

# Genetic determinants of risk in autoimmune pulmonary alveolar proteinosis

Saori Sakaue <sup>1,2,3,4,5,44</sup>, Etsuro Yamaguchi<sup>6,44</sup>, Yoshikazu Inoue <sup>7</sup>, Meiko Takahashi<sup>8</sup>, Jun Hirata<sup>1,9</sup>, Ken Suzuki <sup>1</sup>, Satoru Ito <sup>6</sup>, Toru Arai<sup>7</sup>, Masaki Hirose<sup>7</sup>, Yoshinori Tanino <sup>10</sup>, Takefumi Nikaido<sup>10</sup>, Toshio Ichiwata<sup>11</sup>, Shinya Ohkouchi <sup>12</sup>, Taizou Hirano<sup>13</sup>, Toshinori Takada<sup>14</sup>, Satoru Miyawaki <sup>15</sup>, Shogo Dofuku<sup>15</sup>, Yuichi Maeda<sup>16,17</sup>, Takuro Nii<sup>16,17</sup>, Toshihiro Kishikawa<sup>1,18</sup>, Kotaro Ogawa<sup>1,19</sup>, Tatsuo Masuda<sup>1,20</sup>, Kenichi Yamamoto<sup>1,21</sup>, Kyuto Sonehara<sup>1</sup>, Ryushi Tazawa<sup>22</sup>, Konosuke Morimoto<sup>23</sup>, Masahiro Takaki<sup>24</sup>, Satoshi Konno<sup>25</sup>, Masaru Suzuki<sup>25</sup>, Keisuke Tomii<sup>26</sup>, Atsushi Nakagawa<sup>26</sup>, Tomohiro Handa<sup>27</sup>, Kiminobu Tanizawa<sup>28</sup>, Haruyuki Ishii<sup>29</sup>, Manabu Ishida<sup>29</sup>, Toshiyuki Kato<sup>30</sup>, Naoya Takeda<sup>30</sup>, Koshi Yokomura<sup>31</sup>, Takashi Matsui<sup>31</sup>, Masaki Watanabe<sup>32</sup>, Hiromasa Inoue <sup>32</sup>, Kazuyoshi Imaizumi<sup>33</sup>, Yasuhiro Goto<sup>33</sup>, Hiroshi Kida <sup>16,34</sup>, Tomoyuki Fujisawa<sup>35</sup>, Takafumi Suda<sup>35</sup>, Takashi Yamada<sup>36</sup>, Yasuomi Satake <sup>36</sup>, Hidenori Ibata<sup>37</sup>, Nobuyuki Hizawa<sup>38</sup>, Hideki Mochizuki <sup>19</sup>, Atsushi Kumanogoh <sup>16,39,40</sup>, Fumihiko Matsuda<sup>8</sup>, Koh Nakata<sup>41</sup>, Tomomitsu Hirota<sup>42</sup>, Mayumi Tamari <sup>42</sup> & Yukinori Okada <sup>1,40,43</sup>✉

Pulmonary alveolar proteinosis (PAP) is a devastating lung disease caused by abnormal surfactant homeostasis, with a prevalence of 6–7 cases per million population worldwide. While mutations causing hereditary PAP have been reported, the genetic basis contributing to autoimmune PAP (aPAP) has not been thoroughly investigated. Here, we conducted a genome-wide association study of aPAP in 198 patients and 395 control participants of Japanese ancestry. The common genetic variant, rs138024423 at 6p21, in the major-histocompatibility-complex (MHC) region was significantly associated with disease risk (Odds ratio [OR] = 5.2;  $P = 2.4 \times 10^{-12}$ ). HLA fine-mapping revealed that the common HLA class II allele, HLA-DRB1\*08:03, strongly drove this signal (OR = 4.8;  $P = 4.8 \times 10^{-12}$ ), followed by an additional independent risk allele at HLA-DPβ1 amino acid position 8 (OR = 0.28;  $P = 3.4 \times 10^{-7}$ ). HLA-DRB1\*08:03 was also associated with an increased level of anti-GM-CSF antibody, a key driver of the disease ( $\beta = 0.32$ ;  $P = 0.035$ ). Our study demonstrated a heritable component of aPAP, suggesting an underlying genetic predisposition toward an abnormal antibody production.

