Genetic determinants of risk in autoimmune pulmonary alveolar proteinosis

Saori Sakaue1,2,3,4,44, Etsuro Yamaguchi6,44, Yoshikazu Inoue7, Meiko Takahashi8, Jun Hirata1,9, Ken Suzuki1, Satoru Ito6, Toru Ara7, Masaki Hirose7, Yoshinori Tanino10, Takefumi Nikaido10, Toshio Ichiwata11, Shinya Ohkouchi12, Taizou Hirano13, Toshinori Takada14, Satoru Miyawaki15, Shogo Dofuku15, Yuichi Maeda16,17, Takuro Ni16,17, Toshihiro Kishikawa18, Kotaro Ogawa19, Tatsuo Masuda1,20, Kenichi Yamamoto1,21, Kyuto Sonohara1, Ryuishi Tazawa22, Konosuke Morimoto23, Masahiro Takaki24, Satoshi Konno25, Masaru Suzuki25, Keisuke Tomii26, Atsushi Nakagawa26, Tomohiro Handa27, Kiminori Tanizawa28, Haruyuki Ishii29, Manabu Ishida29, Toshiyuki Kato30, Naoya Takeda30, Koshi Yokomura31, Takashi Matsui31, Masaki Watanabe32, Hiromasa Inoue32, Kazuyoshi Imaizumi33, Yasuhiro Goto33, Hiroshi Kida16,34, Tomoyuki Fujisawa35, Takaumi Suda35, Takashi Yamada36, Yasuomi Satake36, Hidenori Ibata37, Nobuyuki Hizawa38, Hideki Mochizuki19, Atsushi Kumanogoh16,39,40, Fumihiko Matsuda8, Koh Nakata41, Tomomitsu Hirota42, Mayumi Tamari42 & Yukinori Okada1,40,43

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-DPB1 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.
Pulmonary alveolar proteinosis (PAP) is a rare diffuse lung disease with a prevalence of 6–7 cases per million population worldwide. The disease is characterized by the abnormal accumulation of pulmonary surfactant within pulmonary alveoli, resulting in progressive respiratory failure and increased infection risk. PAP has three distinct etiologies: hereditary, autoimmune, and secondary. 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, thereby causing poor surfactant clearance. While rare mutations in the GM-CSF receptor α or β chains (CSF2RA, CSF2RB), the surfactant proteins B or C, ATP-binding cassette subfamily A member 3 (ABCA3), and thyroid transcription factor-1 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, we analyzed 12,153,232 genetic markers that passed the genome-wide significance threshold to maximize the statistical power, given the low prevalence of the disease. We performed a one-stage single-marker association test and confirmed that a suggestive signal at 4q34 had similar effect sizes worldwide. The loci that satisfied the genome-wide significance threshold are colored in red. The loci that satisfied the genome-wide significance threshold are colored in red. MHC, major histocompatibility complex.

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 nationwide collaborative recruitment strategy. 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 (Rsq > 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. 1a). We calculated the genome-wide significance threshold of $P = 5.0 \times 10^{-8}$ to identify the genome-wide association (GWAS) of the genome-wide single-marker association test of aPAP in the Japanese population (Fig. 1a). We observed that the effect sizes of both datasets were consistent with the primary association ($OR = 4.8\times 10^{-8}$). 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).

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   | 69.3 ± 80.2            | N/A                           |
| GM-CSF—µg/ml              |                        |                               |

Plus-minus values are means ± standard deviation.

MHC risk fine-mapping by HLA imputation. To fine map the disease risk within the complex MHC region, we applied HLA imputation method to the genotype data using a high-resolution reference panel of 1120 individuals of Japanese ancestry. 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-DPB1 (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. 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.
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-DPB1 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. 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. 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. 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. 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. The current guidelines for diagnosis and treatment of PAP would have population (OR, 1.59, \( P = 7.4 \times 10^{-8} \)). 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. 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**

The high 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-DPB1 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. 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. 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. 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. 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. The current guidelines for diagnosis and treatment of PAP would have

---

**Figure 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 \(P = 5.0 \times 10^{-5}\). 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 \(P = 5.0 \times 10^{-5}\).
a room for improvement, by optimization for individual clinical outcome and accounting for novel treatment options. 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 pAP 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 μg per milliliter) (see Supplementary Methods). Serum anti-GM-CSF antibody levels were measured by ELISA as described elsewhere. 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 pAP 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 (<80%), including those in close genetic relation (PLINK r2h calculated by PLINK > 0.175). We included samples of the estimated East Asian ancestry, based on the principal component analysis with the samples of HapMap project (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 by the following criteria: (1) genotyping call rate < 98%, (2) P value for Hardy–Weinberg equilibrium < 1.0 × 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 snapshot2 software for haplotype phasing with the use of haplotype reference. After the phasing, we used Minimac3 software for genomewide 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) which were recently constructed and validated for imputation accuracy. We used imputed variants with 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 pAP 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 basis of imputed allelic dosage. We included age, sex, and top 20 principal components as covariates to account for potential population structure. We used the same covariates as in the primary GWAS. For each tested an association with the risk of aPAP, using an additive logistic 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.

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 and 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.

Data availability

We provide an interactive visualization of a Manhattan plot with downloadable GWAS summary statistics at our pheweb.ip website [https://pheweb.ip/phenPAP]. The summary statistics are also deposited at the National Bioscience Database Center (NBDC) Human Database with the accession code hum0197.v2.

