICEK, P., BARKER, J., CLARKE, L., GIL, L., HUNT, S. E., KELMAN, G., KULESHA, E., LEINONEN, R., MCLAREN, W. M., RADHAKRISHNAN, R., ROA, A., SMIRNOV, D., SMITH, R. E., STREETER, I., THORMANN, A., TONEVA, I., VAUGHAN, B., ZHENG-BRADLEY, X., BENTLEY, D. R., GROCOCK, R., HUMPHRAY, S., JAMES, T., KINGSBURY, Z., LEHRACH, H., SUDBRAK, R., ALBRECHT, M. W., et al. 2015. A global reference for human genetic variation. *Nature,* 526**,** 68-74.

BALANOS, G., DORRINGTON, K. & ROBBINS, P. 2002. Desferrioxamine elevates pulmonary vascular resistance in humans: potential for involvement of HIF-1. *J Appl Physiol* 2501–2507.

BAND, G. & MARCHINI, J. qctool. 2.0 ed.

BARNETT, C. F., HSUE, P. Y. & MACHADO, R. F. 2008. Pulmonary HypertensionAn Increasingly Recognized Complication of Hereditary Hemolytic Anemias and HIV Infection. *JAMA,* 299**,** 324-331.

BARREIRO, L. B., TAILLEUX, L., PAI, A. A., GICQUEL, B., MARIONI, J. C. & GILAD, Y. 2012. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. *Proceedings of the National Academy of Sciences,* 109**,** 1204.

BARST, R. J., LANGLEBEN, D., FROST, A., HORN, E. M., OUDIZ, R., SHAPIRO, S., MCLAUGHLIN, V., HILL, N., TAPSON, V. F., ROBBINS, I. M., ZWICKE, D., DUNCAN, B., DIXON, R. A. F. & FRUMKIN, L. R. 2004. Sitaxsentan Therapy for Pulmonary Arterial Hypertension. *American Journal of Respiratory and Critical Care Medicine,* 169**,** 441-447.

BARST, R. J., MCGOON, M., MCLAUGHLIN, V., TAPSON, V., OUDIZ, R., SHAPIRO, S., ROBBINS, I. M., CHANNICK, R., BADESCH, D., RAYBURN, B. K., FLINCHBAUGH, R., SIGMAN, J., ARNESON, C. & JEFFS, R. 2003. Beraprost therapy for pulmonary arterial hypertension. *Journal of the American College of Cardiology,* 41**,** 2119-2125.

BARST, R. J., RUBIN, L. J., LONG, W. A., MCGOON, M. D., RICH, S., BADESCH, D. B., GROVES, B. M., TAPSON, V. F., BOURGE, R. C., BRUNDAGE, B. H., KOERNER, S. K., LANGLEBEN, D., KELLER, C. A., MURALI, S., URETSKY, B. F., CLAYTON, L. M., JÖBSIS, M. M., BLACKBURN, S. D., SHORTINO, D. & CROW, J. W. 1996. A Comparison of Continuous Intravenous Epoprostenol (Prostacyclin) with Conventional Therapy for Primary Pulmonary Hypertension. *New England Journal of Medicine,* 334**,** 296-301.

BATTLE, A., KHAN, Z., WANG, S. H., MITRANO, A., FORD, M. J., PRITCHARD, J. K. & GILAD, Y. 2015. Impact of regulatory variation from RNA to protein. *Science,* 347**,** 664.

BATTLE, A. & MONTGOMERY, S. B. 2014. Determining causality and consequence of expression quantitative trait loci. *Human genetics,* 133**,** 727-735.

BEGUIN, Y. 2003. Soluble transferrin receptor for the evaluation of erythropoiesis and iron status. *Clinica Chimica Acta,* 329**,** 9-22.

BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., KEIRA CHEETHAM, R., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M. J., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M. D., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A., BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CHIARA E. CATENAZZI, M., CHANG, S., NEIL COOLEY, R., CRAKE, N. R., DADA, O. O., DIAKOUMAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J., EGBUJOR, U. C., ELMORE, D. W., ETCHIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FUENTES FAJARDO, K. V., SCOTT FUREY, W., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D.,

HEYER, N. I., HIMS, M. M., HO, J. T., HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-59.

BENYAMIN, B., ESKO, T., RIED, J. S., RADHAKRISHNAN, A., VERMEULEN, S. H., TRAGLIA, M., GOGELE, M., ANDERSON, D., BROER, L., PODMORE, C., LUAN, J., KUTALIK, Z., SANNA, S., VAN DER MEER, P., TANAKA, T., WANG, F., WESTRA, H. J., FRANKE, L., MIHAILOV, E., MILANI, L., HALLDIN, J., WINKELMANN, J., MEITINGER, T., THIERY, J., PETERS, A., WALDENBERGER, M., RENDON, A., JOLLEY, J., SAMBROOK, J., KIEMENEY, L. A., SWEEP, F. C., SALA, C. F., SCHWIENBACHER, C., PICHLER, I., HUI, J., DEMIRKAN, A., ISAACS, A., AMIN, N., STERI, M., WAEBER, G., VERWEIJ, N., POWELL, J. E., NYHOLT, D. R., HEATH, A. C., MADDEN, P. A., VISSCHER, P. M., WRIGHT, M. J., MONTGOMERY, G. W., MARTIN, N. G., HERNANDEZ, D., BANDINELLI, S., VAN DER HARST, P., UDA, M., VOLLENWEIDER, P., SCOTT, R. A., LANGENBERG, C., WAREHAM, N. J., INTERACT, C., VAN DUIJN, C., BEILBY, J., PRAMSTALLER, P. P., HICKS, A. A., OUWEHAND, W. H., OEXLE, K., GIEGER, C., METSPALU, A., CAMASCHELLA, C., TONIOLO, D., SWINKELS, D. W. & WHITFIELD, J. B. 2014. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat Commun,* 5**,** 4926.

BIRNEY, E. & SORANZO, N. 2015. Human genomics: The end of the start for population sequencing. *Nature,* 526**,** 52-53.

BONDER, M. J., KURILSHIKOV, A., TIGCHELAAR, E. F., MUJAGIC, Z., IMHANN, F., VILA, A. V., DEELEN, P., VATANEN, T., SCHIRMER, M., SMEEKENS, S. P., ZHERNAKOVA, D. V., JANKIPERSADSING, S. A., JAEGER, M., OOSTING, M., CENIT, M. C., MASCLEE, A. A. M., SWERTZ, M. A., LI, Y., KUMAR, V., JOOSTEN, L., HARMSEN, H., WEERSMA, R. K., FRANKE, L., HOFKER, M. H., XAVIER, R. J., JONKERS, D., NETEA, M. G., WIJMENGA, C., FU, J. & ZHERNAKOVA, A. 2016. The effect of host genetics on the gut microbiome. *Nature Genetics,* 48**,** 1407-1412.

BOWDEN, J., DAVEY SMITH, G. & BURGESS, S. 2015. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology,* 44**,** 512-525.

BOWDEN, J., DAVEY SMITH, G., HAYCOCK, P. C. & BURGESS, S. 2016. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic epidemiology,* 40**,** 304-314.

BRUNNER, R., DUNBAR-JACOB, J., LEBOFF, M. S., GRANEK, I., BOWEN, D., SNETSELAAR, L. G., SHUMAKER, S. A., OCKENE, J., ROSAL, M., WACTAWSKI-WENDE, J., CAULEY, J., COCHRANE, B., TINKER, L., JACKSON, R., WANG, C. Y. & WU, L. 2009. Predictors of adherence in the Women's Health Initiative Calcium and Vitamin D Trial. *Behavioral medicine (Washington, D.C.),* 34**,** 145-155.

BULTINK, I. E. M., LEMS, W. F., VAN DE STADT, R. J., DINANT, H. J., LEYTE, A., PARK, D. S., DE KONING, M. H. M. T. & DIJKMANS, B. A. C. 2001. Ferritin and serum transferrin receptor predict iron deficiency in anemic patients with rheumatoid arthritis. *Arthritis & Rheumatism,* 44**,** 979-981.

BUNIELLO, A., MACARTHUR, J. A. L., CEREZO, M., HARRIS, L. W., HAYHURST, J., MALANGONE, C., MCMAHON, A., MORALES, J., MOUNTJOY, E., SOLLIS, E., SUVEGES, D., VROUSGOU, O., WHETZEL, P. L., AMODE, R., GUILLEN, J. A., RIAT, H. S., TREVANION, S. J., HALL, P., JUNKINS, H., FLICEK, P., BURDETT, T., HINDORFF, L. A., CUNNINGHAM, F. & PARKINSON, H. 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*.

BURGESS, S. 2014. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *International Journal of Epidemiology,* 43**,** 922-929.

BURGESS, S., BUTTERWORTH, A. & THOMPSON, S. G. 2013. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology,* 37**,** 658-665.

BURGESS, S., FOLEY, C. N., ALLARA, E., STALEY, J. R. & HOWSON, J. M. M. 2020. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nature Communications,* 11**,** 376.

BURGESS, S., FOLEY, C. N. & ZUBER, V. 2017a. Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data. *Annu Rev Genomics Hum Genet*.

BURGESS, S., SCOTT, R. A., TIMPSON, N. J., DAVEY SMITH, G., THOMPSON, S. G. & CONSORTIUM, E.-I. 2015. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol,* 30**,** 543-52.

BURGESS, S., SMALL, D. S. & THOMPSON, S. G. 2017b. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res,* 26**,** 2333-2355.

BYCROFT, C., FREEMAN, C., PETKOVA, D., BAND, G., ELLIOTT, L. T., SHARP, K., MOTYER, A., VUKCEVIC, D., DELANEAU, O., O'CONNELL, J., CORTES, A., WELSH, S., YOUNG, A., EFFINGHAM, M., MCVEAN, G., LESLIE, S., ALLEN, N., DONNELLY, P. & MARCHINI, J. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature,* 562**,** 203-209.

ÇALIŞKAN, M., BAKER, S. W., GILAD, Y. & OBER, C. 2015. Host Genetic Variation Influences Gene Expression Response to Rhinovirus Infection. *PLOS Genetics,* 11**,** e1005111.

CAMASCHELLA, C. 2015. Iron-Deficiency Anemia. *New England Journal of Medicine,* 372**,** 1832-1843.

CASH, J. M. & SEARS, D. A. 1989. The anemia of chronic disease: Spectrum of associated diseases in a series of unselected hospitalized patients. *The American Journal of Medicine,* 87**,** 638-644.

CHANG, C. C., CHOW, C. C., TELLIER, L. C. A. M., VATTIKUTI, S., PURCELL, S. M. & LEE, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience,* 4.

CHARALAMPOPOULOS, A., LEWIS, R., HICKEY, P., DURRINGTON, C., ELLIOT, C., CONDLIFFE, R., SABROE, I. & KIELY, D. G. 2018. Pathophysiology and Diagnosis of Pulmonary Hypertension Due to Left Heart Disease. *Frontiers in medicine,* 5**,** 174-174.

CHARLES-EDWARDS, G., AMARAL, N., SLEIGH, A., AYIS, S., CATIBOG, N., MCDONAGH, T., MONAGHAN, M., AMIN-YOUSSEF, G., KEMP, G. J., SHAH, A. M. & OKONKO, D. O. 2019. Effect of Iron Isomaltoside on Skeletal Muscle Energetics in Patients With Chronic Heart Failure and Iron Deficiency. *Circulation,* 139**,** 2386-2398.

CHARLES, H., GOOD, C. B., HANUSA, B. H., CHANG, C.-C. H. & WHITTLE, J. 2003. Racial differences in adherence to cardiac medications. *Journal of the National Medical Association,* 95**,** 17-27.

CHEADLE, C., BERGER, A. E., MATHAI, S. C., GRIGORYEV, D. N., WATKINS, T. N., SUGAWARA, Y., BARKATAKI, S., FAN, J., BOORGULA, M., HUMMERS, L., ZAIMAN, A. L., GIRGIS, R., MCDEVITT, M. A., JOHNS, R. A., WIGLEY, F., BARNES, K. C. & HASSOUN, P. M. 2012. Erythroid-Specific Transcriptional Changes in PBMCs from Pulmonary Hypertension Patients. *PLOS ONE,* 7**,** e34951.

CHEN, G., ZUO, S., TANG, J., ZUO, C., JIA, D., LIU, Q., LIU, G., ZHU, Q., WANG, Y., ZHANG, J., SHEN, Y., CHEN, D., YUAN, P., QIN, Z., RUAN, C., YE, J., WANG, X.-J., ZHOU, Y., GAO, P., ZHANG, P., LIU, J., JING, Z.-C., LU, A. & YU, Y. 2018. Inhibition of CRTH2-mediated Th2 activation attenuates pulmonary hypertension in mice. *Journal of Experimental Medicine,* 215**,** 2175-2195.

CHOI, S. W., MAK, T. S. H. & O'REILLY, P. 2018. A guide to performing Polygenic Risk Score analyses. *bioRxiv*.