**P**ulmonary alveolar proteinosis (PAP) is a rare diffuse lung disease with a prevalence of 6–7 cases per million population worldwide<sup>1,2</sup>. The disease is characterized by the abnormal accumulation of pulmonary surfactant within pulmonary alveoli, resulting in progressive respiratory failure and increased infection risk<sup>2–5</sup>. PAP has three distinct etiologies: hereditary, autoimmune, and secondary<sup>6</sup>. Approximately 90–95% of cases of PAP are of autoimmune etiology, in which a high level of autoantibodies against granulocyte–macrophage colony-stimulating factor (GM-CSF) neutralize the biologic activity of GM-CSF<sup>7</sup>, thereby causing poor surfactant clearance. While rare mutations in the GM-CSF receptor  $\alpha$  or  $\beta$  chains (*CSF2RA*, *CSF2RB*)<sup>8</sup>, the surfactant proteins B or C<sup>9</sup>, ATP-binding cassette subfamily A member 3 (*ABCA3*)<sup>10</sup>, and thyroid transcription factor-1<sup>11</sup> were identified as causing hereditary PAP, the genetic basis underlying autoimmune PAP (aPAP) has never been thoroughly investigated due to its low prevalence. In particular, given a growing evidence that common genetic variants in major-histocompatibility-complex (MHC) region are associated with rare autoimmune diseases<sup>12–14</sup>, an investigation of human leukocyte antigen (HLA) alleles associated with the risk of aPAP and the production of autoantibodies against GM-CSF is warranted.

Here, we perform a genome-wide association study (GWAS) of aPAP in the Japanese population to investigate the genetic basis in the risk of aPAP. We further perform HLA imputation to fine map the association signal within the MHC locus. We show a significant association of an HLA allele with the risk of aPAP, as well as increased levels of anti-GM-CSF antibody.

## Results

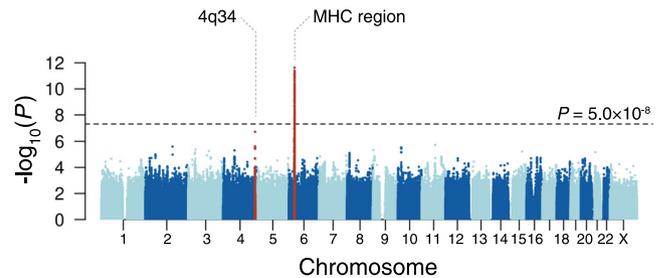
**Study populations.** We studied a total of 198 patients with aPAP of Japanese ancestry and 395 control participants of Japanese ancestry. The patients were enrolled through a nation-wide collaborative recruitment strategy<sup>2,15</sup>. The patient characteristics are summarized in Table 1. All patients were confirmed to have a positive serum anti-GM-CSF antibody level.

**Single-marker association test.** We performed a one-stage GWAS with all the case samples and controls together to maximize the statistical power, given the low prevalence of the disease. We analyzed 12,153,232 genetic markers that passed the stringent post-imputation quality control threshold ( $R_{sq} > 0.7$ ). The MHC locus was identified as associated with the aPAP risk with a genome-wide significance ( $P < 5.0 \times 10^{-8}$ ; Fig. 1; Supplementary Table 1). The lead variant was rs138024423, where the G allele (deletion) was strongly associated with the increased disease risk (Odds ratio [OR], 5.2; 95% confidence interval [CI], 3.3 to 8.2;  $P = 2.4 \times 10^{-12}$ ). The frequency of G allele was higher in cases than in controls (the allele frequency = 0.21 and 0.072, respectively). We additionally identified the risk variant with suggestive significance at 4q34 (rs56125424; OR, 3.0; 95% CI, 2.0–4.6;  $P = 1.9 \times 10^{-7}$ ; Supplementary Fig. 1 for a regional plot). This region is relatively devoid of known genes, and the suggestive variant rs56125424 is located in the intergenic region between the LINC01098 and LINC00290 (1.8 Mb and 1.3 Mb away, respectively). The previously reported genetic associations of this region (i.e., the variants located within 1 Mb from rs56125424) with human complex traits included body mass index<sup>16–18</sup>, waist-to-hip ratio<sup>16</sup>, hand-grip strength<sup>19</sup>, smoking initiation<sup>20</sup>, general risk tolerance<sup>21</sup>, and motion sickness<sup>22</sup>. No autoimmune diseases have been reported to date according to the GWAS catalog<sup>23</sup>. There was no interaction effect between rs138024423 at MHC locus and rs56125424 at 4q34 on the risk of aPAP ( $P = 0.70$ ).

**Table 1 Clinical characteristics of the patients and the controls at baseline.**

Variable	PAP patients (N = 198)	Control participants (N = 395)
Age—yr	57.4 ± 12.9	48.7 ± 21.3
Female sex—no. (%)	79 (40)	185 (47)
Autoantibodies against GM-CSF— $\mu\text{g/ml}$	69.3 ± 80.2	N/A

Plus-minus values are means ± standard deviation.

