

# Association study of human leukocyte antigen (HLA) variants and idiopathic pulmonary fibrosis

Beatriz Guillen-Guio<sup>1,2\*</sup>, Megan L. Paynton<sup>1\*</sup>, Richard J. Allen<sup>1,2</sup>, Daniel P.W. Chin<sup>1,2</sup>, Lauren J. Donoghue<sup>3</sup>, Amy Stockwell<sup>3</sup>, Olivia C. Leavy<sup>1,2</sup>, Tamara Hernandez-Beeftink<sup>1,2</sup>, Carl Reynolds<sup>4</sup>, Paul Cullinan<sup>4</sup>, Fernando Martinez<sup>5</sup>, CleanUP-IPF Investigators of the Pulmonary Trials Cooperative<sup>5</sup>, Helen L. Booth<sup>6</sup>, William A. Fahy<sup>7</sup>, Ian P. Hall<sup>8,9</sup>, Simon P. Hart<sup>10</sup>, Mike R. Hill<sup>11</sup>, Nik Hirani<sup>12</sup>, Richard B. Hubbard<sup>8,9</sup>, Robin J. McAnulty<sup>13</sup>, Ann B. Millar<sup>14</sup>, Vidya Navaratnam<sup>15,16,17</sup>, Eunice Oballa<sup>7</sup>, Helen Parfrey<sup>18</sup>, Gauri Saini<sup>8</sup>, Ian Sayers<sup>19</sup>, Martin D. Tobin<sup>1,2</sup>, Moira K. B. Whyte<sup>12</sup>, Ayodeji Adegunsoye<sup>20</sup>, Naftali Kaminski<sup>21</sup>, Shwu-Fan Ma<sup>22</sup>, Mary E. Streck<sup>20</sup>, Yingze Zhang<sup>23</sup>, Tasha E. Fingerlin<sup>24</sup>, Maria Molina-Molina<sup>25,26,27</sup>, Margaret Neighbors<sup>3</sup>, X. Rebecca Sheng<sup>3</sup>, Justin M. Oldham<sup>28</sup>, Toby M. Maher<sup>29,30</sup>, Philip L. Molyneux<sup>29,31</sup>, Carlos Flores<sup>27,32,33,34</sup>, Imre Noth<sup>22</sup>, David A. Schwartz<sup>35</sup>, Brian L. Yaspan<sup>3</sup>, R. Gisli Jenkins<sup>29</sup>, Louise V. Wain<sup>1,2#</sup>, Edward J. Hollox<sup>36#</sup>.

\*Equal contribution as first authors

#Equal contribution as senior authors

<sup>1</sup>Department of Population Health Sciences, University of Leicester, UK

<sup>2</sup>NIHR Leicester Biomedical Research Centre, Leicester, UK

<sup>3</sup>Genentech, California, USA

<sup>4</sup>Imperial College London, London, UK

<sup>5</sup>Weill Cornell Medicine, New York, USA

<sup>6</sup>University College London Hospitals, London, UK

<sup>7</sup>GlaxoSmithKline, London, UK

<sup>8</sup>University of Nottingham, Nottingham, UK

<sup>9</sup>National Institute for Health Research, Nottingham Biomedical Research Centre, Nottingham, UK

<sup>10</sup>University of Hull, Hull, UK

<sup>11</sup>University of Oxford, Oxford, UK

<sup>12</sup>University of Edinburgh, Edinburgh, UK

<sup>13</sup>University College London, London, UK

<sup>14</sup>University of Bristol, Bristol, UK

<sup>15</sup>Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK

<sup>16</sup>National Institute for Health Research, Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust, Nottingham, UK

<sup>17</sup>Queensland Lung Transplant Service, The Prince Charles Hospital, Brisbane, QLD, Australia

<sup>18</sup>Royal Papworth Hospital NHS Foundation Trust, Cambridge, UK

<sup>19</sup>Centre for Respiratory Research, NIHR Nottingham Biomedical Research Centre, School of Medicine, Biodiscovery Institute, University of Nottingham, Nottingham, UK.

<sup>20</sup>University of Chicago, Chicago, USA

<sup>21</sup>Yale School of Medicine, Connecticut, USA

<sup>22</sup>University of Virginia, Virginia, USA

<sup>23</sup>University of Pittsburgh, Pittsburgh, USA

<sup>24</sup>National Jewish Health, Colorado, USA

<sup>25</sup>Servei de Pneumologia, Laboratori de Pneumologia Experimental, Instituto de Investigació  
NOF Médica de Benicarló (IDIBELL), Barcelona, Spain

This preprint reports a new research that has not been certified by peer review and should not be used to guide clinical practice.