CHRISTMAN, B. W., MCPHERSON, C. D., NEWMAN, J. H., KING, G. A., BERNARD, G. R., GROVES, B. M. & LOYD, J. E. 1992. An Imbalance between the Excretion of Thromboxane and Prostacyclin Metabolites in Pulmonary Hypertension. *New England Journal of Medicine,* 327**,** 70-75.

COLLABORATION, C. R. P. C. H. D. G., WENSLEY, F., GAO, P., BURGESS, S., KAPTOGE, S., DI ANGELANTONIO, E., SHAH, T., ENGERT, J. C., CLARKE, R., DAVEY-SMITH, G., NORDESTGAARD, B. G., SALEHEEN, D., SAMANI, N. J., SANDHU, M., ANAND, S., PEPYS, M. B., SMEETH, L., WHITTAKER, J., CASAS, J. P., THOMPSON, S. G., HINGORANI, A. D. & DANESH, J. 2011. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ (Clinical research ed.),* 342**,** d548-d548.

COLLABORATION, H. L. T. 1998. Lowering blood homocysteine with folic acid based supplements: meta-analysis of randomised trials. *BMJ,* 316**,** 894.

COLLABORATION, H. S. 2002. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. *JAMA*.

COOK, J. D., DASSENKO, S. & SKIKNE, B. S. 1990. Serum transferrin receptor as an index of iron absorption. *British Journal of Haematology,* 75**,** 603-609.

COTRONEO, E., ASHEK, A., WANG, L., WHARTON, J., DUBOIS, O., BOZORGI, S., BUSBRIDGE, M., ALAVIAN, K. N., WILKINS, M. R. & ZHAO, L. 2015. Iron homeostasis and pulmonary hypertension: iron deficiency leads to pulmonary vascular remodeling in the rat. *Circ Res,* 116**,** 1680-90.

CRAWFORD, D. C., CROSSLIN, D. R., TROMP, G., KULLO, I. J., KUIVANIEMI, H., HAYES, M. G., DENNY, J. C., BUSH, W. S., HAINES, J. L., RODEN, D. M., MCCARTY, C. A., JARVIK, G. P. & RITCHIE, M. D. 2014. eMERGEing progress in genomics—the first seven years. *Frontiers in Genetics,* 5.

DALEY, E., EMSON, C., GUIGNABERT, C., DE WAAL MALEFYT, R., LOUTEN, J., KURUP, V. P., HOGABOAM, C., TARASEVICIENE-STEWART, L., VOELKEL, N. F., RABINOVITCH, M., GRUNIG, E. & GRUNIG, G. 2008. Pulmonary arterial remodeling induced by a Th2 immune response. *Journal of Experimental Medicine,* 205**,** 361-372.

DANCIK, G. M. 2015. An online tool for evaluating diagnostic and prognostic gene expression biomarkers in bladder cancer. *BMC urology,* 15**,** 59-59.

DAVEY SMITH, G. & EBRAHIM, S. 2003. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. *International Journal of Epidemiology,* 32**,** 1-22.

DAVID, L., FEIGE, J.-J. & BAILLY, S. 2009. Emerging role of bone morphogenetic proteins in angiogenesis. *Cytokine & Growth Factor Reviews,* 20**,** 203-212.

DAVIE, N., HALEEN, S. J., UPTON, P. D., POLAK, J. M., YACOUB, M. H., MORRELL, N. W. & WHARTON, J. 2002. ETA and ETB Receptors Modulate the Proliferation of Human Pulmonary Artery Smooth Muscle Cells. *American Journal of Respiratory and Critical Care Medicine,* 165**,** 398-405.

DE DOMENICO, I., WARD, D. M. & KAPLAN, J. 2007. Hepcidin regulation: ironing out the details. *The Journal of Clinical Investigation,* 117**,** 1755-1758.

DIB, H., TAMBY, M. C., BUSSONE, G., REGENT, A., BEREZNÉ, A., LAFINE, C., BROUSSARD, C., SIMONNEAU, G., GUILLEVIN, L., WITKO-SARSAT, V., HUMBERT, M. & MOUTHON, L. 2012. Targets of anti-endothelial cell antibodies in pulmonary hypertension and scleroderma. *European Respiratory Journal,* 39**,** 1405.

DURBIN, R. M., ALTSHULER, D., DURBIN, R. M., ABECASIS, G. R., BENTLEY, D. R., CHAKRAVARTI, A., CLARK, A. G., COLLINS, F. S., DE LA VEGA, F. M., DONNELLY, P., EGHOLM, M., FLICEK, P., GABRIEL, S. B., GIBBS, R. A., KNOPPERS, B. M., LANDER, E. S., LEHRACH, H., MARDIS, E. R., MCVEAN, G. A., NICKERSON, D. A., PELTONEN, L., SCHAFER, A. J., SHERRY, S. T., WANG, J., WILSON, R. K., GIBBS, R. A., DEIROS, D., METZKER, M., MUZNY, D., REID, J., WHEELER, D., WANG, J., LI, J., JIAN, M., LI, G., LI, R., LIANG, H., TIAN, G., WANG, B., WANG, J., WANG, W., YANG, H., ZHANG, X., ZHENG, H., LANDER, E. S., ALTSHULER, D., AMBROGIO, L., BLOOM, T., CIBULSKIS, K., FENNELL, T. J., GABRIEL, S. B., JAFFE, D. B., SHEFLER, E., SOUGNEZ, C. L., BENTLEY, D. R., GORMLEY, N., HUMPHRAY, S., KINGSBURY, Z., KOKKO-GONZALES, P., STONE, J., MCKERNAN, K. J., COSTA, G. L., ICHIKAWA, J. K., LEE, C. C., SUDBRAK, R., LEHRACH, H., BORODINA, T. A., DAHL, A., DAVYDOV, A. N., MARQUARDT, P., MERTES, F., NIETFELD, W., ROSENSTIEL, P., SCHREIBER, S., SOLDATOV, A. V., TIMMERMANN, B., TOLZMANN, M., EGHOLM, M., AFFOURTIT, J., ASHWORTH, D., ATTIYA, S., BACHORSKI, M., BUGLIONE, E., BURKE, A., CAPRIO, A., CELONE, C., CLARK, S., CONNERS, D., DESANY, B., GU, L., GUCCIONE, L., KAO, K., KEBBEL, A., KNOWLTON, J., LABRECQUE, M., MCDADE, L., MEALMAKER, C., MINDERMAN, M., NAWROCKI, A., NIAZI, F., et al. 2010. A map of human genome variation from population-scale sequencing. *Nature,* 467**,** 1061-1073.

EDDAHIBI, S., GUIGNABERT, C., BARLIER-MUR, A.-M., DEWACHTER, L., FADEL, E., DARTEVELLE, P., HUMBERT, M., SIMONNEAU, G., HANOUN, N., SAURINI, F., HAMON, M. & ADNOT, S. 2006. Cross Talk Between Endothelial and Smooth Muscle Cells in Pulmonary Hypertension. *Circulation,* 113**,** 1857-1864.

EMANS, M. E., VAN DER PUTTEN, K., VAN ROOIJEN, K. L., KRAAIJENHAGEN, R. J., SWINKELS, D., VAN SOLINGE, W. W., CRAMER, M. J., DOEVENDANS, P. A. F. M., BRAAM, B. & GAILLARD, C. A. J. M. 2011. Determinants of Red Cell Distribution Width (RDW) in Cardiorenal Patients: RDW is Not Related to Erythropoietin Resistance. *Journal of Cardiac Failure,* 17**,** 626-633.

ESCRIBANO-SUBIAS, P., BLANCO, I., LÓPEZ-MESEGUER, M., LOPEZ-GUARCH, C. J., ROMAN, A., MORALES, P., CASTILLO-PALMA, M. J., SEGOVIA, J., GÓMEZ-SANCHEZ, M. A. & BARBERÀ, J. A. 2012. Survival in pulmonary hypertension in Spain: insights from the Spanish registry. *European Respiratory Journal,* 40**,** 596.

EVANGELOU, E., WARREN, H. R., MOSEN-ANSORENA, D., MIFSUD, B., PAZOKI, R., GAO, H., NTRITSOS, G., DIMOU, N., CABRERA, C. P., KARAMAN, I., NG, F. L., EVANGELOU, M., WITKOWSKA, K., TZANIS, E., HELLWEGE, J. N., GIRI, A., VELEZ EDWARDS, D. R., SUN, Y. V., CHO, K., GAZIANO, J. M., WILSON, P. W. F., TSAO, P. S., KOVESDY, C. P., ESKO, T., MÄGI, R., MILANI, L., ALMGREN, P., BOUTIN, T., DEBETTE, S., DING, J., GIULIANINI, F., HOLLIDAY, E. G., JACKSON, A. U., LI-GAO, R., LIN, W.-Y., LUAN, J. A., MANGINO, M., OLDMEADOW, C., PRINS, B. P., QIAN, Y., SARGURUPREMRAJ, M., SHAH, N., SURENDRAN, P., THÉRIAULT, S., VERWEIJ, N., WILLEMS, S. M., ZHAO, J.-H., AMOUYEL, P., CONNELL, J., DE MUTSERT, R., DONEY, A. S. F., FARRALL, M., MENNI, C., MORRIS, A. D., NOORDAM, R., PARÉ, G., POULTER, N. R., SHIELDS, D. C., STANTON, A., THOM, S., ABECASIS, G., AMIN, N., ARKING, D. E., AYERS, K. L., BARBIERI, C. M., BATINI, C., BIS, J. C., BLAKE, T., BOCHUD, M., BOEHNKE, M., BOERWINKLE, E., BOOMSMA, D. I., BOTTINGER, E. P., BRAUND, P. S., BRUMAT, M., CAMPBELL, A., CAMPBELL, H., CHAKRAVARTI, A., CHAMBERS, J. C., CHAUHAN, G., CIULLO, M., COCCA, M., COLLINS, F., CORDELL, H. J., DAVIES, G., DE BORST, M. H., DE GEUS, E. J., DEARY, I. J., DEELEN, J., DEL GRECO M, F., DEMIRKALE, C. Y., DÖRR, M., EHRET, G. B., ELOSUA, R., ENROTH, S., ERZURUMLUOGLU, A. M., FERREIRA, T., FRÅNBERG, M., FRANCO, O. H., GANDIN, I., et al. 2018. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature Genetics,* 50**,** 1412-1425.

EVANS, J. D. W., GIRERD, B., MONTANI, D., WANG, X.-J., GALIÈ, N., AUSTIN, E. D., ELLIOTT, G., ASANO, K., GRÜNIG, E., YAN, Y., JING, Z.-C., MANES, A., PALAZZINI, M., WHEELER, L. A., NAKAYAMA, I., SATOH, T., EICHSTAEDT, C., HINDERHOFER, K., WOLF, M., ROSENZWEIG, E. B., CHUNG, W. K., SOUBRIER, F., SIMONNEAU, G., SITBON, O., GRÄF, S., KAPTOGE, S., DI ANGELANTONIO, E., HUMBERT, M. & MORRELL, N. W. 2016. BMPR2 mutations and survival in pulmonary arterial hypertension: an individual participant data meta-analysis. *The Lancet Respiratory Medicine,* 4**,** 129-137.

EYRE, S., BOWES, J., DIOGO, D., LEE, A., BARTON, A., MARTIN, P., ZHERNAKOVA, A., STAHL, E., VIATTE, S., MCALLISTER, K., AMOS, C. I., PADYUKOV, L., TOES, R. E. M., HUIZINGA, T. W. J., WIJMENGA, C., TRYNKA, G., FRANKE, L., WESTRA, H.-J., ALFREDSSON, L., HU, X., SANDOR, C., DE BAKKER, P. I. W., DAVILA, S., KHOR, C. C., HENG, K. K., ANDREWS, R., EDKINS, S., HUNT, S. E., LANGFORD, C., SYMMONS, D., CONCANNON, P., ONENGUT-GUMUSCU, S., RICH, S. S., DELOUKAS, P., GONZALEZ-GAY, M. A., RODRIGUEZ-RODRIGUEZ, L., ÄRLSETIG, L., MARTIN, J., RANTAPÄÄ-DAHLQVIST, S., PLENGE, R. M., RAYCHAUDHURI, S., KLARESKOG, L., GREGERSEN, P. K., WORTHINGTON, J., BIOLOGICS IN RHEUMATOID ARTHRITIS, G., GENOMICS STUDY, S. & WELLCOME TRUST CASE CONTROL, C. 2012. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature Genetics,* 44**,** 1336-1340.

FAIRFAX, B. P., HUMBURG, P., MAKINO, S., NARANBHAI, V., WONG, D., LAU, E., JOSTINS, L., PLANT, K., ANDREWS, R., MCGEE, C. & KNIGHT, J. C. 2014. Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science,* 343**,** 1246949.

FINOTELLO, F., MAYER, C., PLATTNER, C., LASCHOBER, G., RIEDER, D., HACKL, H., KROGSDAM, A., LONCOVA, Z., POSCH, W., WILFLINGSEDER, D., SOPPER, S., IJSSELSTEIJN, M., BROUWER, T. P., JOHNSON, D., XU, Y., WANG, Y., SANDERS, M. E., ESTRADA, M. V., ERICSSON-GONZALEZ, P., CHAROENTONG, P., BALKO, J., DE MIRANDA, N. F. D. C. C. & TRAJANOSKI, Z. 2019. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Medicine,* 11**,** 34.