Received: 8 April 2020; Accepted: 8 January 2021;
Published online: 15 February 2021

References

1. McCarthy, C., Aventisyan, R., Carey, B. C., Chalk, C. & Trapnell, B. C. Prevalence and healthcare burden of pulmonary alveolar proteinosis. Orphanet J. Rare Dis. 13, 129 (2018).
2. 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).
3. Rosen, S. H., Castleman, B. & Liebow, A. A. Pulmonary alveolar proteinosis. N. Engl. J. Med. 258, 1123–1142 (1958).
4. Trapnell, B. C., Whitsett, J. A. & Nakata, K. Pulmonary Alveolar Proteinosis. N. Engl. J. Med. 349, 2522–2530 (2003).
5. Kumar, A., Abdelnour, S., Inoue, Y. & Culver, D. A. Pulmonary alveolar proteinosis in adults: pathophysiology and clinical approach. Lancet Respiratory Med. 6, 554–565 (2018).
6. Borie, R. et al. Pulmonary alveolar proteinosis. Eur. Respir. Rev. 20, 98–107 (2011).
7. 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).
8. Suzuki, T. et al. Familial pulmonary alveolar proteinosis caused by mutations in CSF2RA. J. Exp. Med. 205, 2703–2710 (2008).
9. Nogee, L. M. Alterations in SP-B and SP-C expression in neonatal lung and mouse models. Annu. Rev. Physiol. 71, 569–593 (2009).
10. Wert, S. E., Whitsett, J. A. & Nogee, L. M. Genetic disorders of surfactant dysfunction. Pediatr. Dev. Pathol. 12, 253–74 (2009).
11. Hamvas, A. et al. Heterogeneous pulmonary phenotypes associated with mutations in the thyroid transcription factor gene NKX2-1. Chest 144, 794–804 (2013).
12. Lyons, P. A. et al. Geneticly distinct subsets within ANCA-associated vasculitis. N. Engl. J. Med. 367, 214–223 (2012).
13. Patsopoulos, N. A. et al. Multiple sclerosis genome map implicates peripheral immune cells and microglia in susceptibility. Science 365, eea27188 (2019).
14. Terao, C. et al. IgG4-related disease in the Japanese population: a genome-wide association study. Lancet Rheumatol. 1, e14–e22 (2019).
15. Tazawa, R. et al. Inhaled GM-CSF for pulmonary alveolar proteinosis. N. Engl. J. Med. 381, 932–933 (2019).
16. Kichaga, 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**, 164–174 (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 associations of 169,741 individuals. *Nat. Genet.* **53**, 237–244 (2019).
20. 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).
21. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–6 (2014).
22. 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).
23. Hirata, J. et al. Genetic and phenotypic landscape of the major histocompatibility complex region in the Japanese population. *Nat. Genet.* **47**, 470–478 (2019).
24. 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).
25. 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).
26. Chen, P. L. et al. Genetic determinants of anti-thyroid drug-induced agranulocytosis by human leukocyte antigen genotyping and genome-wide association study. *Nat. Commun.* **6**, 1–8 (2015).
27. Yasunami, M. et al. Principal contribution of HLA-DQ alleles, DQB1*06:04 and DQB1*06:01, to disease resistance against primary biliary cholangitis in a Japanese population. *Sci. Rep.* **7**, 1–10 (2017).
28. Leonard, C. J. et al. Identification of a systemic lupus erythematosus risk locus spanning ATG16L2, FCHS2D, and P2RY2 in Koreans. *Arthritis Rheumatol.* **68**, 1197–1209 (2016).
29. Nagasawa, J. et al. Systemic lupus erythematosus complicating autoimmune pulmonary alveolar proteinosis that was worsened by immunosuppressive therapy. *Lupus* **22**, 1060–1063 (2013).
30. Silva-Diaz, M., Gonzalez, M. F. & Romero, T. H. Pulmonary alveolar proteinosis in a patient with systemic lupus erythematosus. *J. Rheumatol.* **47**, 779–780 (2020).
31. Anderson, K. et al. Pulmonary alveolar proteinosis: An autoimmune disease lacking an HLA association. *PloS ONE* **14**, e0213179 (2019).
32. Uchida, K. et al. Granulocyte/macrophage-colony-stimulating factor autoantibodies and myeloid cell immune functions in healthy subjects. *Blood* **113**, 2547–2556 (2009).
33. Akira, M. et al. Pulmonary fibrosis on high-resolution CT of patients with pulmonary alveolar proteinosis. *Am. J. Roentgenol.* **207**, 544–551 (2016).
34. Nishimura, M. et al. Clinical significance of serum anti-GM-CSF autoantibody levels in autoimmune pulmonary alveolar proteinosis. *Biomark. Med.* **12**, 151–159 (2018).
35. 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).
36. Belmont, J. W. et al. A haplotype map of the human genome. *Nature* **437**, 1299–1302 (2005).
37. Delanauve, O., Marchini, J. & Zagni, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
38. Dass, S. et al. Next-generation genotyping assay development and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
39. Oikawa, Y. et al. Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. *Nat. Commun.* **9**, 1631 (2018).
40. Akiyama, M. et al. Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat. Commun.* **10**, 4393 (2019).
41. Iia, 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. Hiashi 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., T. Takeda, 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, Kaneo 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, JP20gm0410006, and JP20km0405217), and Takeda Science Foundation.

**Author contributions**

E.Y., Y.I. and Y.O. supervised the study. S.S., E.Y. 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. Hiranot, 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., Y. Yokomura, T. Matsui, M.W., H. Inoue, K.Y., H.K., T.F., T.S., Y.S., H. Ibata, N.H., H.M., A.K. and K.N. collected the samples. M. Takashahi, 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.

© The Author(s) 2021