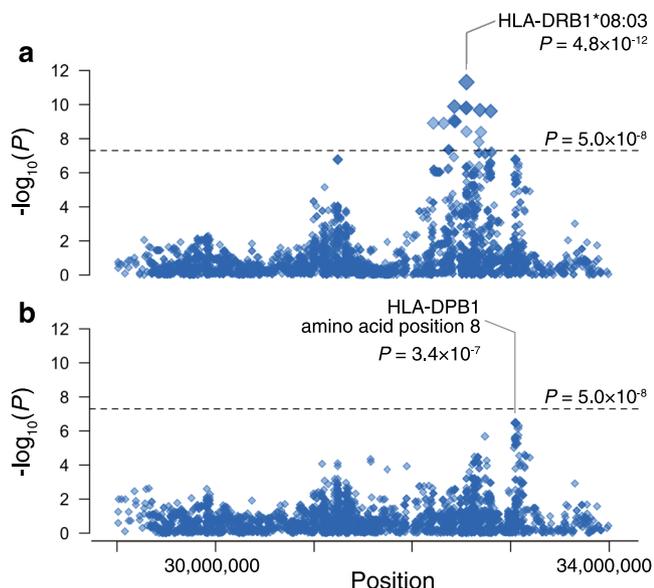


**Fig. 1 Manhattan plot for the genome-wide association study of autoimmune pulmonary alveolar proteinosis.** A Manhattan plot showing  $-\log_{10}(P)$  of the genome-wide single-marker association test of aPAP in the Japanese population.

The dotted horizontal line represents a genome-wide significance threshold of  $P = 5.0 \times 10^{-8}$ . The loci that satisfied the genome-wide significance threshold are colored in red. MHC, major histocompatibility complex.

**MHC risk fine-mapping by HLA imputation.** To fine map the disease risk within the complex MHC region<sup>24,25</sup>, we applied HLA imputation method to the genotype data using a high-resolution reference panel of 1120 individuals of Japanese ancestry<sup>25</sup>. After the stringent post-imputation quality control, we obtained genotype dosages of 101 two-digit, 164 four-digit, and 180 six-digit HLA alleles and 1701 amino acid polymorphisms of classical and nonclassical HLA genes in the entire MHC region. We assessed the association of the imputed HLA alleles and amino acid polymorphisms with the risk of aPAP. We found that the common HLA class II allele, HLA-DRB1\*08:03, strongly drove the signal in the MHC region (Fig. 2a; OR, 4.8; 95% CI, 3.1–7.5;  $P = 4.8 \times 10^{-12}$ ). The frequency of this allele was 0.22 in cases, whereas 0.074 in controls. To rule out the potential uncontrolled bias caused by the heterogeneous distribution of the HLA alleles according to the geographical location, we conducted a sensitivity analysis by stratifying the whole cohort into two datasets based on the location of the recruitment centers, Set 1 and Set 2 (see “Methods”). We observed that the effect sizes of both datasets were consistent with the primary association (OR, 5.6 and 4.7, respectively; Supplementary Fig. 2), and thus concluded that uncontrolled biases should be minimal. We also confirmed that a suggestive signal at 4q34 had similar effect sizes in stratified analyses as well (OR, 3.6 and 3.9, respectively; Supplementary Fig. 3).

To identify additional MHC associations independent of *HLA-DRB1*, we conditioned on HLA-DRB1\*08:03 and tested again the HLA alleles and amino acid polymorphisms. We observed an independent association signal at position 8 of the amino acid polymorphisms in *HLA-DPBI* (Fig. 2b;  $P = 3.4 \times 10^{-7}$ ). At this position, the amino acid residue Val8 was protective against aPAP when compared with the amino acid residue Leu8 (OR, 0.28; 95% CI, 0.17–0.46). After conditioned on HLA-DRB1\*08:03



**Fig. 2 Regional associations of the variants in the major histocompatibility complex (MHC) region with autoimmune pulmonary alveolar proteinosis risk.** Regional associations of the variants in the MHC region with aPAP risk based on the HLA imputation analysis. **a** Each diamond represents the  $-\log_{10}(P)$  of association of the variants, including the single nucleotide polymorphism, two-digit and four-digit HLA alleles, and amino acid polymorphisms of HLA genes. The dotted horizontal line represents the genome-wide significance threshold of  $P = 5.0 \times 10^{-8}$ . The lead HLA allele is labeled. **b** Each diamond represents the  $-\log_{10}(P)$  of the secondary associations after conditioning on the dosage of the lead variant HLA-DRB1\*08:03. The dotted horizontal line represents the genome-wide significance threshold of  $P = 5.0 \times 10^{-8}$ .

and HLA-DP $\beta$ 1 amino acid position 8, no additional MHC associations were observed ( $P > 2.8 \times 10^{-5}$ ).

Given the driver role of the neutralizing anti-GM-CSF antibody in the pathogenesis of aPAP, we finally sought to investigate the association of the risk HLA alleles with serum anti-GM-CSF antibody levels. We observed that the lead risk HLA allele, HLA-DRB1\*08:03, was positively associated with the increased level of anti-GM-CSF antibody (effect size = 0.32,  $P = 0.035$ ) within the cases of aPAP. Although the statistical power in the within-case analysis is limited if we consider the multiple testing, this might suggest that HLA-DRB1\*08:03 is genetically associated with the increased production of anti-GM-CSF antibody, thereby predisposing individuals harboring this allele to the increased risk of developing aPAP.