FRISE, M. C., CHENG, H.-Y., NICKOL, A. H., CURTIS, M. K., POLLARD, K. A., ROBERTS, D. J., RATCLIFFE, P. J., DORRINGTON, K. L. & ROBBINS, P. A. 2016. Clinical iron deficiency disturbs normal human responses to hypoxia. *The Journal of clinical investigation,* 126**,** 2139-2150.

FROST, A. E., LANGLEBEN, D., OUDIZ, R., HILL, N., HORN, E., MCLAUGHLIN, V., ROBBINS, I. M., SHAPIRO, S., TAPSON, V. F., ZWICKE, D., DEMARCO, T., SCHILZ, R., RUBENFIRE, M. & BARST, R. J. 2005. The 6-min walk test (6MW) as an efficacy endpoint in pulmonary arterial hypertension clinical trials: Demonstration of a ceiling effect. *Vascular Pharmacology,* 43**,** 36-39.

GAENZER, H., MARSCHANG, P., STURM, W., NEUMAYR, G. Ü., VOGEL, W., PATSCH, J. & WEISS, G. Ü. 2002. Association between increased iron stores and impaired endothelial function in patients with hereditary hemochromatosis. *Journal of the American College of Cardiology,* 40**,** 2189-2194.

GALIÈ, N., GHOFRANI, H. A., TORBICKI, A., BARST, R. J., RUBIN, L. J., BADESCH, D., FLEMING, T., PARPIA, T., BURGESS, G., BRANZI, A., GRIMMINGER, F., KURZYNA, M. & SIMONNEAU, G. 2005. Sildenafil Citrate Therapy for Pulmonary Arterial Hypertension. *New England Journal of Medicine,* 353**,** 2148-2157.

GALIE, N., HUMBERT, M., VACHIERY, J. L., GIBBS, S., LANG, I., TORBICKI, A., SIMONNEAU, G., PEACOCK, A., VONK NOORDEGRAAF, A., BEGHETTI, M., GHOFRANI, A., GOMEZ SANCHEZ, M. A., HANSMANN, G., KLEPETKO, W., LANCELLOTTI, P., MATUCCI, M., MCDONAGH, T., PIERARD, L. A., TRINDADE, P. T., ZOMPATORI, M., HOEPER, M. & GROUP, E. S. C. S. D. 2016. 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,* 37**,** 67-119.

GAMAZON, E. R., SEGRÈ, A. V., VAN DE BUNT, M., WEN, X., XI, H. S., HORMOZDIARI, F., ONGEN, H., KONKASHBAEV, A., DERKS, E. M., AGUET, F., QUAN, J., NICOLAE, D. L., ESKIN, E., KELLIS, M., GETZ, G., MCCARTHY, M. I., DERMITZAKIS, E. T., COX, N. J., ARDLIE, K. G. & CONSORTIUM, G. T. 2018. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature Genetics,* 50**,** 956-967.

GERMAIN, M., EYRIES, M., MONTANI, D., POIRIER, O., GIRERD, B., DORFMÜLLER, P., COULET, F., NADAUD, S., MAUGENRE, S., GUIGNABERT, C., CARPENTIER, W., VONK-NOORDEGRAAF, A., LÉVY, M., CHAOUAT, A., LAMBERT, J.-C., BERTRAND, M., DUPUY, A.-M., LETENNEUR, L., LATHROP, M., AMOUYEL, P., DE RAVEL, T. J. L., DELCROIX, M., AUSTIN, E. D., ROBBINS, I. M., HEMNES, A. R., LOYD, J. E., BERMAN-ROSENZWEIG, E., BARST, R. J., CHUNG, W. K., SIMONNEAU, G., TRÉGOUËT, D. A., HUMBERT, M. & SOUBRIER, F. 2013. Genome-wide association analysis identifies a susceptibility locus for pulmonary arterial hypertension. *Nature genetics,* 45**,** 518-521.

GHOFRANI, H.-A., GALIÈ, N., GRIMMINGER, F., GRÜNIG, E., HUMBERT, M., JING, Z.-C., KEOGH, A. M., LANGLEBEN, D., KILAMA, M. O., FRITSCH, A., NEUSER, D. & RUBIN, L. J. 2013. Riociguat for the Treatment of Pulmonary Arterial Hypertension. *New England Journal of Medicine,* 369**,** 330-340.

GHOFRANI, H. A., PEPKE-ZABA, J., BARBERA, J. A., CHANNICK, R., KEOGH, A. M., GOMEZ-SANCHEZ, M. A., KNEUSSL, M. & GRIMMINGER, F. 2004. Nitric oxide pathway and phosphodiesterase inhibitors in pulmonary arterial hypertension. *Journal of the American College of Cardiology,* 43**,** S68.

GIAID, A., YANAGISAWA, M., LANGLEBEN, D., MICHEL, R. P., LEVY, R., SHENNIB, H., KIMURA, S., MASAKI, T., DUGUID, W. P. & STEWART, D. J. 1993. Expression of Endothelin-1 in the Lungs of Patients with Pulmonary Hypertension. *New England Journal of Medicine,* 328**,** 1732-1739.

GILL, D., DEL GRECO, M. F., WALKER, A. P., SRAI, S. K. S., LAFFAN, M. A. & MINELLI, C. 2017. The Effect of Iron Status on Risk of Coronary Artery Disease: A Mendelian Randomization Study-Brief Report. *Arterioscler Thromb Vasc Biol,* 37**,** 1788-1792.

GILL, D., **MONORI, G.**, **TZOULAKI, I.** & **DEHGAN, A.** 2018. Iron Status and Risk of Stroke- A Mendelian Randomization Study. *Stroke*.

GONORAZKY, H. D., NAUMENKO, S., RAMANI, A. K., NELAKUDITI, V., MASHOURI, P., WANG, P., KAO, D., OHRI, K., VITHTHIYAPASKARAN, S., TARNOPOLSKY, M. A., MATHEWS, K. D., MOORE, S. A., OSORIO, A. N., VILLANOVA, D., KEMALADEWI, D. U., COHN, R. D., BRUDNO, M. & DOWLING, J. J. 2019. Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *The American Journal of Human Genetics,* 104**,** 466-483.

GOV, E. & ARGA, K. Y. 2017. Differential co-expression analysis reveals a novel prognostic gene module in ovarian cancer. *Scientific Reports,* 7**,** 4996.

GRAF, S., HAIMEL, M., BLEDA, M., HADINNAPOLA, C., SOUTHGATE, L., LI, W., HODGSON, J., LIU, B., SALMON, R. M., SOUTHWOOD, M., MACHADO, R. D., MARTIN, J. M., TREACY, C. M., YATES, K., DAUGHERTY, L. C., SHAMARDINA, O., WHITEHORN, D., HOLDEN, S., ALDRED, M., BOGAARD, H. J., CHURCH, C., COGHLAN, G., CONDLIFFE, R., CORRIS, P. A., DANESINO, C., EYRIES, M., GALL, H., GHIO, S., GHOFRANI, H. A., GIBBS, J. S. R., GIRERD, B., HOUWELING, A. C., HOWARD, L., HUMBERT, M., KIELY, D. G., KOVACS, G., MACKENZIE ROSS, R. V., MOLEDINA, S., MONTANI, D., NEWNHAM, M., OLSCHEWSKI, A., OLSCHEWSKI, H., PEACOCK, A. J., PEPKE-ZABA, J., PROKOPENKO, I., RHODES, C. J., SCELSI, L., SEEGER, W., SOUBRIER, F., STEIN, D. F., SUNTHARALINGAM, J., SWIETLIK, E. M., TOSHNER, M. R., VAN HEEL, D. A., VONK NOORDEGRAAF, A., WAISFISZ, Q., WHARTON, J., WORT, S. J., OUWEHAND, W. H., SORANZO, N., LAWRIE, A., UPTON, P. D., WILKINS, M. R., TREMBATH, R. C. & MORRELL, N. W. 2018. Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nat Commun,* 9**,** 1416.

GRAY, R. & WHEATLEY, K. 1991. How to avoid bias when comparing bone marrow transplantation with chemotherapy. . *Bone Marrow Transplant* 7**,** 9-12.

GROUP, H. P. S. C. 2002. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20 536 high-risk individuals: a randomised placebo-controlled trial. *The Lancet,* 360**,** 23-33.

GRUNDBERG, E., ADOUE, V., KWAN, T., GE, B., DUAN, Q. L., LAM, K. C. L., KOKA, V., KINDMARK, A., WEISS, S. T., TANTISIRA, K., MALLMIN, H., RABY, B. A., NILSSON, O. & PASTINEN, T. 2011. Global Analysis of the Impact of Environmental Perturbation on cis-Regulation of Gene Expression. *PLOS Genetics,* 7**,** e1001279.

GTEXPORTAL. *Dataset Summary of Analysis Samples* [Online]. Available: https://gtexportal.org/home/tissueSummaryPage [Accessed 18/02 2020].

HALPERN, S. D. & TAICHMAN, D. B. 2009. Misclassification of Pulmonary Hypertension Due to Reliance on Pulmonary Capillary Wedge Pressure Rather Than Left Ventricular End-Diastolic Pressure. *Chest,* 136**,** 37-43.

HAMPOLE, C. V., MEHROTRA, A. K., THENAPPAN, T., GOMBERG-MAITLAND, M. & SHAH, S. J. 2009. Usefulness of red cell distribution width as a prognostic marker in pulmonary hypertension. *Am J Cardiol,* 104**,** 868-72.

HANNON, E., WEEDON, M., BRAY, N., O'DONOVAN, M. & MILL, J. 2017. Pleiotropic Effects of Trait-Associated Genetic Variation on DNA Methylation: Utility for Refining GWAS Loci. *The American Journal of Human Genetics,* 100**,** 954-959.

HARRISON, R. E., BERGER, R., HAWORTH SHEILA, G., TULLOH, R., MACHE CHRISTOPH, J., MORRELL NICHOLAS, W., ALDRED MICHEALA, A. & TREMBATH RICHARD, C. 2005. Transforming Growth Factor-β Receptor Mutations and Pulmonary Arterial Hypertension in Childhood. *Circulation,* 111**,** 435-441.

HARRISON, R. E., FLANAGAN, J. A., SANKELO, M., ABDALLA, S. A., ROWELL, J., MACHADO, R. D., ELLIOTT, C. G., ROBBINS, I. M., OLSCHEWSKI, H., MCLAUGHLIN, V., GRUENIG, E., KERMEEN, F., LAITINEN, T., MORRELL, N. W. & TREMBATH, R. C. 2003. Molecular and functional analysis identifies ALK-1 as the predominant cause of pulmonary hypertension related to hereditary haemorrhagic telangiectasia. *Journal of Medical Genetics,* 40**,** 865.

HARST, P. V. D. & VERWEIJ, N. 2018. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circulation Research,* 122**,** 433-443.

HASIN, Y., SELDIN, M. & LUSIS, A. 2017. Multi-omics approaches to disease. *Genome Biology,* 18**,** 83.

HASSOUN, P. M. 2014. Inflammation in pulmonary arterial hypertension: is it time to quell the fire? *European Respiratory Journal,* 43**,** 685.

HEAD, S. R., KOMORI, H. K., LAMERE, S. A., WHISENANT, T., VAN NIEUWERBURGH, F., SALOMON, D. R. & ORDOUKHANIAN, P. 2014. Library construction for next-generation sequencing: overviews and challenges. *BioTechniques,* 56**,** 61-passim.

HEMANI, G., ZHENG, J., ELSWORTH, B., WADE, K. H., HABERLAND, V., BAIRD, D., LAURIN, C., BURGESS, S., BOWDEN, J., LANGDON, R., TAN, V. Y., YARMOLINSKY, J., SHIHAB, H. A., TIMPSON, N. J., EVANS, D. M., RELTON, C., MARTIN, R. M., DAVEY SMITH, G., GAUNT, T. R. & HAYCOCK, P. C. 2018. The MR-Base platform supports systematic causal inference across the human phenome. *eLife,* 7**,** e34408.

HENNEKENS, C. H., BURING, J. E., MANSON, J. E., STAMPFER, M., ROSNER, B., COOK, N. R., BELANGER, C., LAMOTTE, F., GAZIANO, J. M., RIDKER, P. M., WILLETT, W. & PETO, R. 1996. Lack of Effect of Long-Term Supplementation with Beta Carotene on the Incidence of Malignant Neoplasms and Cardiovascular Disease. *New England Journal of Medicine,* 334**,** 1145-1149.

HENTZE, M. W., MUCKENTHALER, M. U., GALY, B. & CAMASCHELLA, C. 2010. Two to Tango: Regulation of Mammalian Iron Metabolism. *Cell,* 142**,** 24-38.

