Received: 29 November 2020 | Revised: 11 March 2021 | Accepted: 28 March 2021

DOI: 10.1002/jcla.23782

RESEARCH ARTICLE

WILEY

# A nonsynonymous polymorphism (rs117179004, T392M) of hyaluronidase 1 (*HYAL1*) is associated with increased risk of idiopathic pulmonary fibrosis in Southern Han Chinese

Juan Liu[1] | Yanhan Deng[2] | Zheng Wang[3] | Biwen Mo[4] | Jianghong Wei[4] | Zhenshun Cheng[5] | Qingzhen Peng[6] | Guang Wei[6] | Jingping Li[7] | Ying Shu[7] | Hua Yang[8] | Shirong Fang[8] | Guangwei Luo[9] | Shuo Yang[9] | Yingnan Wang[10] | Jing Zhu[10] | Jingping Yang[11] | Ming Wu[11] | Xuyan Xu[12] | Renying Ge[12] | Xiaoju Zhang[3] | Weining Xiong[1,13] | Xiaomei Wang[14] | Zongzhe Li[15]

[1]Key Laboratory of Pulmonary Diseases of Health Ministry, Key Cite of National Clinical Research Center for Respiratory Disease, Department of Respiratory and Critical Care Medicine, Wuhan Clinical Medical Research Center for Chronic Airway Diseases, Tongji Hospital, Tongji Medical College, Huazhong University of Sciences and Technology, Wuhan, China

[2]Departments of Rheumatology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Respiratory Medicine, Henan Provincial People's Hospital & The People's Hospital of Zhengzhou University, Zhengzhou, China

[4]Department of Respiratory Medicine, Affiliated Hospital of Guilin Medical University, Guilin, China

[5]Department of Respiratory Medicine, Zhongnan Hospital of Wuhan University, Wuhan University, Wuhan, China

[6]Department of Respiratory Medicine, Xiaogan Central Hospital, Xiaogan, China

[7]Department of Respiratory Medicine, Qianjiang Central Hospital, Qianjiang, China

[8]Department of Respiratory Medicine, University Hospital of Hubei University for Nationalities, Enshi, China

[9]Department of Respiratory Medicine, Wuhan No. 1 Hospital, Wuhan, China

[10]Department of Respiratory and Critical Care Medicine, Renmin Hospital of Three Gorges University, Yichang, China

[11]Department of Respiratory and Critical Care Medicine, The Third Affiliated Hospital of Inner Mongolia Medical University, Baotou, China

[12]Department of Respiratory Medicine, Xianning Center Hospital, The First Affiliated Hospital of Hubei University of Science and Technology, Xianning, China

[13]Department of Respiratory and Critical Care Medicine, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[14]Department of Geriatrics, The First Affiliated Hospital of Army Medical University, Chongqing, China

[15]Department of Internal Medicine and Genetic Diagnosis Center, Tongji Hospital, Tongji Medical College, Huazhong University of Sciences and Technology, Wuhan, China

**Correspondence**
Zongzhe Li, Department of Internal Medicine and Genetic diagnosis Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430030, China.
Email: marklee0208@163.com

Xiaomei Wang, Department of Geriatrics, The First Affiliated Hospital of Army Medical University, Chongqing, 400038, China.
Email: wxm091@163.com

## Abstract

**Background:** Idiopathic pulmonary fibrosis (IPF) is a genetic heterogeneous disease with high mortality and poor prognosis. Hyaluronidase 1 (*HYAL1*) was found to be up-regulated in fibroblasts from IPF patients, and overexpression of *HYAL1* could prevent human fetal lung fibroblast proliferation. However, the genetic correlation between the *HYAL1* and IPF or connective tissue diseases related interstitial lung disease (CTD-ILD) has not been determined.

**Methods:** A two-stage study was conducted in Southern Han Chinese population. We sequenced the coding regions and flanking regulatory regions of *HYAL1* in stage one

(253 IPF cases and 125 controls). A statistically significant variant was further genotyped in stage two (162 IPF cases, 182 CTD-ILD cases, and 225 controls).

**Results:** We identified a nonsynonymous polymorphism (rs117179004, T392M) significantly associated with increased IPF risk (dominant model: OR = 2.239, 95% CI = 1.212–4.137, $p = 0.010$ in stage one; OR = 2.383, 95% CI = 1.376–4.128, $p = 0.002$ in stage two). However, we did not observe this association in CTD-ILD (OR = 1.401, 95% CI = 0.790–2.485, $p = 0.248$).