The allele frequency of HLA-DRB1\*08:03 is common in Asian population (e.g., 8.3% in Japanese, 5.8% in Chinese, and 5.4% in Vietnamese), but very rare in other populations including Europeans (e.g., 0% in British, 0.29% in German, 0% in Italian, and 0.24% in European Caucasian in USA) according to the Allele Frequency Net Database<sup>26</sup>. The frequency heterogeneity was replicated in the UK Biobank resource of global populations<sup>27</sup>. Previous reports regarding the genetic association of HLA-DRB1\*08:03 with human diseases and traits are thus limited to the Asian populations. Chen et al.<sup>28</sup> reported that HLA-DRB1\*08:03 was associated with increased risk of drug-induced agranulocytosis in patients with Graves' disease (OR, 4.36,  $P = 1.8 \times 10^{-9}$ ) in Taiwanese. While not reaching the genome-wide significance threshold, there had been a suggestive association of HLA-DRB1\*08:03 with primary biliary cholangitis in the Japanese population (OR, 1.75,  $P = 1.0 \times 10^{-7}$ )<sup>29</sup>, and one with systemic lupus erythematosus (SLE) in the Korean

population (OR, 1.59,  $P = 7.4 \times 10^{-8}$ )<sup>30</sup>. Intriguingly, although there exists no systematic epidemiological observation of the comorbidity of aPAP and these autoimmune diseases, there have been two case reports of comorbidity of SLE and aPAP<sup>31,32</sup>. We speculated that the Asian specific allele of HLA-DRB1\*08:03 might have a shared etiological basis for abnormal self- or drug-derived antigen presentation and production of autoantibodies.

## Discussion

In this study, we provided evidence of significant genetic contribution to the pathogenesis of aPAP. The strong genetic risk was identified within the MHC region. We further fine mapped the association signals, and revealed that HLA-DRB1\*08:03 significantly drove the genetic risk, followed by an additional risk at HLA-DP $\beta$ 1 amino acid position 8. We finally showed that this allele was also associated with the increased production of anti-GM-CSF antibody, a key driver of the disease. The effect size of the identified genetic risk was much larger than those of any of the previously suggested epidemiological risk factors for aPAP, such as cigarette smoking.

Given its autoimmune etiology, the association of variants in the MHC region with the risk of aPAP has long been expected, while no significant association had been reported to date<sup>33</sup>. Our study provided the first evidence of HLA association in the susceptibility to aPAP with genome-wide significance, which was enabled by a nation-wide collaborative patient recruitment strategy and a clear diagnostic criterion for enrollment<sup>2,5</sup>. This effort led to the recruitment and genotyping of around one fourth of the estimated patients with PAP in Japan. We performed the one-stage GWAS in this study because of the extremely low prevalence of the disease, which hampered the recruitment of completely independent patients as a replication dataset. In addition, the lead allele, HLA-DRB1\*08:03, is Asian-specific, which makes a replication in other populations difficult. Although we mitigated a risk of spurious associations caused by population stratification by conducting stratified analyses based on the geographical location of recruitment centers, a future GWAS in other Asian cohorts is warranted to confirm the replicability of the association. Furthermore, GWAS in other populations would elucidate the comprehensive genetic architecture of the disease across ancestries.

Intriguingly, the lead HLA allele that strongly increased the risk of the disease was common (MAF > 5%) in the Japanese population, despite the extremely low prevalence of the disease. We speculated that the risk variants within HLA class II alleles led to the presentation of self-peptide of GM-CSF to immunocompetent cells. Previously, Uchida et al.<sup>34</sup> reported that anti-GM-CSF antibodies were detected in healthy individuals but in far lower level than in aPAP, and that anti-GM-CSF autoantibodies were associated with impaired GM-CSF-dependent myeloid functions at levels above a critical threshold. Our results showing that the disease risk allele of HLA-DRB1\*08:03 was also associated with an increased level of anti-GM-CSF antibodies that might support the hypothetical pathogenesis of aPAP. Further analyses incorporating rare variants from whole-genome sequencing (WGS) or environmental interactions might decipher the pathogenetic architecture of this disease.

The aPAP is characterized by large variations in disease severity and clinical prognosis. For patients with severe symptoms, approved treatments are limited to whole lung lavage. The clinical outcome of patients with comorbidities is unsatisfactory<sup>35</sup>. Of note, a randomized control trial recently showed a significant effect on the laboratory outcome of inhaled rhGM-CSF (sargramostim) in patients with aPAP but no clinically important changes in outcomes in a limited sample size<sup>15</sup>. The current guidelines for diagnosis and treatment of PAP would have

a room for improvement, by optimization for individual clinical outcome and accounting for novel treatment options<sup>5</sup>. Our study might suggest a potential clinical utility of genetic risk in explaining the differences in the severity or treatment response and in constructing efficient treatment strategies.