HIRATA, Y., EMORI, T., EGUCHI, S., KANNO, K., IMAI, T., OHTA, K. & MARUMO, F. 1993. Endothelin receptor subtype B mediates synthesis of nitric oxide by cultured bovine endothelial cells. *The Journal of clinical investigation,* 91**,** 1367-1373.

HOEPER MARIUS, M., MAYER, E., SIMONNEAU, G. & RUBIN LEWIS, J. 2006. Chronic Thromboembolic Pulmonary Hypertension. *Circulation,* 113**,** 2011-2020.

HOEPER, M. M., HUMBERT, M., SOUZA, R., IDREES, M., KAWUT, S. M., SLIWA-HAHNLE, K., JING, Z.-C. & GIBBS, J. S. R. 2016. A global view of pulmonary hypertension. *The Lancet Respiratory Medicine,* 4**,** 306-322.

HOEPER, M. M., HUSCHER, D., GHOFRANI, H. A., DELCROIX, M., DISTLER, O., SCHWEIGER, C., GRUNIG, E., STAEHLER, G., ROSENKRANZ, S., HALANK, M., HELD, M., GROHÉ, C., LANGE, T. J., BEHR, J., KLOSE, H., WILKENS, H., FILUSCH, A., GERMANN, M., EWERT, R., SEYFARTH, H.-J., OLSSON, K. M., OPITZ, C. F., GAINE, S. P., VIZZA, C. D., VONK-NOORDEGRAAF, A., KAEMMERER, H., GIBBS, J. S. R. & PITTROW, D. 2013. Elderly patients diagnosed with idiopathic pulmonary arterial hypertension: Results from the COMPERA registry. *International Journal of Cardiology,* 168**,** 871-880.

HOU, L., CHEN, M., ZHANG, C. K., CHO, J. & ZHAO, H. 2014. Guilt by rewiring: gene prioritization through network rewiring in genome wide association studies. *Human molecular genetics,* 23**,** 2780-2790.

HOWARD, B. V., ARAGAKI, A. K., TINKER, L. F., ALLISON, M., HINGLE, M. D., JOHNSON, K. C., MANSON, J. E., SHADYAB, A. H., SHIKANY, J. M., SNETSELAAR, L. G., THOMSON, C. A., ZASLAVSKY, O. & PRENTICE, R. L. 2018. A Low-Fat Dietary Pattern and Diabetes: A Secondary Analysis From the Women's Health Initiative Dietary Modification Trial. *Diabetes Care,* 41**,** 680.

HOWARD, B. V., VAN HORN, L., HSIA, J., MANSON, J. E., STEFANICK, M. L., WASSERTHEIL-SMOLLER, S., KULLER, L. H., LACROIX, A. Z., LANGER, R. D., LASSER, N. L., LEWIS, C. E., LIMACHER, M. C., MARGOLIS, K. L., MYSIW, W. J., OCKENE, J. K., PARKER, L. M., PERRI, M. G., PHILLIPS, L., PRENTICE, R. L., ROBBINS, J., ROSSOUW, J. E., SARTO, G. E., SCHATZ, I. J., SNETSELAAR, L. G., STEVENS, V. J., TINKER, L. F., TREVISAN, M., VITOLINS, M. Z., ANDERSON, G. L., ASSAF, A. R., BASSFORD, T., BERESFORD, S. A. A., BLACK, H. R., BRUNNER, R. L., BRZYSKI, R. G., CAAN, B., CHLEBOWSKI, R. T., GASS, M., GRANEK, I., GREENLAND, P., HAYS, J., HEBER, D., HEISS, G., HENDRIX, S. L., HUBBELL, F. A., JOHNSON, K. C. & KOTCHEN, J. M. 2006. Low-Fat Dietary Pattern and Risk of Cardiovascular DiseaseThe Women's Health Initiative Randomized Controlled Dietary Modification Trial. *JAMA,* 295**,** 655-666.

HOWARD, L. S. G. E., WATSON, G. M. J., WHARTON, J., RHODES, C. J., CHAN, K., KHENGAR, R., ROBBINS, P. A., KIELY, D. G., CONDLIFFE, R., ELLIOTT, C. A., PEPKE-ZABA, J., SHEARES, K., MORRELL, N. W., DAVIES, R., ASHBY, D., GIBBS, J. S. R. & WILKINS, M. R. 2013. Supplementation of Iron in Pulmonary Hypertension: Rationale and Design of a Phase II Clinical Trial in Idiopathic Pulmonary Arterial Hypertension. *Pulmonary Circulation,* 3**,** 100-107.

INTERLEUKIN-6 RECEPTOR MENDELIAN RANDOMISATION ANALYSIS, C., SWERDLOW, D. I., HOLMES, M. V., KUCHENBAECKER, K. B., ENGMANN, J. E. L., SHAH, T., SOFAT, R., GUO, Y., CHUNG, C., PEASEY, A., PFISTER, R., MOOIJAART, S. P., IRELAND, H. A., LEUSINK, M., LANGENBERG, C., LI, K. W., PALMEN, J., HOWARD, P., COOPER, J. A., DRENOS, F., HARDY, J., NALLS, M. A., LI, Y. R., LOWE, G., STEWART, M., BIELINSKI, S. J., PETO, J., TIMPSON, N. J., GALLACHER, J., DUNLOP, M., HOULSTON, R., TOMLINSON, I., TZOULAKI, I., LUAN, J. A., BOER, J. M. A., FOROUHI, N. G., ONLAND-MORET, N. C., VAN DER SCHOUW, Y. T., SCHNABEL, R. B., HUBACEK, J. A., KUBINOVA, R., BACEVICIENE, M., TAMOSIUNAS, A., PAJAK, A., TOPOR-MADRY, R., MALYUTINA, S., BALDASSARRE, D., SENNBLAD, B., TREMOLI, E., DE FAIRE, U., FERRUCCI, L., BANDENELLI, S., TANAKA, T., MESCHIA, J. F., SINGLETON, A., NAVIS, G., MATEO LEACH,

I., BAKKER, S. J. L., GANSEVOORT, R. T., FORD, I., EPSTEIN, S. E., BURNETT, M. S., DEVANEY, J. M., JUKEMA, J. W., WESTENDORP, R. G. J., JAN DE BORST, G., VAN DER GRAAF, Y., DE JONG, P. A., MAILAND-VAN DER ZEE, A.-H., KLUNGEL, O. H., DE BOER, A., DOEVENDANS, P. A., STEPHENS, J. W., EATON, C. B., ROBINSON, J. G., MANSON, J. E., FOWKES, F. G., FRAYLING, T. M., PRICE, J. F., WHINCUP, P. H., MORRIS, R. W., LAWLOR, D. A., SMITH, G. D., BEN-SHLOMO, Y., REDLINE, S., LANGE, L. A., KUMARI, M., WAREHAM, N. J., VERSCHUREN, W. M. M., BENJAMIN, E. J., WHITTAKER, J. C., HAMSTEN, A., DUDBRIDGE, F., DELANEY, J. A. C., WONG, A., KUH, D., HARDY, R., CASTILLO, B. A., CONNOLLY, J. J., VAN DER HARST, P., et al. 2012. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet (London, England),* 379**,** 1214-1224.

IVAN, M., KONDO, K. & YANG, H. 2001. HIFa targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science***,** 464–468.

JAMIESON, S. W., KAPELANSKI, D. P., SAKAKIBARA, N., MANECKE, G. R., THISTLETHWAITE, P. A., KERR, K. M., CHANNICK, R. N., FEDULLO, P. F. & AUGER, W. R. 2003. Pulmonary endarterectomy: experience and lessons learned in 1,500 cases. *The Annals of Thoracic Surgery,* 76**,** 1457-1464.

JANKOWSKA, E. A., ROZENTRYT P FAU - WITKOWSKA, A., WITKOWSKA A FAU - NOWAK, J., NOWAK J FAU - HARTMANN, O., HARTMANN O FAU - PONIKOWSKA, B., PONIKOWSKA B FAU - BORODULIN-NADZIEJA, L., BORODULIN-NADZIEJA L FAU - BANASIAK, W., BANASIAK W FAU - POLONSKI, L., POLONSKI L FAU - FILIPPATOS, G., FILIPPATOS G FAU - MCMURRAY, J. J. V., MCMURRAY JJ FAU - ANKER, S. D., ANKER SD FAU - PONIKOWSKI, P. & PONIKOWSKI, P. 2010. Iron deficiency: an ominous sign in patients with systolic chronic heart failure. *Eur Heart J*.

JING, Z.-C., PARIKH, K., PULIDO, T., JERJES-SANCHEZ, C., WHITE, R. J., ALLEN, R., TORBICKI, A., XU, K.-F., YEHLE, D., LALIBERTE, K., ARNESON, C. & RUBIN LEWIS, J. 2013. Efficacy and Safety of Oral Treprostinil Monotherapy for the Treatment of Pulmonary Arterial Hypertension. *Circulation,* 127**,** 624-633.

JOEHANES, R., ZHANG, X., HUAN, T., YAO, C., YING, S. X., NGUYEN, Q. T., DEMIRKALE, C. Y., FEOLO, M. L., SHAROPOVA, N. R., STURCKE, A., SCHAFFER, A. A., HEARD-COSTA, N., CHEN, H., LIU, P. C., WANG, R., WOODHOUSE, K. A., TANRIVERDI, K., FREEDMAN, J. E., RAGHAVACHARI, N., DUPUIS, J., JOHNSON, A. D., O'DONNELL, C. J., LEVY, D. & MUNSON, P. J. 2017. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol,* 18**,** 16.

JOHNSON, T., KUTALIK, Z., MANNING, A. & CHEN, H. 2009. QUICKTEST. 1.1 ed.

JUNCÀ, J., FERNÁNDEZ-AVILÉS, F., ORIOL, A., NAVARRO, J. T., MILLÁ, F., SANCHO, J. M. & FELIU, E. 1998. The usefulness of the serum transferrin receptor in detecting iron deficiency in the anemia of chronic disorders. *Haematologica,* 83**,** 676-680.

KANEHISA, M., ARAKI, M., GOTO, S., HATTORI, M., HIRAKAWA, M., ITOH, M., KATAYAMA, T., KAWASHIMA, S., OKUDA, S., TOKIMATSU, T. & YAMANISHI, Y. 2007. KEGG for linking genomes to life and the environment. *Nucleic Acids Research,* 36**,** D480-D484.

KERSTJENS-FREDERIKSE, W. S., BONGERS, E. M. H. F., ROOFTHOOFT, M. T. R., LETER, E. M., DOUWES, J. M., VAN DIJK, A., VONK-NOORDEGRAAF, A., DIJK-BOS, K. K., HOEFSLOOT, L. H., HOENDERMIS, E. S., GILLE, J. J. P., SIKKEMA-RADDATZ, B., HOFSTRA, R. M. W. & BERGER, R. M. F. 2013. TBX4 mutations (small patella syndrome) are associated with childhood-onset pulmonary arterial hypertension. *Journal of Medical Genetics,* 50**,** 500.

KHATRI, P., SIROTA, M. & BUTTE, A. J. 2012. Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *PLOS Computational Biology,* 8**,** e1002375.

KICHAEV, G., BHATIA, G., LOH, P.-R., GAZAL, S., BURCH, K., FREUND, M. K., SCHOECH, A., PASANIUC, B. & PRICE, A. L. 2019. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *The American Journal of Human Genetics,* 104**,** 65-75.

KING, E. A., DAVIS, J. W. & DEGNER, J. F. 2019. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *BioRxiv*.

KLERK, M., VERHOEF P FAU - CLARKE, R., CLARKE R FAU - BLOM, H. J., BLOM HJ FAU - KOK, F. J., KOK FJ FAU - SCHOUTEN, E. G. & SCHOUTEN, E. G. 2002. MTHFR 677C-->T polymorphism and risk of coronary heart disease: a meta-analysis. *JAMA*.

KOMMIREDDY, S., BHYRAVAVAJHALA, S., KURIMETI, K., CHENNAREDDY, S., KANCHINADHAM, S., RAJENDRA VARA PRASAD, I. & RAJASEKHAR, L. 2015. Pulmonary arterial hypertension in systemic lupus erythematosus may benefit by addition of immunosuppression to vasodilator therapy: an observational study. *Rheumatology,* 54**,** 1673-1679.

KRASUSKI, R. A., HART, S. A., SMITH, B., WANG, A., HARRISON, J. K. & BASHORE, T. M. 2011. Association of anemia and long-term survival in patients with pulmonary hypertension. *International Journal of Cardiology,* 150**,** 291-295.

KUKURBA, K. R. & MONTGOMERY, S. B. 2015. RNA Sequencing and Analysis. *Cold Spring Harbor protocols,* 2015**,** 951-969.

KUMAR, R. & GRAHAM, B. 2018. How does inflammation contribute to pulmonary hypertension? *European Respiratory Journal,* 51**,** 1702403.

LAM, C. S. P., DONAL, E., KRAIGHER-KRAINER, E. & VASAN, R. S. 2011. Epidemiology and clinical course of heart failure with preserved ejection fraction. *European Journal of Heart Failure,* 13**,** 18-28.