<sup>26</sup>Campus de Bellvitge, Universitat de Barcelona, Barcelona, Spain

<sup>27</sup>Centro de Investigacion Biomedica en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain

<sup>28</sup>University of Michigan, Michigan, USA

<sup>29</sup>National Heart and Lung Institute, Imperial College London, London, UK

<sup>30</sup>Division of Pulmonary and Critical Care Medicine, University of Southern California, Los Angeles, USA

<sup>31</sup>Royal Brompton and Harefield Hospitals, Guy's and St Thomas' NHS Foundation Trust, London, UK

<sup>32</sup>Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain

<sup>33</sup>Genomics Division, Instituto Tecnológico y de Energias Renovables, Santa Cruz de Tenerife, Spain

<sup>34</sup>Facultad de Ciencias de la Salud, Universidad Fernando Pessoa Canarias, Las Palmas de Gran Canaria, Spain

<sup>35</sup>University of Colorado Medicine, Colorado, USA

<sup>36</sup>Department of Genetics and Genome Biology, University of Leicester, UK.

### **Corresponding author**

Dr. Edward J. Hollox

Department of Genetics and Genome Biology, University of Leicester, Leicester, UK

Email: [ejh33@leicester.ac.uk](mailto:ejh33@leicester.ac.uk)

## ABSTRACT

### Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic interstitial pneumonia marked by progressive lung fibrosis and a poor prognosis. Recent studies have highlighted the potential role of infection in the pathogenesis of IPF and a prior association of the *HLA-DQB1* gene with idiopathic fibrotic interstitial pneumonia (including IPF) has been reported. Due to the important role that the Human Leukocyte Antigen (HLA) region plays in the immune response, here we evaluated if HLA genetic variation was associated specifically with IPF risk.

### Methods

We performed a meta-analysis of associations of the HLA region with IPF risk in individuals of European ancestry from seven independent case-control studies of IPF (comprising a total of 5,159 cases and 27,459 controls, including the prior study of fibrotic interstitial pneumonia). Single nucleotide polymorphisms, classical HLA alleles and amino acids were analysed and signals meeting a region-wide association threshold  $p < 4.5 \times 10^{-4}$  and a posterior probability of replication  $> 90\%$  were considered significant. We sought to replicate the previously reported *HLA-DQB1* association in the subset of studies independent of the original report.

### Results

The meta-analysis of all seven studies identified four significant independent single nucleotide polymorphisms associated with IPF risk. However, none met the posterior probability for replication criterion. The *HLA-DQB1* association was not replicated in the independent IPF studies.

### Conclusion

Variation in the HLA region was not consistently associated with risk in studies of IPF. However, this does not preclude the possibility that other genomic regions linked to the immune response may be involved in the aetiology of IPF.

## INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a lung disease characterised by progressive scarring of the alveoli leading to impaired gas exchange. The median survival after IPF diagnosis is 2-5 years and there are limited drug treatments for patients (1). Several studies have reported both environmental and genetic risk factors related to IPF (2). However, the cause of IPF development remains unclear.

Genetic variation in the Human Leukocyte Antigen (HLA) region, also known as the major histocompatibility complex (MHC) region, has been reported as associated with inflammatory and respiratory diseases (3,4). This includes fibrotic idiopathic interstitial pneumonia (fIIP), where *HLA-DQB1\*06:02* has been associated with increased risk of disease (5). Additionally, *HLA-DRB1\*15:01* has been found to be more prevalent among IPF patients than controls (6). The mechanism behind these associations is unclear. However, it has been previously suggested that respiratory infections, including coronavirus infectious disease 2019 (COVID-19), could trigger the development and progression of interstitial lung diseases, including IPF (7-9). Indeed, antiviral drugs against herpesviruses have been proposed to attenuate disease progression in IPF patients (10). Given the important role of the MHC receptors encoded by HLA genes in presenting viral antigens to the host immune system, genetic variation at the HLA region could influence the response to these infections and therefore IPF pathophysiology.

Recent studies on large biobanks have emphasised the pleiotropy of the HLA region (11). This region is highly polymorphic and gene-dense, making the interpretation of single nucleotide polymorphisms (SNP) associations difficult, and its study requires specific imputation techniques (12). In this sense, the latest SNP imputation panels enable improved imputation accuracy within the HLA region. Additionally, specific HLA imputation panels also allow the imputation of classical HLA alleles and amino acid alleles across the HLA region, which are biologically informative (13).