## Methods

**Study populations.** Patients with PAP were recruited from major hospitals throughout Japan as a nationwide collaborative project. We enrolled patients with a diagnosis of aPAP based on findings on high-resolution computed tomography (CT) and biopsy, cytologic findings on bronchoalveolar lavage, or both, with a positive serum anti-GM-CSF antibody level ( $>1.0 \mu\text{g}$  per milliliter) (see Supplementary Methods). Serum anti-GM-CSF antibody levels were measured by ELISA as described elsewhere<sup>36</sup>. Control participants were recruited at Osaka University or related institutions (see Supplementary Methods). All participants provided written informed consent approved by the institutional review board of each participating hospital or institution. This study was approved by the ethical committee of Aichi Medical University, the Jikei University School of Medicine, and Osaka University.

**Genotyping and quality control.** Hundred and ninety-eight patients with aPAP and 395 control participants were genotyped using Infinium Asian Screening Array. This genotyping array was built using an East Asian reference panel including WGS, which enabled efficient genotyping in East Asian populations.

For sample quality control (QC), we excluded samples with low genotyping call rates (call rate  $<98\%$ ) and in close genetic relation (PI\_HAT calculated by PLINK<sup>37</sup>  $>0.175$ ). We included samples of the estimated East Asian ancestry, based on the principal component analysis with the samples of HapMap project<sup>38</sup> (Supplementary Fig. 4). We confirmed that no sample had the heterozygosity rate greater than the mean  $+3$  SD of all the individuals. For variant QC, we excluded variants meeting any of the following criteria: (1) genotyping call rate  $<98\%$ , (2)  $P$  value for Hardy–Weinberg equilibrium  $<1.0 \times 10^{-6}$ , and (3) minor allele count  $<5$ . In addition we excluded variants with  $>10\%$  frequency difference with the imputation reference panel. This QC process yielded 498,097 autosomal and 16,207 X chromosomal scaffold variants.

**Whole-genome imputation.** We used shapeit2 software<sup>39</sup> for haplotype phasing with the use of haplotype reference. After the phasing, we used Minimac3 software<sup>40</sup> for genotype imputation. As an imputation reference, we used the reference haplotypes of 1000 Genomes Project Phase 3 version 5 genotype ( $n = 2,504$ ) and Japanese WGS data ( $n = 1037$ )<sup>41</sup>, which were recently constructed and validated for imputation accuracy<sup>42</sup>. We used imputed variants with  $\text{Rsq} > 0.7$  in the association analysis.

**Genome-wide association study.** We performed GWAS by using logistic regression to test single-marker genetic associations with the risk of aPAP on the basis of imputed allelic dosage. We included age, sex, and top 20 principal components as covariates to account for potential population structure. We used genetically determined sex as a covariate. We considered a genome-wide significance threshold in the association as  $P < 5.0 \times 10^{-8}$ . We set a suggestive significance threshold and a significance threshold in a conditional analysis as  $P < 5.0 \times 10^{-7}$ . We assessed the genotyping accuracy of the lead SNP and the suggestive SNP (i.e., rs138024423 at MHC locus and rs56125424 at 4q34) by comparing the SNP-array based genotype with WGS data for part of the control individuals ( $n = 181$ ). We confirmed that the concordance between the SNP-array based genotype and the WGS based genotype was high (99.5% and 100%, respectively).

**HLA imputation and association test.** We imputed classical and non-classical HLA alleles (2-, 4-, and 6-digits) and corresponding amino acid sequences using the reference panel recently constructed from 1120 individuals of Japanese ancestry<sup>25</sup>. Briefly, we encoded all variants in the reference panel as biallelic markers. We extracted the genotyped variants in the entire MHC region (24–36 Mb on chromosome 6, NCBI Build 37). Using these variants as a scaffold, we imputed the classical and non-classical HLA alleles (2-, 4-, and 6-digits) and corresponding amino acid sequences using SNP2HLA software<sup>43</sup>. We applied post-imputation quality control to keep the imputed variants with minor allele frequency (MAF)  $\geq 0.5\%$  and  $\text{Rsq} > 0.7$ . For these binary markers that indicated the presence or absence of an investigated HLA allele or an amino acid sequence, we tested an association with the risk of aPAP, using an additive logistic regression model. We included the same covariates as in the primary GWAS. For each marker, we used probabilistic genotypes that took any uncertainty in imputation into account.

**Sensitivity analysis by stratifying the cohort based on the geographical location of recruitment centers.** We split the study cohort into two datasets, based on the geographical location of recruitment centers as a sensitivity analysis.