LANE, K. B., MACHADO, R. D., PAUCIULO, M. W., THOMSON, J. R., PHILLIPS, J. A., LOYD, J. E., NICHOLS, W. C. & TREMBATH, R. C. 2000. Heterozygous germline mutations in BMPR2, encoding a TGF-β receptor, cause familial primary pulmonary hypertension. *Nature Genetics,* 26**,** 81.

LANGE, A. W., HAITCHI, H. M., LECRAS, T. D., SRIDHARAN, A., XU, Y., WERT, S. E., JAMES, J., UDELL, N., THURNER, P. J. & WHITSETT, J. A. 2014. Sox17 is required for normal pulmonary vascular morphogenesis. *Developmental Biology,* 387**,** 109-120.

LARKIN, E. K., NEWMAN, J. H., AUSTIN, E. D., HEMNES, A. R., WHEELER, L., ROBBINS, I. M., WEST, J. D., PHILLIPS, J. A., HAMID, R. & LOYD, J. E. 2012. Longitudinal Analysis Casts Doubt on the Presence of Genetic Anticipation in Heritable Pulmonary Arterial Hypertension. *American Journal of Respiratory and Critical Care Medicine,* 186**,** 892-896.

LAU, E. M. T., GIANNOULATOU, E., CELERMAJER, D. S. & HUMBERT, M. 2017. Epidemiology and treatment of pulmonary arterial hypertension. *Nat Rev Cardiol,* 14**,** 603-614.

LAWLOR, D. A., HARBORD, R. M., STERNE, J. A., TIMPSON, N. & DAVEY SMITH, G. 2008. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med,* 27**,** 1133-63.

LEUNG, C. C., MOONDRA, V., CATHERWOOD, E. & ANDRUS, B. W. 2010. Prevalence and Risk Factors of Pulmonary Hypertension in Patients With Elevated Pulmonary Venous Pressure and Preserved Ejection Fraction. *The American Journal of Cardiology,* 106**,** 284-286.

LEWIS, S. J., ARAYA, R., SMITH, G. D., FREATHY, R., GUNNELL, D., PALMER, T. & MUNAFÒ, M. 2011. Smoking is associated with, but does not cause, depressed mood in pregnancy--a mendelian randomization study. *PloS one,* 6**,** e21689-e21689.

LIU, D. J., PELOSO, G. M., YU, H., BUTTERWORTH, A. S., WANG, X., MAHAJAN, A., SALEHEEN, D., EMDIN, C., ALAM, D., ALVES, A. C., AMOUYEL, P., DI ANGELANTONIO, E., ARVEILER, D., ASSIMES, T. L., AUER, P. L., BABER, U., BALLANTYNE, C. M., BANG, L. E., BENN, M., BIS, J. C., BOEHNKE, M., BOERWINKLE, E., BORK-JENSEN, J., BOTTINGER, E. P., BRANDSLUND, I., BROWN, M., BUSONERO, F., CAULFIELD, M. J., CHAMBERS, J. C., CHASMAN, D. I., CHEN, Y. E., CHEN, Y.-D. I., CHOWDHURY, R., CHRISTENSEN, C., CHU, A. Y., CONNELL, J. M., CUCCA, F., CUPPLES, L. A., DAMRAUER, S. M., DAVIES, G., DEARY, I. J., DEDOUSSIS, G., DENNY, J. C., DOMINICZAK, A., DUBÉ, M.-P., EBELING, T., EIRIKSDOTTIR, G., ESKO, T., FARMAKI, A.-E., FEITOSA, M. F., FERRARIO, M., FERRIERES, J., FORD, I., FORNAGE, M., FRANKS, P. W., FRAYLING, T. M., FRIKKE-SCHMIDT, R., FRITSCHE, L. G., FROSSARD, P., FUSTER, V., GANESH, S. K., GAO, W., GARCIA, M. E., GIEGER, C., GIULIANINI, F., GOODARZI, M. O., GRALLERT, H., GRARUP, N., GROOP, L., GROVE, M. L., GUDNASON, V., HANSEN, T., HARRIS, T. B., HAYWARD, C., HIRSCHHORN, J. N., HOLMEN, O. L., HUFFMAN, J., HUO, Y., HVEEM, K., JABEEN, S., JACKSON, A. U., JAKOBSDOTTIR, J., JARVELIN, M.-R., JENSEN, G. B., JØRGENSEN, M. E., JUKEMA, J. W., JUSTESEN, J. M., KAMSTRUP, P. R., KANONI, S., KARPE, F., KEE, F., KHERA, A. V., KLARIN, D., KOISTINEN, H. A., KOONER, J. S., KOOPERBERG, C., KUULASMAA, K., KUUSISTO, J., LAAKSO, M., LAKKA, T., et al. 2017. Exome-wide association study of plasma lipids in >300,000 individuals. *Nature Genetics,* 49**,** 1758-1766.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15**,** 550.

MA, L., ROMAN-CAMPOS, D., AUSTIN, E. D., EYRIES, M., SAMPSON, K. S., SOUBRIER, F., GERMAIN, M., TRÉGOUËT, D.-A., BORCZUK, A., ROSENZWEIG, E. B., GIRERD, B., MONTANI, D., HUMBERT, M., LOYD, J. E., KASS, R. S. & CHUNG, W. K. 2013. A Novel Channelopathy in Pulmonary Arterial Hypertension. *New England Journal of Medicine,* 369**,** 351-361.

MACHADO, R. D., PAUCIULO, M. W., THOMSON, J. R., LANE, K. B., MORGAN, N. V., WHEELER, L., PHILLIPS III, J. A., NEWMAN, J., WILLIAMS, D., GALIÈ, N., MANES, A., MCNEIL, K., YACOUB, M., MIKHAIL, G., ROGERS, P., CORRIS, P., HUMBERT, M., DONNAI, D., MARTENSSON, G., TRANEBJAERG, L., LOYD, J. E., TREMBATH, R. C. & NICHOLS, W. C. 2001. BMPR2 Haploinsufficiency as the Inherited Molecular

Mechanism for Primary Pulmonary Hypertension. *The American Journal of Human Genetics,* 68**,** 92-102.

MACHADO, R. D., SOUTHGATE, L., EICHSTAEDT, C. A., ALDRED, M. A., AUSTIN, E. D., BEST, D. H., CHUNG, W. K., BENJAMIN, N., ELLIOTT, C. G., EYRIES, M., FISCHER, C., GRÄF, S., HINDERHOFER, K., HUMBERT, M., KEILES, S. B., LOYD, J. E., MORRELL, N. W., NEWMAN, J. H., SOUBRIER, F., TREMBATH, R. C., VIALES, R. R. & GRÜNIG, E. 2015. Pulmonary Arterial Hypertension: A Current Perspective on Established and Emerging Molecular Genetic Defects. *Human mutation,* 36**,** 1113-1127.

MAHAJAN, A., TALIUN, D., THURNER, M., ROBERTSON, N. R., TORRES, J. M., RAYNER, N. W., PAYNE, A. J., STEINTHORSDOTTIR, V., SCOTT, R. A., GRARUP, N., COOK, J. P., SCHMIDT, E. M., WUTTKE, M., SARNOWSKI, C., MÄGI, R., NANO, J., GIEGER, C., TROMPET, S., LECOEUR, C., PREUSS, M. H., PRINS, B. P., GUO, X., BIELAK, L. F., BELOW, J. E., BOWDEN, D. W., CHAMBERS, J. C., KIM, Y. J., NG, M. C. Y., PETTY, L. E., SIM, X., ZHANG, W., BENNETT, A. J., BORK-JENSEN, J., BRUMMETT, C. M., CANOUIL, M., EC KARDT, K.-U., FISCHER, K., KARDIA, S. L. R., KRONENBERG, F., LÄLL, K., LIU, C.-T., LOCKE, A. E., LUAN, J. A., NTALLA, I., NYLANDER, V., SCHÖNHERR, S., SCHURMANN, C., YENGO, L., BOTTINGER, E. P., BRANDSLUND, I., CHRISTENSEN, C., DEDOUSSIS, G., FLOREZ, J. C., FORD, I., FRANCO, O. H., FRAYLING, T. M., GIEDRAITIS, V., HACKINGER, S., HATTERSLEY, A. T., HERDER, C., IKRAM, M. A., INGELSSON, M., JØRGENSEN, M. E., JØRGENSEN, T., KRIEBEL, J., KUUSISTO, J., LIGTHART, S., LINDGREN, C. M., LINNEBERG, A., LYSSENKO, V., MAMAKOU, V., MEITINGER, T., MOHLKE, K. L., MORRIS, A. D., NADKARNI, G., PANKOW, J. S., PETERS, A., SATTAR, N., STANČÁKOVÁ, A., STRAUCH, K., TAYLOR, K. D., THORAND, B., THORLEIFSSON, G., THORSTEINSDOTTIR, U., TUOMILEHTO, J., WITTE, D. R., DUPUIS, J., PEYSER, P. A., ZEGGINI, E., LOOS, R. J. F., FROGUEL, P., INGELSSON, E., LIND, L., GROOP, L., LAAKSO, M., COLLINS, F. S., JUKEMA, J. W., PALMER, C. N. A., GRALLERT, H., METSPALU, A., et al. 2018. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics,* 50**,** 1505-1513.

MAST, A. E., BLINDER, M. A., GRONOWSKI, A. M., CHUMLEY, C. & SCOTT, M. G. 1998. Clinical utility of the soluble transferrin receptor and comparison with serum ferritin in several populations. *Clinical Chemistry,* 44**,** 45-51.

MASTON, L. D., JONES, D. T., GIERMAKOWSKA, W., HOWARD, T. A., CANNON, J. L., WANG, W., WEI, Y., XUAN, W., RESTA, T. C. & GONZALEZ BOSC, L. V. 2017. Central role of T helper 17 cells in chronic hypoxia-induced pulmonary hypertension. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 312**,** L609-L624.

MAURANO, M. T., HUMBERT, R., RYNES, E., THURMAN, R. E., HAUGEN, E., WANG, H., REYNOLDS, A. P., SANDSTROM, R., QU, H., BRODY, J., SHAFER, A., NERI, F., LEE, K., KUTYAVIN, T., STEHLING-SUN, S., JOHNSON, A. K., CANFIELD, T. K., GISTE, E., DIEGEL, M., BATES, D., HANSEN, R. S., NEPH, S., SABO, P. J., HEIMFELD, S., RAUBITSCHEK, A., ZIEGLER, S., COTSAPAS, C., SOTOODEHNIA, N., GLASS, I., SUNYAEV, S. R., KAUL, R. & STAMATOYANNOPOULOS, J. A. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, N.Y.),* 337**,** 1190-1195.

MCGOON, M. D., BENZA, R. L., ESCRIBANO-SUBIAS, P., JIANG, X., MILLER, D. P., PEACOCK, A. J., PEPKE-ZABA, J., PULIDO, T., RICH, S., ROSENKRANZ, S., SUISSA, S. & HUMBERT, M. 2013. Pulmonary Arterial Hypertension: Epidemiology and Registries. *Journal of the American College of Cardiology,* 62**,** D51-D59.

MCVICKER, G., VAN DE GEIJN, B., DEGNER, J. F., CAIN, C. E., BANOVICH, N. E., RAJ, A., LEWELLEN, N., MYRTHIL, M., GILAD, Y. & PRITCHARD, J. K. 2013. Identification of Genetic Variants That Affect Histone Modifications in Human Cells. *Science,* 342**,** 747.

MEANS JR, R. T., ALLEN, J., SEARS, D. A. & SCHUSTER, S. J. 1999. Serum soluble transferrin receptor and the prediction of marrow aspirate iron results in a heterogeneous group of patients. *Clinical and Laboratory Haematology,* 21**,** 161-168.

MEHMOOD, M., AGARWAL, R., RAINA, A., CORREA-JAQUE, P. & BENZA, R. L. 2016. Hemodynamic response to treatment of iron deficiency anemia in pulmonary arterial hypertension: longitudinal insights from an implantable hemodynamic monitor. *Pulm Circ,* 6**,** 616-618.

MENDEL, G. A. 1961. Studies on iron absorption. I. The relationships between the rate of erythropoiesis, hypoxia and iron absorption.

MERLA, G., BRUNETTI-PIERRI, N., PICCOLO, P., MICALE, L. & LOVIGLIO MARIA, N. 2012. Supravalvular Aortic Stenosis. *Circulation: Cardiovascular Genetics,* 5**,** 692-696.

MICHAELSON, J. J., LOGUERCIO, S. & BEYER, A. 2009. Detection and interpretation of expression quantitative trait loci (eQTL). *Methods,* 48**,** 265-276.

MIYAMOTO, S., NAGAYA, N., SATOH, T., KYOTANI, S., SAKAMAKI, F., FUJITA, M., NAKANISHI, N. & MIYATAKE, K. 2000. Clinical Correlates and Prognostic Significance of Six-minute Walk Test in Patients with Primary Pulmonary Hypertension. *American Journal of Respiratory and Critical Care Medicine,* 161**,** 487-492.