Given the potential role of infection response in IPF pathophysiology, and the suggested association of *HLA-DQB1\*06:02* with disease, here we present an analysis of the HLA region across seven independent case-control cohorts for IPF, with a total of over 5,000 cases and 27,000 controls. We used an HLA-specific imputation approach to impute SNP variation, classical HLA alleles and amino acid alleles (hereafter to be known as variants) with the aim of identifying novel IPF risk loci within the HLA region.

## METHODS

### Sample, genotyping and quality controls

We analysed genomic data from seven previously described independent case-control studies, named here as CleanUP-UCD [Study of Clinical Efficacy of Antimicrobial Therapy Strategy Using Pragmatic Design - University of California Davis] (14,15), Colorado (16), Genentech (17,18), IPF-JES [Job Exposures Study] (19), UK (20), US (21) and UUS [United States, United Kingdom, and Spain] (22). CleanUP-UCD, IPF-JES, UK, US and UUS studies included patients diagnosed with IPF as cases and population controls. The Colorado study used cases of fIIP (including IPF cases among other conditions) and population controls and was the study that previously reported the fIIP association for *HLA-DQB1\*06:02* (5). The Genentech study included patients with IPF and controls from non-IPF clinical trials of age-related macular degeneration, diabetic macular oedema, multiple sclerosis, asthma, and inflammatory bowel disease. In all seven studies, cases were diagnosed according to the American Thoracic Society and European Respiratory Society guidelines (23,24). The studies were

performed in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) and approved by the appropriate institutional review or Research Ethics Committee. Further details about the seven studies can be found in the Supplement.

In Colorado, IPF-JES, UK, US, UUS, and CleanUP-UCD studies, individuals were genotyped using SNP arrays (Affymetrix and Illumina Inc.). Genotyping and quality control procedures have been previously detailed (20,22,25). In summary for the six studies, individuals were filtered based on low call rate estimates, sex mismatches, excess heterozygosity, relatedness, and non-European ancestry (based on principal components [PC] clustering with European individuals from The 1000 Genomes project). Individuals overlapping between studies were removed. Genotypes from individuals in the Genentech study were obtained from whole-genome sequencing using HiSeq X Ten platform (Illumina Inc.) to an average read depth of 30X. Related individuals and those with genotype missingness >10% or excess heterozygosity were excluded. Ancestry was determined using ADMIXTURE v1.23 (26), and individuals with a European genetic ancestry score > 0.7 were included in the analyses.

### **Variant imputation in the HLA region and association tests**

For each study, classical HLA alleles, amino acid alleles (AA) and SNPs located within chr6:28,510,120-33,480,577 (GRCh38) were imputed to capture genetic variation in the HLA region.

For Chicago, Colorado, UK, UUS, CleanUP-UCD, and IPF-JES studies, phasing and SNP Imputation to the TOPMed reference panel was performed using the TOPMed Imputation Server (6). Classical HLA alleles from class I genes (-A, -B, C) and class II genes (-DPA, -DPB, -DQA, -DQB, -DRB), amino acid (AA) alleles, and additional SNPs were imputed to the type one diabetes genomics consortium (T1DGC) panel using IMPUTE2 (V2.3.2) after chromosome phasing with SHAPE-IT v2.837 (27). After SNP imputation using both TOPMed and T1DGC panels, duplicated SNPs were removed, prioritising those imputed with TOPMed. For the Genentech study, variants with a minor allele frequency (MAF)<1% or ambiguous and SNPs absent from the TOPMed reference panel were removed. Classical HLA alleles and amino acid alleles were imputed using HLA-HD (28), and data was analysed using MiDAS (29).