The Set 1 included 84 cases from Aichi Medical University and National Hospital Organization Kinki-Chuo Chest Medical Center (located in the west part of Japan) and 252 controls from the west part of Japan. The Set 2 included the remaining 114 cases and 143 controls from the east part of Japan. We then tested the association of HLA-DRB1\*08:03 and rs56125424 with the risk of aPAP within each of the two datasets, using the logistic regression model with the same covariates used for the primary analysis.

**Conditional analysis in the MHC locus.** To investigate secondary association signals, we used the same additive logistic regression model, additionally including an allelic dosage of the primary lead signal as a covariate.

**Quantitative association test for autoantibody levels.** To assess the contribution of each HLA marker to the quantity of serum anti-GM-CSF antibody within the 198 cases, we tested the association of probabilistic genotypes of HLA alleles or amino acid sequences with the normalized value of serum anti-GM-CSF antibody levels, using an additive linear regression model. For the normalization of serum anti-GM-CSF antibody levels, we performed rank-based inverse normal transformation of the residuals, which we obtained from a linear regression model of log-transformed serum anti-GM-CSF antibody levels adjusted for age, sex, and top 20 principal components. The distribution of anti-GM-CSF antibody levels across patients is shown in Supplementary Fig. 5.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

We provide an interactive visualization of a Manhattan plot with downloadable GWAS summary statistics at our pheweb.jp website [<https://pheweb.jp/pheno/PAP>]. The summary statistics are also deposited at the National Bioscience Database Center (NBDC) Human Database with the accession code [hum0197.v2](https://www.nbdb.ac.jp/human/).

Received: 8 April 2020; Accepted: 8 January 2021;

Published online: 15 February 2021

## References

- McCarthy, C., Avetisyan, R., Carey, B. C., Chalk, C. & Trapnell, B. C. Prevalence and healthcare burden of pulmonary alveolar proteinosis. *Orphanet J. Rare Dis.* **13**, 129 (2018).
- Inoue, Y. et al. Characteristics of a large cohort of patients with autoimmune pulmonary alveolar proteinosis in Japan. *Am. J. Respir. Crit. Care Med.* **177**, 752–762 (2008).
- Rosen, S. H., Castleman, B. & Liebow, A. A. Pulmonary alveolar proteinosis. *N. Engl. J. Med.* **258**, 1123–1142 (1958).
- Trapnell, B. C., Whitsett, J. A. & Nakata, K. Pulmonary Alveolar Proteinosis. *N. Engl. J. Med.* **349**, 2527–2539 (2003).
- Kumar, A., Abdelmalak, B., Inoue, Y. & Culver, D. A. Pulmonary alveolar proteinosis in adults: pathophysiology and clinical approach. *Lancet Respiratory Med.* **6**, 554–565 (2018).
- Borie, R. et al. Pulmonary alveolar proteinosis. *Eur. Respiratory Rev.* **20**, 98–107 (2011).
- Kitamura, T. et al. Idiopathic pulmonary alveolar proteinosis as an autoimmune disease with neutralizing antibody against granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **190**, 875–880 (1999).
- Suzuki, T. et al. Familial pulmonary alveolar proteinosis caused by mutations in CSF2RA. *J. Exp. Med.* **205**, 2703–2710 (2008).
- Nogee, L. M. Alterations in SP-B and SP-C expression in neonatal lung disease. *Annu. Rev. Physiol.* **66**, 601–23 (2004).
- Wert, S. E., Whitsett, J. A. & Nogee, L. M. Genetic disorders of surfactant dysfunction. *Pediatr. Dev. Pathol.* **12**, 253–74 (2009).
- Hamvas, A. et al. Heterogeneous pulmonary phenotypes associated with mutations in the thyroid transcription factor gene NKX2-1. *Chest* **144**, 794–804 (2013).
- Lyons, P. A. et al. Genetically distinct subsets within ANCA-associated vasculitis. *N. Engl. J. Med.* **367**, 214–223 (2012).
- Patsopoulos, N. A. et al. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, eaav7188 (2019).
- Terao, C. et al. IgG4-related disease in the Japanese population: a genome-wide association study. *Lancet Rheumatol.* **1**, e14–e22 (2019).
- Tazawa, R. et al. Inhaled GM-CSF for pulmonary alveolar proteinosis. *N. Engl. J. Med.* **381**, 923–932 (2019).
- Kichaev, G. et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).