MORABIA, A. 1997. Case–Control Studies in Clinical Research: Mechanism and Prevention of Selection Bias. *Preventive Medicine,* 26**,** 674-677.

NASIM, M. T., OGO, T., AHMED, M., RANDALL, R., CHOWDHURY, H. M., SNAPE, K. M., BRADSHAW, T. Y., SOUTHGATE, L., LEE, G. J., JACKSON, I., LORD, G. M., GIBBS, J. S. R., WILKINS, M. R., OHTA-OGO, K., NAKAMURA, K., GIRERD, B., COULET, F., SOUBRIER, F., HUMBERT, M., MORRELL, N. W., TREMBATH, R. C. & MACHADO, R. D. 2011. Molecular genetic characterization of SMAD signaling molecules in pulmonary arterial hypertension. *Human Mutation,* 32**,** 1385-1389.

NATH, A. P., RITCHIE, S. C., BYARS, S. G., FEARNLEY, L. G., HAVULINNA, A. S., JOENSUU, A., KANGAS, A. J., SOININEN, P., WENNERSTRÖM, A., MILANI, L., METSPALU, A., MÄNNISTÖ, S., WÜRTZ, P., KETTUNEN, J., RAITOHARJU, E., KÄHÖNEN, M., JUONALA, M., PALOTIE, A., ALA-KORPELA, M., RIPATTI, S., LEHTIMÄKI, T., ABRAHAM, G., RAITAKARI, O., SALOMAA, V., PEROLA, M. & INOUYE, M. 2017. An interaction map of circulating metabolites, immune gene networks, and their genetic regulation. *Genome Biology,* 18**,** 146.

NIELSEN, O. J., ANDERSEN, L. S., HANSEN, N. E. & HANSEN, T. M. 1994. Serum transferrin receptor levels in anaemic patients with rheumatoid arthritis. *Scandinavian Journal of Clinical and Laboratory Investigation,* 54**,** 75-82.

OKONKO, D. O., GRZESLO, A., WITKOWSKI, T., MANDAL, A. K. J., SLATER, R. M., ROUGHTON, M., FOLDES, G., THUM, T., MAJDA, J., BANASIAK, W., MISSOURIS, C. G., POOLE-WILSON, P. A., ANKER, S. D. & PONIKOWSKI, P. 2008. Effect of Intravenous Iron Sucrose on Exercise Tolerance in Anemic and Nonanemic Patients With Symptomatic Chronic Heart Failure and Iron Deficiency: FERRIC-HF: A Randomized, Controlled, Observer-Blinded Trial. *Journal of the American College of Cardiology,* 51**,** 103-112.

OLSCHEWSKI, H., SIMONNEAU, G., GALIÈ, N., HIGENBOTTAM, T., NAEIJE, R., RUBIN, L. J., NIKKHO, S., SPEICH, R., HOEPER, M. M., BEHR, J., WINKLER, J., SITBON, O., POPOV, W., GHOFRANI, H. A., MANES, A., KIELY, D. G., EWERT, R., MEYER, A., CORRIS, P. A., DELCROIX, M., GOMEZ-SANCHEZ, M., SIEDENTOP, H. & SEEGER, W. 2002. Inhaled Iloprost for Severe Pulmonary Hypertension. *New England Journal of Medicine,* 347**,** 322-329.

OUDIZ, R. J. 2007. Pulmonary Hypertension Associated with Left-Sided Heart Disease. *Clinics in Chest Medicine,* 28**,** 233-241.

PAK, M., LOPEZ, M. A., GABAYAN, V., GANZ, T. & RIVERA, S. 2006. Suppression of hepcidin during anemia requires erythropoietic activity. *Blood,* 108**,** 3730-3735.

PALMER, T. M., NORDESTGAARD, B. G., BENN, M., TYBJÆRG-HANSEN, A., DAVEY SMITH, G., LAWLOR, D. A. & TIMPSON, N. J. 2013. Association of plasma uric acid with ischaemic heart disease and blood pressure: mendelian randomisation analysis of two large cohorts. *BMJ (Clinical research ed.),* 347**,** f4262-f4262.

PARKER, J. S., MULLINS, M., CHEANG, M. C. U., LEUNG, S., VODUC, D., VICKERY, T., DAVIES, S., FAURON, C., HE, X., HU, Z., QUACKENBUSH, J. F., STIJLEMAN, I. J., PALAZZO, J., MARRON, J. S., NOBEL, A. B., MARDIS, E., NIELSEN, T. O., ELLIS, M. J., PEROU, C. M. & BERNARD, P. S. 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* 27**,** 1160-1167.

PATRO, R., DUGGAL, G., LOVE, M. I., IRIZARRY, R. A. & KINGSFORD, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods,* 14**,** 417-419.

PEYSSONNAUX, C., NIZET, V. & JOHNSON, R. S. 2008. Role of the hypoxia inducible factors HIF in iron metabolism. *Cell Cycle,* 7**,** 28-32.

PICHLER, I., DEL GRECO M, F., GÖGELE, M., LILL, C. M., BERTRAM, L., DO, C. B., ERIKSSON, N., FOROUD, T., MYERS, R. H., CONSORTIUM, P. G., NALLS, M., KELLER, M. F., INTERNATIONAL PARKINSON'S DISEASE GENOMICS, C., WELLCOME TRUST CASE CONTROL, C., BENYAMIN, B., WHITFIELD, J. B., GENETICS OF IRON STATUS, C., PRAMSTALLER, P. P., HICKS, A. A., THOMPSON, J. R. & MINELLI, C. 2013. Serum Iron Levels and the Risk of Parkinson Disease: A Mendelian Randomization Study. *PLOS Medicine,* 10**,** e1001462.

PIERCE, B. L., AHSAN, H. & VANDERWEELE, T. J. 2011. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *International Journal of Epidemiology,* 40**,** 740-752.

PLESNER, L. L., SCHOOS, M. M., DALSGAARD, M., GOETZE, J. P., KJØLLER, E., VESTBO, J. & IVERSEN, K. 2017. Iron Deficiency in COPD Associates with Increased Pulmonary Artery Pressure Estimated by Echocardiography. *Heart, Lung and Circulation,* 26**,** 101-104.

POMS, A. D., TURNER, M., FARBER, H. W., MELTZER, L. A. & MCGOON, M. D. 2013. Comorbid Conditions and Outcomes in Patients With Pulmonary Arterial Hypertension: A REVEAL Registry Analysis. *Chest,* 144**,** 169-176.

POWELL, J. E., HENDERS, A. K., MCRAE, A. F., WRIGHT, M. J., MARTIN, N. G., DERMITZAKIS, E. T., MONTGOMERY, G. W. & VISSCHER, P. M. 2012. Genetic control of gene expression in whole blood and lymphoblastoid cell lines is largely independent. *Genome Research,* 22**,** 456-466.

PRENTICE, R. L., CAAN, B., CHLEBOWSKI, R. T., PATTERSON, R., KULLER, L. H., OCKENE, J. K., MARGOLIS, K. L., LIMACHER, M. C., MANSON, J. E., PARKER, L. M., PASKETT, E., PHILLIPS, L., ROBBINS, J., ROSSOUW, J. E., SARTO, G. E., SHIKANY, J. M., STEFANICK, M. L., THOMSON, C. A., VAN HORN, L., VITOLINS, M. Z., WACTAWSKI-WENDE, J., WALLACE, R. B., WASSERTHEIL-SMOLLER, S., WHITLOCK, E., YANO, K., ADAMS-CAMPBELL, L., ANDERSON, G. L., ASSAF, A. R., BERESFORD, S. A. A., BLACK, H. R., BRUNNER, R. L., BRZYSKI, R. G., FORD, L., GASS, M., HAYS, J., HEBER, D., HEISS, G., HENDRIX, S. L., HSIA, J., HUBBELL, F. A., JACKSON, R. D., JOHNSON, K. C., KOTCHEN, J. M., LACROIX, A. Z., LANE, D. S., LANGER, R. D., LASSER, N. L. & HENDERSON, M. M. 2006. Low-Fat Dietary Pattern and Risk of Invasive Breast CancerThe Women's Health Initiative Randomized Controlled Dietary Modification Trial. *JAMA,* 295**,** 629-642.

PUNNONEN, K., IRJALA K FAU - RAJAMAKI, A. & RAJAMAKI, A. 1994. Iron-deficiency anemia is associated with high concentrations of transferrin receptor in serum. *Clin Chem*.

PUTT, M. & CHINCHILLI, V. M. 1999. A mixed effects model for the analysis of repeated measures cross-over studies. *Stat Med*.

R'ZIK, S. & BEGUIN, Y. 2001. Serum soluble transferrin receptor concentration is an accurate estimate of the mass of tissue receptors. *Experimental Hematology,* 29**,** 677-685.

RABINOVITCH, M., GUIGNABERT, C., HUMBERT, M. & NICOLLS MARK, R. 2014. Inflammation and Immunity in the Pathogenesis of Pulmonary Arterial Hypertension. *Circulation Research,* 115**,** 165-175.

RAMAKRISHNAN, L., PEDERSEN, S. L., TOE, Q. K., WEST, L. E., MUMBY, S., CASBOLT, H., ISSITT, T., GARFIELD, B., LAWRIE, A., WORT, S. J. & QUINLAN, G. J. 2018. The Hepcidin/Ferroportin axis modulates proliferation of pulmonary artery smooth muscle cells. *Scientific Reports,* 8**,** 12972.

REDDON, H., GUÉANT, J.-L. & MEYRE, D. 2016. The importance of gene-environment interactions in human obesity. *Clinical science (London, England : 1979),* 130**,** 1571-1597.

REMACHA, A. F., SARDA, M. P., PARELLADA, M., UBEDA, J. & MANTEIGA, R. 1998. The role of serum transferrin receptor in the diagnosis of iron deficiency. *Haematologica,* 83**,** 963.

RHODES, C. J., BATAI, K., BLEDA, M., HAIMEL, M., SOUTHGATE, L., GERMAIN, M., PAUCIULO, M. W., HADINNAPOLA, C., AMAN, J., GIRERD, B., ARORA, A., KNIGHT, J., HANSCOMBE, K. B., KARNES, J. H., KAAKINEN, M., GALL, H., ULRICH, A., HARBAUM, L., CEBOLA, I., FERRER, J., LUTZ, K., SWIETLIK, E. M., AHMAD, F., AMOUYEL, P., ARCHER, S. L., ARGULA, R., AUSTIN, E. D., BADESCH, D., BAKSHI, S., BARNETT, C., BENZA, R., BHATT, N., BOGAARD, H. J., BURGER, C. D., CHAKINALA, M., CHURCH, C., COGHLAN, J. G., CONDLIFFE, R., CORRIS, P. A., DANESINO, C., DEBETTE, S., ELLIOTT, C. G., ELWING, J., EYRIES, M., FORTIN, T., FRANKE, A., FRANTZ, R. P., FROST, A., GARCIA, J. G. N., GHIO, S., GHOFRANI, H.-A., GIBBS, J. S. R., HARLEY, J., HE, H., HILL, N. S., HIRSCH, R., HOUWELING, A. C., HOWARD, L. S., IVY, D., KIELY, D. G., KLINGER, J., KOVACS, G., LAHM, T., LAUDES, M., MACHADO, R. D., ROSS, R. V. M., MARSOLO, K., MARTIN, L. J., MOLEDINA, S., MONTANI, D., NATHAN, S. D., NEWNHAM, M., OLSCHEWSKI, A., OLSCHEWSKI, H., OUDIZ, R. J., OUWEHAND, W. H., PEACOCK, A. J., PEPKE-ZABA, J., REHMAN, Z., ROBBINS, I., RODEN, D. M., ROSENZWEIG, E. B., SAYDAIN, G., SCELSI, L., SCHILZ, R., SEEGER, W., SHAFFER, C. M., SIMMS, R. W., SIMON, M., SITBON, O., SUNTHARALINGAM, J., TANG, H., TCHOURBANOV, A. Y., THENAPPAN, T., TORRES, F., TOSHNER, M. R., TREACY, C. M., NOORDEGRAAF, A. V., WAISFISZ, Q., WALSWORTH, A. K., et al. 2018. Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *The Lancet Respiratory Medicine*.

RHODES, C. J., HOWARD, L. S., BUSBRIDGE, M., ASHBY, D., KONDILI, E., GIBBS, J. S., WHARTON, J. & WILKINS, M. R. 2011a. Iron deficiency and raised hepcidin in idiopathic pulmonary arterial hypertension: clinical prevalence, outcomes, and mechanistic insights. *J Am Coll Cardiol,* 58**,** 300-9.

RHODES, C. J., WHARTON, J., HOWARD, L. S., GIBBS, J. S. & WILKINS, M. R. 2011b. Red cell distribution width outperforms other potential circulating biomarkers in predicting survival in idiopathic pulmonary arterial hypertension. *Heart,* 97**,** 1054-60.

RODEN, D. M., PULLEY, J. M., BASFORD, M. A., BERNARD, G. R., CLAYTON, E. W., BALSER, J. R. & MASYS, D. R. 2008. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clinical pharmacology and therapeutics,* 84**,** 362-369.