Logistic regression analyses were performed in each study separately assuming an additive model. For Chicago, Colorado, UK, UUS, CleanUP-UCD, and IPF-JES studies, models were adjusted for the ten leading PCs to correct for population stratification. For the Genentech study, sex, age, and five genetic-ancestry PCs were included as covariates and reverse regression (30) was applied to flag any associations driven by the differential allele frequency of a control indication(s) rather than IPF. In all studies, low frequency variants (MAF < 1%) and variants with a poor imputation quality ( $r^2 < 0.3$ ) were removed from the analyses. A fixed-effect weighted meta-analysis combining the association results of the seven studies was performed using METAL (31) to establish the genetic variants associated with IPF. Variants were required to be present in at least two studies to be included in the analysis. The significance threshold was declared at  $p=4.50 \times 10^{-4}$  after Bonferroni correction based on the number of independent SNPs, amino acid changes, and classical HLA alleles in this region. The number of independent signals in the region was identified using LD-prune from PLINK v1.9 ( $r^2=0.2$ ) (32). Conditional analyses were performed with COJO-GCTA (33) to identify independent sentinel variants. We used the Meta-Analysis Model-Based Assessment of Replicability (MAMBA) to assess the consistency of association results across studies (34). This tool calculates a posterior probability of replicability (PPR) that a given SNP has a non-zero replicable effect, indicating the likelihood of a specific SNP to replicate. Associations with a PPR >90% were considered consistent and likely to replicate. The Manhattan and conditional analyses plots were obtained using HLA-TAPAS (35) and R

v4.0.0, respectively. Forest plots were performed with the *forestplot* R v4.1.3 package. Replication of the previously association with *HLA-DQB1\*06:02* was sought using the 6 independent datasets (i.e., excluding Colorado study). Sensitivity analyses of all significant signals were conducted excluding the Colorado study, given that the case definition used included other fIIPs in addition to IPF.

## RESULTS

A total of 32,618 unrelated individuals of European ancestry (5,159 cases and 27,459 controls, Table S1) and 44,713 common variants within the HLA region (43,544 SNPs, 174 classical alleles, and 995 AA) were included in the analyses. After meta-analysis, four independent loci met the HLA-wide significance threshold ( $p=4.50 \times 10^{-4}$ ) (Table 1, Table S2, Figures 1 and 2).

**Table 1: Association analysis results for sentinel variants**

Variant ID	Position (b38)	Nearest gene/s	Non-effect allele	Effect allele	EAfreq	Direction*	Study $p \leq 0.05$ *	OR [95%CI]	P	MAMBA (PPR)
rs4947344	chr6:2710069	<i>DQB1/HLA-DQA2</i>	C	T	0.286	+++++	YYNNNNY	1.14 [1.09,1.20]	$5.60 \times 10^{-8}$	0.85%
rs9266618	chr6:31378389	<i>HLA-B/MICA</i>	A	C	0.093	+++++	NYNNNNN	1.20 [1.11,1.29]	$2.73 \times 10^{-6}$	0.63%
rs116450899	chr6:28764432	<i>ZBED9/TRIM27</i>	G	A	0.090	+++++	NYNNNNN	1.19 [1.10,1.28]	$4.65 \times 10^{-6}$	1.31%
rs138188178	chr6:30632465	<i>ATAT1</i>	G	T	0.043	+++++	NNNNNNY	1.21 [1.09,1.34]	$4.14 \times 10^{-4}$	0.28%

EAfreq, frequency of the effect allele across the entire sample. MAMBA, Meta-Analysis Model-Based Assessment of Replicability; PPR, posterior probability of replication. Order for effect Direction and Study significance is: **CleanUP-UCD – Colorado – Genentech – IPF-JES – UK – US – UUS** (Y/N: Yes/No).

The top sentinel variant was the SNP rs4947344, an intergenic variant located between *HLA-DQB1* and *HLA-DQA2* that was associated with IPF risk in our study (odds ratio (OR) = 1.14, 95% confidence interval (CI) = 1.09-1.20,  $p=5.60 \times 10^{-8}$ ). After conditional analyses, we found that this variant was in linkage disequilibrium with the allele *HLA-DQB1\*06:02*, previously associated with fIIP (5) (Figure S1).

The 3 remaining sentinel variants identified in the meta-analysis results were SNPs rs9266618, rs116450899, and rs138188178 (Table 1, Figures 1 and 2). rs9266618 and rs116450899 are intergenic variants located between *HLA-B* and the MHC Class I Polypeptide-Related Sequence A (*MICA*), and between the zinc finger BED-type containing 9 (*ZBED9*) and the tripartite motif containing 27 (*TRIM27*) genes, respectively, while rs138188178 is intronic to the gene encoding the Alpha Tubulin Acetyltransferase 1 (*ATAT1*).

None of the variants met nominal significance ( $p < 0.05$ ) in all seven contributing studies and the replicability assessment with MAMBA revealed that none of these association signals were considered likely to replicate (PPR < 90% in all cases) (Table 1, Figure 2).