17. Pulit, S. L. et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
18. Zhu, Z. et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J. Allergy Clin. Immunol.* **145**, 537–549 (2020).
19. Tikkanen, E. et al. Biological insights into muscular strength: genetic findings in the UK Biobank. *Sci. Rep.* **8**, 6451 (2018).
20. Liu, M. et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
21. Karlsson Linnér, R. et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245–257 (2019).
22. Hromatka, B. S. et al. Genetic variants associated with motion sickness point to roles for inner ear development, neurological processes and glucose homeostasis. *Hum. Mol. Genet.* **24**, 2700–2708 (2015).
23. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–6 (2014).
24. Raychaudhuri, S. et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44**, 291–296 (2012).
25. Hirata, J. et al. Genetic and phenotypic landscape of the major histocompatibility complex region in the Japanese population. *Nat. Genet.* **51**, 470–480 (2019).
26. Gonzalez-Galarza, F. F. et al. Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* **48**, D783–D788 (2020).
27. Tanigawa, Y. & Rivas, M. A. Initial review and analysis of COVID-19 host genetics and associated phenotypes. *Preprints* <https://doi.org/10.20944/preprints202003.0356.v1> (2020).
28. Chen, P. L. et al. Genetic determinants of antithyroid drug-induced agranulocytosis by human leukocyte antigen genotyping and genome-wide association study. *Nat. Commun.* **6**, 1–8 (2015).
29. Yasunami, M. et al. Principal contribution of HLA-DQ alleles, DQB1\*06:04 and DQB1\*03:01, to disease resistance against primary biliary cholangitis in a Japanese population. *Sci. Rep.* **7**, 1–10 (2017).
30. Lessard, C. J. et al. Identification of a systemic lupus erythematosus risk locus spanning ATG16L2, FCHSD2, and P2RY2 in Koreans. *Arthritis Rheumatol.* **68**, 1197–1209 (2016).
31. Nagasawa, J. et al. Systemic lupus erythematosus complicating autoimmune pulmonary alveolar proteinosis that was worsened by immunosuppressive therapy. *Lupus* **22**, 1060–1063 (2013).
32. Silva-Díaz, M., González, M. F. & Romero, T. H. Pulmonary alveolar proteinosis in a patient with systemic lupus erythematosus. *J. Rheumatol.* **47**, 779–780 (2020).
33. Anderson, K. et al. Pulmonary alveolar proteinosis: An autoimmune disease lacking an HLA association. *PLoS ONE* **14**, e0213179 (2019).
34. Uchida, K. et al. Granulocyte/macrophage-colony-stimulating factor autoantibodies and myeloid cell immune functions in healthy subjects. *Blood* **113**, 2547–2556 (2009).
35. Akira, M. et al. Pulmonary fibrosis on high-resolution ct of patients with pulmonary alveolar proteinosis. *Am. J. Roentgenol.* **207**, 544–551 (2016).
36. Nishimura, M. et al. Clinical significance of serum anti-GM-CSF autoantibody levels in autoimmune pulmonary alveolar proteinosis. *Biomark. Med.* **12**, 151–159 (2018).
37. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
38. Belmont, J. W. et al. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
39. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
40. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
41. Okada, Y. et al. Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. *Nat. Commun.* **9**, 1631 (2018).
42. Akiyama, M. et al. Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat. Commun.* **10**, 4393 (2019).
43. Jia, X. et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).

## Acknowledgements

We sincerely thank all participants of this study for their invaluable contributions. We thank Prof. Hisashi Arase and Dr. Akihiko Sakoguchi for their critical input on the functional interpretation of HLA associations. This research was supported by the Japan Agency for Medical Research and Development (AMED; JP20ek0109413 [E.Y., Y.I. and Y.O.] and JP19ek0109268 [Y.I., E.Y., T.A., M.H., T.I., S.O., T.T., R.T., K.M., M. Takaki., M.S., K. Tomii., H. Ishii., H.K. and K.N.]) and Bioinformatics Initiative of Osaka University Graduate School of Medicine, Osaka University Center of Medical Data Science. S.S. was in part supported by The Mochida Memorial Foundation for Medical Pharmaceutical Research, Kanai Foundation for the promotion of medical science, and Astellas Foundation for Research on Metabolic Disorders. Y.O. was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (19H01021 and 20K21834), AMED (JP20km0405211, JP20gm4010006, and JP20km0405217), and Takeda Science Foundation.

## Author contributions

E.Y., Y.I. and Y.O. supervised the study. S.S., E.Y., Y.I. and Y.O. wrote the manuscript. S.S., M. Takahashi, J.H., K. Sonehara, and Y.O. conducted data analysis. E.Y., Y.I., J.H., S.I., T.A., M.H., Y.T., T.N., T.I., S.O., T. Hirano, T.T., S.M., S.D., Y.M., T.N., T. Kishikawa, K.O., T. Masuda, K. Yamamoto, R.T., K.M., M. Takaki, S.K., M.S., K. Tomii, A.N., T. Handa, K. Tanizawa, H. Ishii, M.I., T. Kato, N.T., K. Yokomura, T. Matsui, M.W., H. Inoue, K.I., Y.G., H.K., T.F., T.S., T.Y., Y.S., H. Ibata, N.H., H.M., A.K. and K.N. collected the samples. M. Takahashi, J.H., K. Suzuki, F.M., T. Hirota, and M. Tamari constructed the data.

## Competing interests

J.H. is an employee of TEIJIN PHARMA LIMITED. The authors declare that no other conflicts of interest exist.