ROTHMAN, K. J. 1976. CAUSES. *American Journal of Epidemiology,* 104**,** 587-592.

ROTHMAN, K. J. & GREENLAND, S. 2005. Causation and causal inference in epidemiology. *Am J Public Health,* 95 Suppl 1**,** S144-50.

RUBIN, L. J., BADESCH, D. B., BARST, R. J., GALIÈ, N., BLACK, C. M., KEOGH, A., PULIDO, T., FROST, A., ROUX, S., LECONTE, I., LANDZBERG, M. & SIMONNEAU, G. 2002. Bosentan therapy for pulmonary arterial hypertension. *New England Journal of Medicine,* 346**,** 896-903.

RUITER, G., LANKHORST, S., BOONSTRA, A., POSTMUS, P. E., ZWEEGMAN, S., WESTERHOF, N., VAN DER LAARSE, W. J. & VONK-NOORDEGRAAF, A. 2011. Iron deficiency is common in idiopathic pulmonary arterial hypertension. *Eur Respir J,* 37**,** 1386-91.

RUITER, G., MANDERS, E., HAPPÉ, C. M., SCHALIJ, I., GROEPENHOFF, H., HOWARD, L. S., WILKINS, M. R., BOGAARD, H. J., WESTERHOF, N., VAN DER LAARSE, W. J., DE MAN, F. S. & VONK-NOORDEGRAAF, A. 2015. Intravenous iron therapy in patients with idiopathic pulmonary arterial hypertension and iron deficiency. *Pulmonary Circulation,* 5**,** 466-472.

SANDERSON, E., DAVEY SMITH, G., WINDMEIJER, F. & BOWDEN, J. 2018. An examination of Multivariable Mendelian Randomization in the single sample and two-sample summary data settings.

SCHERMULY, R. T., GHOFRANI, H. A., WILKINS, M. R. & GRIMMINGER, F. 2011. Mechanisms of disease: pulmonary arterial hypertension. *Nat Rev Cardiol,* 8**,** 443-55.

SCOTT, R. A., LAGOU, V., WELCH, R. P., WHEELER, E., MONTASSER, M. E., LUAN, J. A., MÄGI, R., STRAWBRIDGE, R. J., REHNBERG, E., GUSTAFSSON, S., KANONI, S., RASMUSSEN-TORVIK, L. J., YENGO, L., LECOEUR, C., SHUNGIN, D., SANNA, S., SIDORE, C., JOHNSON, P. C. D., JUKEMA, J. W., JOHNSON, T., MAHAJAN, A., VERWEIJ, N., THORLEIFSSON, G., HOTTENGA, J.-J., SHAH, S., SMITH, A. V., SENNBLAD, B., GIEGER, C., SALO, P., PEROLA, M., TIMPSON, N. J., EVANS, D. M., POURCAIN, B. S., WU, Y., ANDREWS, J. S., HUI, J., BIELAK, L. F., ZHAO, W., HORIKOSHI, M., NAVARRO, P., ISAACS, A., O'CONNELL, J. R., STIRRUPS, K., VITART, V., HAYWARD, C., ESKO, T., MIHAILOV, E., FRASER, R. M., FALL, T., VOIGHT, B. F., RAYCHAUDHURI, S., CHEN, H., LINDGREN, C. M., MORRIS, A. P., RAYNER, N. W., ROBERTSON, N., RYBIN, D., LIU, C.-T., BECKMANN, J. S., WILLEMS, S. M., CHINES, P. S., JACKSON, A. U., KANG, H. M., STRINGHAM, H. M., SONG, K., TANAKA, T., PEDEN, J. F., GOEL, A., HICKS, A. A., AN, P., MÜLLER-NURASYID, M., FRANCO-CERECEDA, A., FOLKERSEN, L., MARULLO, L., JANSEN, H., OLDEHINKEL, A. J., BRUINENBERG, M., PANKOW, J. S., NORTH, K. E., FOROUHI, N. G., LOOS, R. J. F., EDKINS, S., VARGA, T. V., HALLMANS, G., OKSA, H., ANTONELLA, M., NAGARAJA, R., TROMPET, S., FORD, I., BAKKER, S. J. L., KONG, A., KUMARI, M., GIGANTE, B., HERDER, C., MUNROE, P. B., CAULFIELD, M., ANTTI, J., MANGINO, M., SMALL, K., MILJKOVIC, I., et al. 2012. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics,* 44**,** 991-1005.

SHIH, Y. J., BAYNES, R. D., HUDSON, B. G., FLOWERS, C. H., SKIKNE, B. S. & COOK, J. D. 1990. Serum transferrin receptor is a truncated form of tissue receptor. *Journal of Biological Chemistry,* 265**,** 19077-19081.

SIMONNEAU, G., GATZOULIS, M. A., ADATIA, I., CELERMAJER, D., DENTON, C., GHOFRANI, A., GOMEZ SANCHEZ, M. A., KRISHNA KUMAR, R., LANDZBERG, M., MACHADO, R. F., OLSCHEWSKI, H., ROBBINS, I. M. & SOUZA, R. 2013. Updated Clinical Classification of Pulmonary Hypertension. *Journal of the American College of Cardiology,* 62**,** D34-D41.

SIMONNEAU, G., MONTANI, D., CELERMAJER, D. S., DENTON, C. P., GATZOULIS, M. A., KROWKA, M., WILLIAMS, P. G. & SOUZA, R. 2019. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *European Respiratory Journal,* 53**,** 1801913.

SIMPSON, R. J. 1992. Effect of hypoxic exposure on iron absorption in heterozygous hypotransferrinaemic mice.

SINGH, A. J., RAMSEY, S. A., FILTZ, T. M. & KIOUSSI, C. 2018. Differential gene regulatory networks in development and disease. *Cellular and Molecular Life Sciences,* 75**,** 1013-1025.

SITBON, O., CHANNICK, R., CHIN, K. M., FREY, A., GAINE, S., GALIÈ, N., GHOFRANI, H.-A., HOEPER, M. M., LANG, I. M., PREISS, R., RUBIN, L. J., DI SCALA, L., TAPSON, V., ADZERIKHO, I., LIU, J., MOISEEVA, O., ZENG, X., SIMONNEAU, G. & MCLAUGHLIN, V. V. 2015. Selexipag for the Treatment of Pulmonary Arterial Hypertension. *New England Journal of Medicine,* 373**,** 2522-2533.

SMITH, G. D. & EBRAHIM, S. 2002. Data dredging, bias or confounding. *BMJ,* 325**,** 21-28.

SMITH, T. G., BALANOS, G. M., CROFT, Q. P. P., TALBOT, N. P., DORRINGTON, K. L., RATCLIFFE, P. J. & ROBBINS, P. A. 2008. The increase in pulmonary arterial pressure caused by hypoxia depends on iron status. *The Journal of Physiology,* 586**,** 5999-6005.

SMITH, T. G., TALBOT, N. P., PRIVAT, C. & ET AL. 2009. Effects of iron supplementation and depletion on hypoxic pulmonary hypertension: Two randomized controlled trials. *JAMA,* 302**,** 1444-1450.

SONESON, C., LOVE, M. & ROBINSON, M. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*.

SONNWEBER, T., NAIRZ, M., THEURL, I., PETZER, V., TYMOSZUK, P., HASCHKA, D., RIEGER, E., KAESSMANN, B., DERI, M., WATZINGER, K., STERINGER-MASCHERBAUER, R., TANCEVSKI, I., WEISS, G. & LÖFFLER-RAGG, J. 2018. The crucial impact of iron deficiency definition for the course of precapillary pulmonary hypertension. *PLOS ONE,* 13**,** e0203396.

SOON, E., TREACY, C. M., TOSHNER, M. R., MACKENZIE-ROSS, R., MANGLAM, V., BUSBRIDGE, M., SINCLAIR-MCGARVIE, M., ARNOLD, J., SHEARES, K. K., MORRELL, N. W. & PEPKE-ZABA, J. 2011. Unexplained iron deficiency in idiopathic and heritable pulmonary arterial hypertension. *Thorax,* 66**,** 326-332.

SOUTHGATE, L., MACHADO, R. D., GRÄF, S. & MORRELL, N. W. 2020. Molecular genetic framework underlying pulmonary arterial hypertension. *Nature Reviews Cardiology,* 17**,** 85-95.

STACHER, E., GRAHAM, B. B., HUNT, J. M., GANDJEVA, A., GROSHONG, S. D., MCLAUGHLIN, V. V., JESSUP, M., GRIZZLE, W. E., ALDRED, M. A., COOL, C. D. & TUDER, R. M. 2012. Modern age pathology of pulmonary arterial hypertension. *American journal of respiratory and critical care medicine,* 186**,** 261-272.

STEARNS, F. W. 2010. One Hundred Years of Pleiotropy: A Retrospective. *Genetics,* 186**,** 767.

STEINER, M. K., SYRKINA OLGA, L., KOLLIPUTI, N., MARK EUGENE, J., HALES CHARLES, A. & WAXMAN AARON, B. 2009. Interleukin-6 Overexpression Induces Pulmonary Hypertension. *Circulation Research,* 104**,** 236-244.

SUN, X.-G., HANSEN JAMES, E., OUDIZ RONALD, J. & WASSERMAN, K. 2001. Exercise Pathophysiology in Patients With Primary Pulmonary Hypertension. *Circulation,* 104**,** 429-435.

SUOMINEN, P., MÖTTÖNEN, T., RAJAMÄKI, A. & IRJALA, K. 2000. Single values of serum transferrin receptor and transferrin receptor ferritin index can be used to detect true and functional iron deficiency in rheumatoid arthritis patients with anemia. *Arthritis & Rheumatism,* 43**,** 1016-1020.

TAM, V., PATEL, N., TURCOTTE, M., BOSSÉ, Y., PARÉ, G. & MEYRE, D. 2019. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics,* 20**,** 467-484.

TAMOSIUNIENE, R., TIAN, W., DHILLON, G., WANG, L., SUNG YON, K., GERA, L., PATTERSON ANDREW, J., AGRAWAL, R., RABINOVITCH, M., AMBLER, K., LONG CARLIN, S., VOELKEL NORBERT, F. & NICOLLS MARK, R. 2011. Regulatory T Cells Limit Vascular Endothelial Injury and Prevent Pulmonary Hypertension. *Circulation Research,* 109**,** 867-879.

THAYER, T. E., HUANG, S., LEVINSON, R. T., FARBER-EGER, E., ASSAD, T. R., HUSTON, J. H., MOSLEY, J. D., WELLS, Q. S. & BRITTAIN, E. L. 2019. Unbiased Phenome-wide Association Studies of Red Cell Distribution Width Identifies Key Associations with Pulmonary Hypertension. *Annals of the American Thoracic Society*.

THE HAPLOTYPE REFERENCE, C., MCCARTHY, S., DAS, S., KRETZSCHMAR, W., DELANEAU, O., WOOD, A. R., TEUMER, A., KANG, H. M., FUCHSBERGER, C., DANECEK, P., SHARP, K., LUO, Y., SIDORE, C., KWONG, A., TIMPSON, N., KOSKINEN, S., VRIEZE, S., SCOTT, L. J., ZHANG, H., MAHAJAN, A., VELDINK, J., PETERS, U., PATO, C., VAN DUIJN, C. M., GILLIES, C. E., GANDIN, I., MEZZAVILLA, M., GILLY, A., COCCA, M., TRAGLIA, M., ANGIUS, A., BARRETT, J. C., BOOMSMA, D., BRANHAM, K., BREEN, G., BRUMMETT, C. M., BUSONERO, F., CAMPBELL, H., CHAN, A., CHEN, S., CHEW, E., COLLINS, F. S., CORBIN, L. J., SMITH, G. D., DEDOUSSIS, G., DORR, M., FARMAKI, A.-E., FERRUCCI, L., FORER, L., FRASER, R. M., GABRIEL, S., LEVY, S., GROOP, L., HARRISON, T., HATTERSLEY, A., HOLMEN, O. L., HVEEM, K., KRETZLER, M., LEE, J. C., MCGUE, M., MEITINGER, T., MELZER, D., MIN, J. L., MOHLKE, K. L., VINCENT, J. B., NAUCK, M., NICKERSON, D., PALOTIE, A., PATO, M., PIRASTU, N., MCINNIS, M., RICHARDS, J. B., SALA, C., SALOMAA, V., SCHLESSINGER, D., SCHOENHERR, S., SLAGBOOM, P. E., SMALL, K., SPECTOR, T., STAMBOLIAN, D., TUKE, M., TUOMILEHTO, J., VAN DEN BERG, L. H., VAN RHEENEN, W., VOLKER, U., WIJMENGA, C., TONIOLO, D., ZEGGINI, E., GASPARINI, P., SAMPSON, M. G., WILSON, J. F., FRAYLING, T., DE BAKKER, P. I. W., SWERTZ, M. A., MCCARROLL, S., KOOPERBERG, C., DEKKER, A., ALTSHULER, D., WILLER, C., IACONO, W., et al. 2016. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics,* 48**,** 1279.