After sensitivity analyses excluding the Colorado study, three signal associations reached the Bonferroni P threshold, although the likelihood of replication remained very small for all three (Table S3). The association of *HLA-DQB1\*06:02* with IPF was not independently replicated in the six independent IPF datasets ( $p=0.043$ , PPR= 0.09%) (Table S4, Figure S2).

## DISCUSSION

In this study, we used genotype data and imputation of SNPs, amino acids and classical HLA alleles from seven independent case-control cohorts to test for association with IPF susceptibility. Although we identified four significant associations, including a signal correlated with a previously reported association for *HLA-DQB1\*06:02*, the effects were inconsistent across the contributing studies and this heterogeneity was reflected in a poor posterior probability for replication for all signals. The reason for this heterogeneity is unclear, but could be due to a broader definition of cases (including non-IPF fIIP) in the Colorado study (5) or use of a different methodology (whole genome sequencing) in the Genentech study, among other possibilities. While control subjects in the Genentech study were selected from clinical trials of non-IPF diseases, including those strongly associated with variation in the HLA region (e.g., asthma), reverse regression on these four associations did not detect that the control cohort composition was driving the IPF effect estimates (18).

We acknowledge several limitations. First, our study was restricted to individuals of European genetic ancestries. Thus, we cannot exclude the possibility that HLA variation might contribute to IPF susceptibility in individuals of other genetic ancestries. Next, the heterogeneity of signals across cohorts might suggest that HLA variation could contribute to forms of pulmonary fibrosis that do not meet the strict diagnostic criteria for IPF. Due to technology limitations, we were unable to evaluate all types of classical HLA alleles, which could have concealed potential IPF associations. Finally, the HLA is a complex region with known genetic and environmental interactions. Therefore, our approach of testing individual HLA variants with disease may be too simplistic to capture the complex interplay of these proteins with disease risk.

Our meta-analysis approach has the advantage of increasing the sample size for genetic studies of IPF. We restricted our analysis to studies of clinically defined IPF to avoid the issues of imprecise electronic healthcare record coding and resulting effect size attenuation that has been previously reported (36,37). By combining all available datasets, we were able to improve power to detect associations of large and modest effect sizes but at the expense of reserving samples for independent replication. To mitigate this, we used an approach that considers heterogeneity of association results across the meta-analysis to provide a posterior probability of replication. This approach allowed us to provide a quantitative indicator of the combined strength of the associations across studies rather than relying on a single combined P value which can be influenced by effects in individual studies.

Our results strongly suggest that individual genetic variants within the HLA are not associated with susceptibility to IPF in individuals with European ancestries, although they may have a role in other IIPs. Future studies investigating the role of infection in the aetiology of IPF should aim to assess the relationship of interactions of immune system genes.

## FUNDING

BGG is supported by Wellcome Trust grant 221680/Z/20/Z. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. MLP was funded by a University of Leicester College of Life Sciences PhD studentship. JMO reports National Institutes of Health (NIH) National Heart, Lung, and Blood Institute grants R56HL158935 and K23HL138190. CF is supported by the Instituto de Salud Carlos III (PI20/00876) and the Spanish Ministry of Science and Innovation (grant RTC-2017-6471-1), co-financed by the European Regional Development Funds (A way of making Europe) from the EU. RGJ and LVW report funding from the Medical Research Council (MR/V00235X/1). LVW holds a GlaxoSmithKline / Asthma + Lung UK Chair in Respiratory Research (C17-1). This work was partially supported by the National Institute for Health Research (NIHR) Leicester Biomedical Research Centre; the views expressed are those of the author(s) and not necessarily those of the National Health Service, the NIHR, or the Department of Health. This research includes use of the UK Biobank through application 8389 and used the SPECTRE High Performance Computing Facility at the University of Leicester.

## COMPETING INTERESTS

LJD, AS, MN, XRS and BLY are full-time employees of Genentech; AS, MN, XRS, and BLY and hold stock options in Roche. JMO reports personal fees from Boehringer Ingelheim, Genentech, United Therapeutics, AmMax Bio and Lupin Pharmaceuticals outside of the submitted work. DAS is the founder and chief scientific officer of Eleven P15, a company focused on the early detection and treatment of pulmonary fibrosis. RGJ reports honoraria from Chiesi, Roche, PatientMPower, AstraZeneca, GSK, Boehringer Ingelheim, and consulting fees from Bristol Myers Squibb, Daewoong, Veracyte, Resolution Therapeutics, RedX, Pliant, Chiesi. LVW reports research funding from GlaxoSmithKline, Roche and Orion Pharma, and consultancy for GlaxoSmithKline and Galapagos, outside of the submitted work. The other authors declare no competing interests.