## Additional information

The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21011-y>.

**Correspondence** and requests for materials should be addressed to Y.O.

**Peer review information** *Nature Communications* thanks Peter Gregersen, Louise Wain, and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

<sup>1</sup>Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. <sup>2</sup>Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan. <sup>3</sup>Center for Data Sciences, Harvard Medical School, Boston, USA. <sup>4</sup>Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA. <sup>5</sup>Program in

Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, USA. <sup>6</sup>Division of Respiratory Medicine and Allergology, Department of Internal Medicine, School of Medicine, Aichi Medical University, Aichi, Japan. <sup>7</sup>Clinical Research Center, National Hospital Organization Kinki-Chuo Chest Medical Center, Sakai, Osaka, Japan. <sup>8</sup>Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. <sup>9</sup>Pharmaceutical Discovery Research Laboratories, TEIJIN PHARMA LIMITED, Hino, Japan. <sup>10</sup>Department of Pulmonary Medicine, Fukushima Medical University, Fukushima, Japan. <sup>11</sup>Department Respiratory Medicine, Tokyo Medical University, Tokyo, Japan. <sup>12</sup>Occupational Health, Graduate School of Medicine, Tohoku University, Miyagi, Japan. <sup>13</sup>Respiratory Medicine, School of Medicine, Tohoku University, Miyagi, Japan. <sup>14</sup>Uonuma Institute of Community Medicine, Niigata University Medical and Dental Hospital, Niigata, Japan. <sup>15</sup>Department of Neurosurgery, Faculty of Medicine, the University of Tokyo, Tokyo, Japan. <sup>16</sup>Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, Suita, Japan. <sup>17</sup>Laboratory of Immune Regulation, Department of Microbiology and Immunology, Osaka University Graduate School of Medicine, Suita, Japan. <sup>18</sup>Department of Otorhinolaryngology - Head and Neck Surgery, Osaka University Graduate School of Medicine, Suita, Japan. <sup>19</sup>Department of Neurology, Osaka University Graduate School of Medicine, Suita, Japan. <sup>20</sup>Department of Obstetrics and Gynecology, Osaka University Graduate School of Medicine, Suita, Japan. <sup>21</sup>Department of Pediatrics, Osaka University Graduate School of Medicine, Suita, Japan. <sup>22</sup>Student Support and Health Administration Organization, Tokyo Medical and Dental University, Tokyo, Japan. <sup>23</sup>Department of Clinical Medicine, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan. <sup>24</sup>Department of Infectious Diseases, Nagasaki University Hospital, Nagasaki University, Nagasaki, Japan. <sup>25</sup>Department of Respiratory Medicine, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo, Japan. <sup>26</sup>Department of Respiratory Medicine, Kobe City Medical Center General Hospital, Kobe, Japan. <sup>27</sup>Department of Advanced Medicine for Respiratory Failure, Graduate School of Medicine, Kyoto University, Kyoto, Japan. <sup>28</sup>Department of Respiratory Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan. <sup>29</sup>Department of Respiratory Medicine, Kyorin University, Mitaka, Japan. <sup>30</sup>Department of Respiratory Medicine and Allergology, Kariya Toyota General Hospital, Kariya, Japan. <sup>31</sup>Department of Respiratory Medicine, Respiratory Disease Center, Seirei Mikatahara General Hospital, Hamamatsu, Japan. <sup>32</sup>Department of Pulmonary Medicine, Graduate School of Medical & Dental Sciences, Kagoshima University, Kagoshima, Japan. <sup>33</sup>Department of Respiratory Medicine, Fujita Health University School of Medicine, Aichi, Japan. <sup>34</sup>Department of Respiratory Medicine, National Hospital Organization Osaka Toneyama Medical Center, Toyonaka, Japan. <sup>35</sup>Second Division, Department of Internal Medicine, Hamamatsu University School of Medicine, Hamamatsu, Japan. <sup>36</sup>Department of Respiratory Medicine, Shizuoka City Shizuoka Hospital, Shizuoka, Japan. <sup>37</sup>Department of Respiratory Medicine, National Hospital Organization Mie Chuo Medical Center, Tsu, Japan. <sup>38</sup>Department of Pulmonary Medicine, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. <sup>39</sup>Laboratory of Immunopathology, World Premier International Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan. <sup>40</sup>Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Japan. <sup>41</sup>Division of Advanced Medical Development, Niigata University Medical and Dental Hospital, Niigata, Japan. <sup>42</sup>Division of Molecular Genetics, the Jikei University School of Medicine, Research Center for Medical Science, Tokyo, Japan. <sup>43</sup>Laboratory of Statistical Immunology, World Premier International Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan. <sup>44</sup>These authors contributed equally: S. Sakae, E. Yamaguchi. <sup>✉</sup>email: [yokada@sg.med.osaka-u.ac.jp](mailto:yokada@sg.med.osaka-u.ac.jp)