THOMSON, C. A., VAN HORN, L., CAAN, B. J., ARAGAKI, A. K., CHLEBOWSKI, R. T., MANSON, J. E., ROHAN, T. E., TINKER, L. F., KULLER, L. H., HOU, L., LANE, D. S., JOHNSON, K. C., VITOLINS, M. Z. & PRENTICE, R. L.

2014. Cancer Incidence and Mortality during the Intervention and Postintervention Periods of the Women's Health Initiative Dietary Modification Trial. *Cancer Epidemiology Biomarkers & Prevention,* 23**,** 2924.

TIAN, W., JIANG, X., SUNG YON, K., SHUFFLE, E., WU, T.-H., KAO PETER, N., TU ALLEN, B., DORFMÜLLER, P., CAO, A., WANG, L., PENG, G., KIM, Y., ZHANG, P., CHAPPELL, J., PASUPNETI, S., DAHMS, P., MAGUIRE, P., CHAIB, H., ZAMANIAN, R., PETERS-GOLDEN, M., SNYDER MICHAEL, P., VOELKEL NORBERT, F., HUMBERT, M., RABINOVITCH, M. & NICOLLS MARK, R. 2019. Phenotypically Silent Bone Morphogenetic Protein Receptor 2 Mutations Predispose Rats to Inflammation-Induced Pulmonary Arterial Hypertension by Enhancing the Risk for Neointimal Transformation. *Circulation,* 140**,** 1409-1425.

TOBLLI, J. E., LOMBRAÑA, A., DUARTE, P. & DI GENNARO, F. 2007. Intravenous Iron Reduces NT-Pro-Brain Natriuretic Peptide in Anemic Patients With Chronic Heart Failure and Renal Insufficiency. *Journal of the American College of Cardiology,* 50**,** 1657-1665.

TUDER, R. M., GROVES, B., BADESCH, D. B. & VOELKEL, N. F. 1994. Exuberant endothelial cell growth and elements of inflammation are present in plexiform lesions of pulmonary hypertension. *The American journal of pathology,* 144**,** 275-285.

TYNDALL, A. J., BANNERT, B., VONK, M., AIRÒ, P., COZZI, F., CARREIRA, P. E., BANCEL, D. F., ALLANORE, Y., MÜLLER-LADNER, U., DISTLER, O., IANNONE, F., PELLERITO, R., PILECKYTE, M., MINIATI, I., ANANIEVA, L., GURMAN, A. B., DAMJANOV, N., MUELLER, A., VALENTINI, G., RIEMEKASTEN, G., TIKLY, M., HUMMERS, L., HENRIQUES, M. J. S., CARAMASCHI, P., SCHEJA, A., ROZMAN, B., TON, E., KUMÁNOVICS, G., COLEIRO, B., FEIERL, E., SZUCS, G., VON MÜHLEN, C. A., RICCIERI, V., NOVAK, S., CHIZZOLINI, C., KOTULSKA, A., DENTON, C., COELHO, P. C., KÖTTER, I., SIMSEK, I., DE LA PENA LEFEBVRE, P. G., HACHULLA, E., SEIBOLD, J. R., REDNIC, S., ŠTORK, J., MOROVIC-VERGLES, J. & WALKER, U. A. 2010. Causes and risk factors for death in systemic sclerosis: a study from the EULAR Scleroderma Trials and Research (EUSTAR) database. *Annals of the Rheumatic Diseases,* 69**,** 1809.

ULRICH, A., WHARTON, J., THAYER, T. E., SWIETLIK, E. M., ASSAD, T. R., DESAI, A. A., GRÄF, S., HARBAUM, L., HUMBERT, M., MORRELL, N. W., NICHOLS, W. C., SOUBRIER, F., SOUTHGATE, L., TRÉGOUËT, D.-A., TREMBATH, R. C., BRITTAIN, E. L., WILKINS, M. R., PROKOPENKO, I. & RHODES, C. J. 2019. Mendelian randomisation analysis of red cell distribution width in pulmonary arterial hypertension. *European Respiratory Journal***,** 1901486.

VAN DE BRUAENE, A., DELCROIX, M., PASQUET, A., DE BACKER, J., DE PAUW, M., NAEIJE, R., VACHIÉRY, J.-L., PAELINCK, B., MORISSENS, M. & BUDTS, W. 2011. Iron deficiency is associated with adverse outcome in Eisenmenger patients. *European Heart Journal,* 32**,** 2790-2799.

VAN HORN, L., DOLECEK, T. A., GRANDITS, G. A. & SKWERES, L. 1997. Adherence to dietary recommendations in the special intervention group in the Multiple Risk Factor Intervention Trial. *The American Journal of Clinical Nutrition,* 65**,** 289S-304S.

VANDERWEELE, T. J., TCHETGEN TCHETGEN, E. J., CORNELIS, M. & KRAFT, P. 2014. Methodological challenges in mendelian randomization. *Epidemiology (Cambridge, Mass.),* 25**,** 427-435.

VERBANCK, M., CHEN, C.-Y., NEALE, B. & DO, R. 2018. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics,* 50**,** 693-698.

VIETHEN, T., GERHARDT, F., DUMITRESCU, D., KNOOP-BUSCH, S., TEN FREYHAUS, H., RUDOLPH, T. K., BALDUS, S. & ROSENKRANZ, S. 2014. Ferric carboxymaltose improves exercise capacity and quality of life in patients with pulmonary arterial hypertension and iron deficiency: A pilot study. *International Journal of Cardiology,* 175**,** 233-239.

VISSCHER, P. M., BROWN, M. A., MCCARTHY, M. I. & YANG, J. 2012. Five years of GWAS discovery. *Am J Hum Genet,* 90**,** 7-24.

VISSCHER, P. M., WRAY, N. R., ZHANG, Q., SKLAR, P., MCCARTHY, M. I., BROWN, M. A. & YANG, J. 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics,* 101**,** 5-22.

VOELKEL, N. F. & GOMEZ-ARROYO, J. 2014. The Role of Vascular Endothelial Growth Factor in Pulmonary Arterial Hypertension. The Angiogenesis Paradox. *American Journal of Respiratory Cell and Molecular Biology,* 51**,** 474-484.

WANG, J., VASAIKAR, S., SHI, Z., GREER, M. & ZHANG, B. 2017. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Research,* 45**,** W130-W137.

WANG, K., LI, M. & HAKONARSON, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research,* 38**,** e164-e164.

WANG, X.-J., LIAN, T.-Y., JIANG, X., LIU, S.-F., LI, S.-Q., JIANG, R., WU, W.-H., YE, J., CHENG, C.-Y., DU, Y., XU, X.-Q., WU, Y., PENG, F.-H., SUN, K., MAO, Y.-M., YU, H., LIANG, C., SHYY, JOHN Y.-J., ZHANG, S.-Y., ZHANG, X. & JING, Z.-C. 2019. Germline BMP9 mutation causes idiopathic pulmonary arterial hypertension. *European Respiratory Journal,* 53**,** 1801609.

WEISS, G., WERNER-FELMAYER, G., WERNER, E. R., GRÜNEWALD, K., WACHTER, H. & HENTZE, M. W. 1994. Iron regulates nitric oxide synthase activity by controlling nuclear transcription. *The Journal of experimental medicine,* 180**,** 969-976.

WESTRA, H. J., PETERS, M. J., ESKO, T., YAGHOOTKAR, H., SCHURMANN, C., KETTUNEN, J., CHRISTIANSEN, M. W., FAIRFAX, B. P., SCHRAMM, K., POWELL, J. E., ZHERNAKOVA, A., ZHERNAKOVA, D. V., VELDINK, J. H., VAN DEN BERG, L. H., KARJALAINEN, J., WITHOFF, S., UITTERLINDEN, A. G., HOFMAN, A., RIVADENEIRA, F., HOEN, P. A. C., REINMAA, E., FISCHER, K., NELIS, M., MILANI, L., MELZER, D., FERRUCCI, L., SINGLETON, A. B., HERNANDEZ, D. G., NALLS, M. A., HOMUTH, G., NAUCK, M., RADKE, D., VOLKER, U., PEROLA, M., SALOMAA, V., BRODY, J., SUCHY-DICEY, A., GHARIB, S. A., ENQUOBAHRIE, D. A., LUMLEY, T., MONTGOMERY, G. W., MAKINO, S., PROKISCH, H., HERDER, C., RODEN, M., GRALLERT, H., MEITINGER, T., STRAUCH, K., LI, Y., JANSEN, R. C., VISSCHER, P. M., KNIGHT, J. C., PSATY, B. M., RIPATTI, S., TEUMER, A., FRAYLING, T. M., METSPALU, A., VAN MEURS, J. B. J. & FRANKE, L. 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet,* 45**,** 1238-1243.

WILKEN, E., BENNJI, S., SYMONS, G., WILLIAMNS, P. G. & ALLWOOD, B. 2019. Treatment of pulmonary arterial hypertension: A review of drugs available for advanced therapy. *AJTCCM,* 25.

WU, D. & YOTNDA, P. 2011. Induction and testing of hypoxia in cell culture. *Journal of visualized experiments : JoVE***,** 2899.

XU, C., TACHMAZIDOU, I., WALTER, K., CIAMPI, A., ZEGGINI, E., GREENWOOD, C. M. T. & THE, U. K. K. C. 2014. Estimating Genome-Wide Significance for Whole-Genome Sequencing Studies. *Genetic Epidemiology,* 38**,** 281-290.

YU, P. B., BEPPU H FAU - KAWAI, N., KAWAI N FAU - LI, E., LI E FAU - BLOCH, K. D. & BLOCH, K. D. 2005. Bone morphogenetic protein (BMP) type II receptor deletion reveals BMP ligand-specific gain of signaling in pulmonary artery smooth muscle cells.

ZHENG, J., BAIRD, D., BORGES, M.-C., BOWDEN, J., HEMANI, G., HAYCOCK, P., EVANS, D. M. & SMITH, G. D. 2017. Recent Developments in Mendelian Randomization Studies. *Current Epidemiology Reports,* 4**,** 330-345.

ZHERNAKOVA, D. V., DEELEN, P., VERMAAT, M., VAN ITERSON, M., VAN GALEN, M., ARINDRARTO, W., VAN 'T HOF, P., MEI, H., VAN DIJK, F., WESTRA, H. J., BONDER, M. J., VAN ROOIJ, J., VERKERK, M., JHAMAI, P. M., MOED, M., KIELBASA, S. M., BOT, J., NOOREN, I., POOL, R., VAN DONGEN, J., HOTTENGA, J. J., STEHOUWER, C. D., VAN DER KALLEN, C. J., SCHALKWIJK, C. G., ZHERNAKOVA, A., LI, Y., TIGCHELAAR, E. F., DE KLEIN, N., BEEKMAN, M., DEELEN, J., VAN HEEMST, D., VAN DEN BERG, L. H., HOFMAN, A., UITTERLINDEN, A. G., VAN GREEVENBROEK, M. M., VELDINK, J. H., BOOMSMA, D. I., VAN DUIJN, C. M., WIJMENGA, C., SLAGBOOM, P. E., SWERTZ, M. A., ISAACS, A., VAN MEURS, J. B., JANSEN, R., HEIJMANS, B. T., T HOEN, P. A. & FRANKE, L. 2017. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet,* 49**,** 139-145.

ZHU, L., DING, Y., CHEN, C.-Y., WANG, L., HUO, Z., KIM, S., SOTIRIOU, C., OESTERREICH, S. & TSENG, G. C. 2016. MetaDCN: meta-analysis framework for differential co-expression network detection with an application in breast cancer. *Bioinformatics,* 33**,** 1121-1129.

ZHU, N., GONZAGA-JAUREGUI, C., WELCH CARRIE, L., MA, L., QI, H., KING ALEJANDRA, K., KRISHNAN, U., ROSENZWEIG ERIKA, B., IVY, D. D., AUSTIN ERIC, D., HAMID, R., NICHOLS WILLIAM, C., PAUCIULO MICHAEL, W., LUTZ KATIE, A., SAWLE, A., REID JEFFREY, G., OVERTON JOHN, D., BARAS, A., DEWEY, F., SHEN, Y. & CHUNG WENDY, K. 2018. Exome Sequencing in Children With Pulmonary Arterial Hypertension Demonstrates Differences Compared With Adults. *Circulation: Genomic and Precision Medicine,* 11**,** e001887.

ZHU, Z., WANG, X., LI, X., LIN, Y., SHEN, S., LIU, C. L., HOBBS, B. D., HASEGAWA, K., LIANG, L., BOEZEN, H. M., CAMARGO, C. A., JR., CHO, M. H. & CHRISTIANI, D. C. 2019. Genetic overlap of chronic obstructive pulmonary disease and cardiovascular disease-related traits: a large-scale genome-wide cross-trait analysis. *Respir Res*.

ZOLI, A., ALTOMONTE, L., MIRONE, L., MAGARÓ, M., RICERCA, B. M., STORTI, S., CANDIDO, A. & BIZZI, M. 1994. Serum transferrin receptors in rheumatoid arthritis. *Annals of the Rheumatic Diseases,* 53**,** 699.