## AUTHORS CONTRIBUTION

LVW and EJH designed and supervised the study. BG-G, MLP, RJA, DPC, LJD, AS, OCL and TH-B performed the analyses. CR, PC, FM, CleanUP-IPF Investigators of the Pulmonary Trials Cooperative, HLB, WAF, IPH, SPH, MRH, NH, RBH, RJM, ABM, VN, EO, HP, GS, IS, MDT, MKBW, AA, NK, S-FM, MES, YZ, TEF, MM-M, MN, XRS, JMO, TMM, PLM, CF, IN, DAS, BLY, RGJ and LVW participated in data collection. BG-G, RJA, LVW and EJH wrote the first draft of the manuscript. All authors revised and approved the final version of the manuscript.

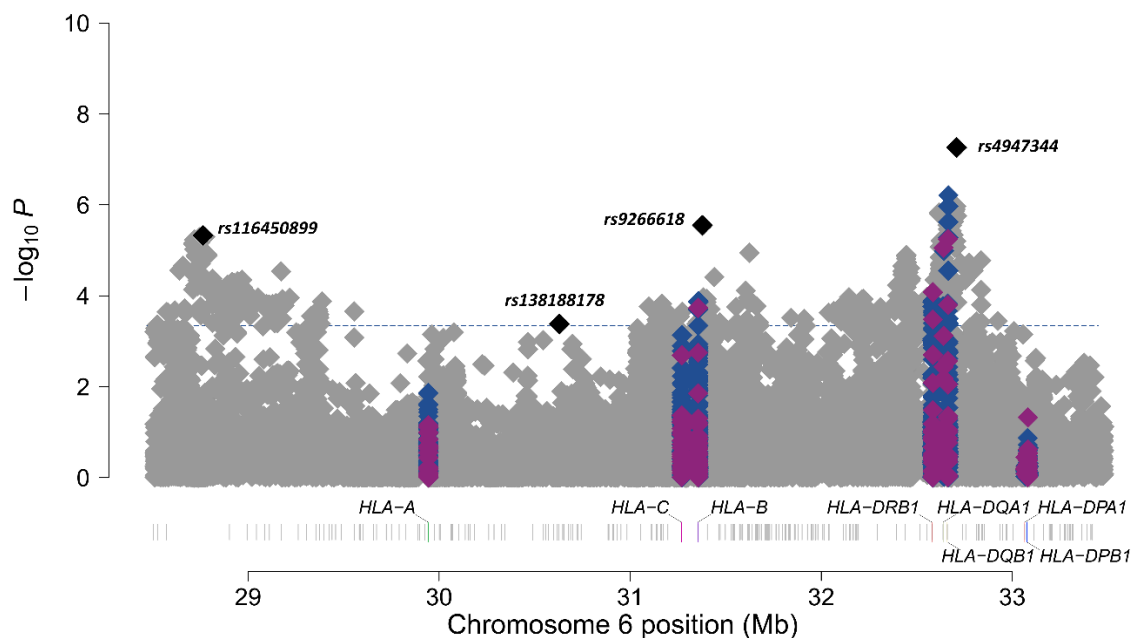


## REFERENCES

1. Fujimoto H, Kobayashi T, Azuma A. Idiopathic Pulmonary Fibrosis: Treatment and Prognosis. *Clin Med Insights Circ Respir Pulm Med* 2015;**9**(Suppl 1):179-85.
2. Michalski JE, Schwartz DA. Genetic Risk Factors for Idiopathic Pulmonary Fibrosis: Insights into Immunopathogenesis. *J Inflamm Res* 2020;**13**:1305-18.
3. Obeidat M, Hall IP. Genetics of complex respiratory diseases: implications for pathophysiology and pharmacology studies. *Br J Pharmacol* 2011;**163**(1):96-105.
4. Dendrou CA, Petersen J, Rossjohn J, et al. HLA variation and disease. *Nat Rev Immunol* 2018;**18**(5):325-39.
5. Fingerlin TE, Zhang W, Yang IV, et al. Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet* 2016;**17**(1):74.
6. Xue J, Gochuico BR, Alawad AS, et al. The HLA class II Allele DRB1\*1501 is over-represented in patients with idiopathic pulmonary fibrosis. *PLoS One* 2011;**6**(2):e14715.
7. Sheng G, Chen P, Wei Y, et al. Viral Infection Increases the Risk of Idiopathic Pulmonary Fibrosis: A Meta-Analysis. *Chest* 2020;**157**(5):1175-87.
8. Magro CM, Allen J, Pope-Harman A, et al. The role of microvascular injury in the evolution of idiopathic pulmonary fibrosis. *Am J Clin Pathol* 2003;**119**(4):556-67.
9. Mehta P, Rosas IO, Singer M. Understanding post-COVID-19 interstitial lung disease (ILD): a new fibroinflammatory disease entity. *Intensive Care Med* 2022;**48**(12):1803-06.
10. Egan JJ, Adamali HI, Lok SS, et al. Ganciclovir antiviral therapy in advanced idiopathic pulmonary fibrosis: an open pilot study. *Pulm Med* 2011;**2011**:240805.
11. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* 2021;**53**(10):1415-24.
12. Choo SY. The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med J* 2007;**48**(1):11-23.
13. Meyer D, Nunes K. HLA imputation, what is it good for? *Hum Immunol* 2017;**78**(3):239-41.
14. Martinez FJ, Yow E, Flaherty KR, et al. Effect of Antimicrobial Therapy on Respiratory Hospitalization or Death in Adults With Idiopathic Pulmonary Fibrosis: The CleanUP-IPF Randomized Clinical Trial. *JAMA* 2021;**325**(18):1841-51.
15. Allen RJ, Oldham JM, Jenkins DA, et al. Longitudinal lung function and gas transfer in individuals with idiopathic pulmonary fibrosis: a genome-wide association study. *Lancet Respir Med* 2023;**11**(1):65-73.
16. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013;**45**(6):613-20.
17. Dressen A, Abbas AR, Cabanski C, et al. Analysis of protein-altering variants in telomerase genes and their association with MUC5B common variant status in patients with idiopathic pulmonary fibrosis: a candidate gene sequencing study. *Lancet Respir Med* 2018;**6**(8):603-14.
18. Donoghue LJ, Stockwell AD, Neighbors M, et al. Identification of a Genetic Susceptibility Locus for Idiopathic Pulmonary Fibrosis in the 16p Subtelomere Using Whole-Genome Sequencing. *Am J Respir Crit Care Med* 2023;**207**(7):941-44.
19. Reynolds CJ, Sisodia R, Barber C, et al. What role for asbestos in idiopathic pulmonary fibrosis? Findings from the IPF job exposures case-control study. *Occup Environ Med* 2023;**80**(2):97-103.
20. Allen RJ, Porte J, Braybrooke R, et al. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir Med* 2017;**5**(11):869-80.
21. Noth I, Zhang Y, Ma SF, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013;**1**(4):309-17.

22. Allen RJ, Guillen-Guio B, Oldham JM, et al. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2020;**201**(5):564-74.
23. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2018;**198**(5):e44-e68.
24. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;**183**(6):788-824.
25. Allen RJ, Stockwell A, Oldham JM, et al. Genome-wide association study across five cohorts identifies five novel loci associated with idiopathic pulmonary fibrosis. *Thorax* 2022;**77**(8):829-33.
26. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;**19**(9):1655-64.
27. Delaneau O, Coulonges C, Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics* 2008;**9**:540.
28. Kawaguchi S, Higasa K, Shimizu M, et al. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat* 2017;**38**(7):788-97.
29. Migdal M, Ruan DF, Forrest WF, et al. MiDAS-Meaningful Immunogenetic Data at Scale. *PLoS Comput Biol* 2021;**17**(7):e1009131.
30. Tom J, Chang D, Wuster A, et al. Enabling genome-wide association testing with multiple diseases and no healthy controls. *Gene* 2019;**684**:118-23.
31. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;**26**(17):2190-1.
32. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;**4**:7.
33. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012;**44**(4):369-75, S1-3.
34. McGuire D, Jiang Y, Liu M, et al. Model-based assessment of replicability for genome-wide association meta-analysis. *Nat Commun* 2021;**12**(1):1964.
35. Luo Y, Kanai M, Choi W, et al. A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nat Genet* 2021;**53**(10):1504-16.
36. Leavy OC, Allen RJ, Kraven LM, et al. The Use of Genetic Information to Define Idiopathic Pulmonary Fibrosis in UK Biobank. *Chest* 2023;**163**(2):362-65.
37. Partanen JJ, Happola P, Zhou W, et al. Leveraging global multi-ancestry meta-analysis in the study of idiopathic pulmonary fibrosis genetics. *Cell Genom* 2022;**2**(10):100181.

## FIGURES



**Figure 1. Manhattan plot of meta-analysis results at chromosome 6 region.** SNPs are shown in grey, amino acids in blue, and classical HLA alleles in purple. The four sentinel variants are highlighted in black. The y-axis shows the transformed p-values ( $-\log_{10}$ ) while the x-axis represents chromosome positions in Mb (GRCh38/hg38). The horizontal line corresponds to the significance threshold of the study after Bonferroni correction ( $p=4.50 \times 10^{-4}$ ).

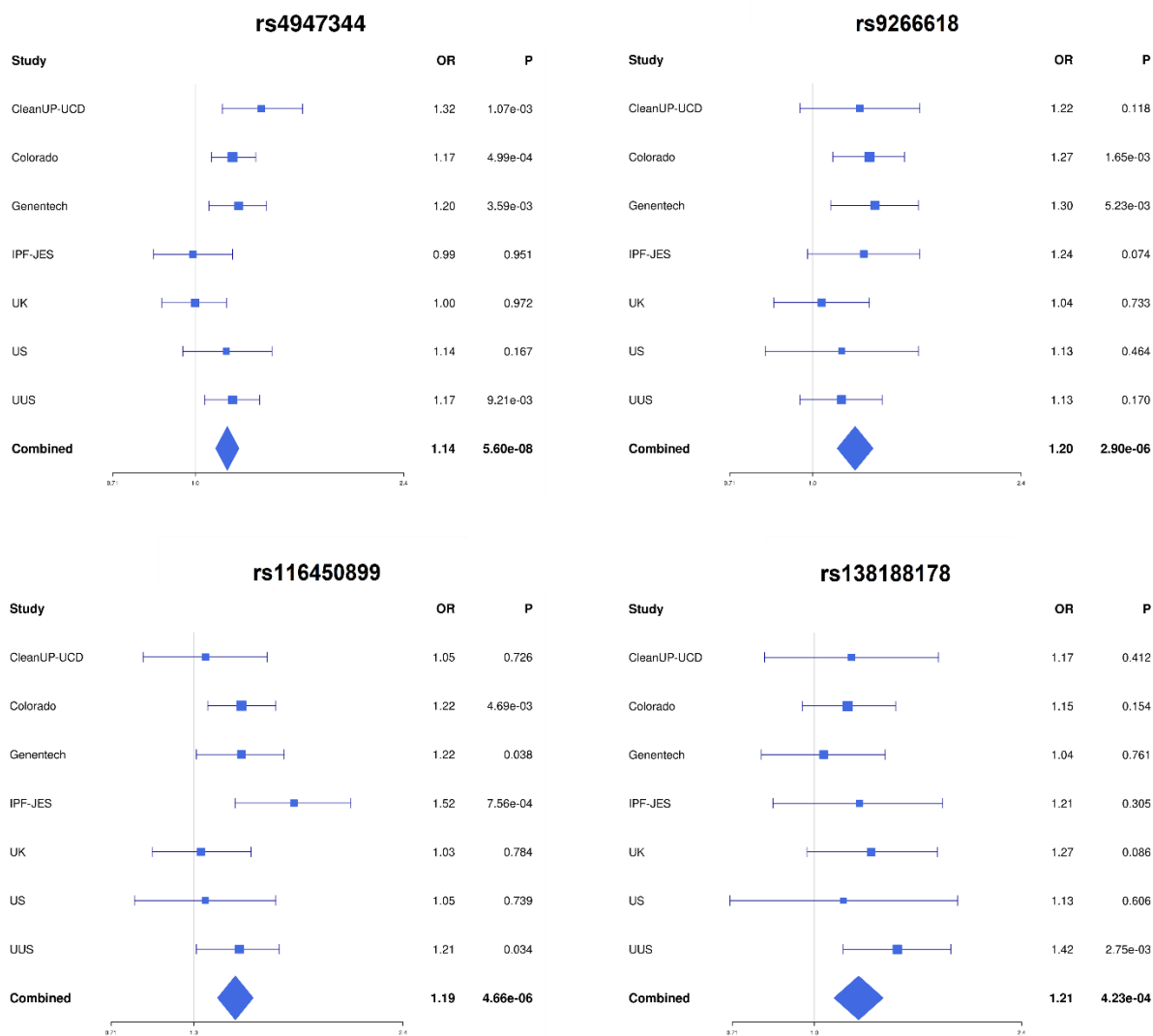


Figure 2. Forest plots of the association analysis results for the four sentinel variants.