**Conclusion:** Our findings suggest that the nonsynonymous polymorphism (rs117179004, T392M) may confer susceptibility to IPF in Southern Han Chinese, but is not associated with susceptibility to CTD-ILD.

**KEYWORDS**
Chinese, hyaluronidase 1, idiopathic pulmonary fibrosis, polymorphism

## 1 | INTRODUCTION

Interstitial lung disease (ILD) is a heterogeneous group of disorders that diffusely affects the lung parenchyma while having variable etiologies, clinical presentations, radiographic patterns, and histological appearances.[1] ILD causes an irreversible architectural distortion and then impairs gas exchange. A group of ILDs presents with underlying connective tissue diseases (CTD), including systemic sclerosis (SSc), rheumatoid arthritis (RA), and inflammatory myositis, and these types of ILD are referred to CTD-ILD.[2]

Idiopathic pulmonary fibrosis (IPF), a specific form of ILD characterized by pulmonary fibrosis or progressive alveolar interstitial lesions with an unknown cause, occurs primarily in elderly people. Associated with a poor prognosis, the median survival for patients affected by IPF varies from 2 to 5 years, and the patients exhibit variable disease courses and prognoses.[3]

Although the direct mechanism underlying IPF is not completely understood, a genetic predisposition has been considered one of the important causes of this disease.[4] In recent years, there has been growing evidence that genetic factors play an important role in both sporadic and familial IPF cases. Recent independent studies have shown that up to 20% of IPF patients have a family history and can present earlier, indicating that both the frequency of familial pulmonary fibrosis and the genetic risk of sporadic IPF could be underestimated.[5] Current data suggest that at least one third of the sporadic and familial IPF can be explained by common genetic variants identified in large GWASs; meanwhile, some of the allele frequency differ across populations, and some associated with disease prognosis or response to treatment.[6]

As a structural role in the extracellular matrix, hyaluronan (HA) in the development of inflammatory diseases has been emphasized.[7] The concentration of HA in the lungs increases several fold in bleomycin-induced and radiation pulmonary fibrosis, which is attributed to parenchymal damage and inflammation.[8] The hyaluronidases (HYALs) are a group of enzymes that regulate HA metabolism and consequently remodel the extracellular matrix.[9] In the bleomycin lungs, HYAL treatment potently blocked lung injury while decreasing transforming growth factor (TGF)-β production and collagen deposition and thus prevents the development of fibrosis.[8] Hyaluronidase1 (HYAL1) is one of the hyaluronidases that involve in the degradation of HA. And we found that the expression of *HYAL1* was upregulated in fibroblasts from IPF patients (Gene Expression Omnibus accession number: GSE48149). In addition, recent study showed that *HYAL1* expression level in IPF fibroblasts was significantly upregulated at the mRNA level, but not altered at the protein level.[10] In vitro, overexpression of *HYAL1* could prevent human fetal lung fibroblast HFL-1 cell line from fibroproliferation.[10] Thus, *HYAL1* is a biological candidate gene in the development of IPF.

In this study, we carried out a two-step design study to explore the correlation between the *HYAL1* gene and IPF. We first sequenced the coding regions and flanking regulatory regions of *HYAL1* in 253 IPF cases and 125 controls. A statistically significant variant was further genotyped in 162 IPF cases, 182 CTD-ILD cases, and 225 controls.

## 2 | MATERIALS AND METHODS

### 2.1 | Subjects

This study included 415 IPF patients (253 in stage one, 162 in stage two), 182 CTD-ILD patients (stage two), and 350 healthy control subjects (125 in stage one, 225 in stage two). Participants were consecutively recruited from Department of Respiratory and Critical Care Medicine, Tongji Hospital, mainly between March 2014 and December 2018. Clinical data including age, gender, ethnicity, family history, smoking status, medical history, occupational exposure history, physical examination findings, and laboratory results were collected. The identification of IPF was based on evidence-based guidelines, which included the exclusion of other known causes of ILD, the presence of a usual interstitial pneumonia pattern on high-resolution computed tomography (HRCT) in patients not subjected to surgical lung biopsy, specific combinations of HRCT and surgical lung biopsy findings in patients subjected to surgical lung biopsy, and abnormalities of lung function tests.[11] CTD-ILD diagnoses included rheumatologic and pulmonary evaluation. Subjects who suffered from acute inflammation, tuberculosis, or cancers were excluded. The control subjects were

selected from a general health check-up program, and none of them had any clinical evidences of pulmonary diseases, or any other severe diseases. All subjects were unrelated and southern Han Chinese. Each participant provided written informed consent and a peripheral blood sample. This study was approved by the ethics review committee of the Medical Research Institute, Tongji Hospital, Tongji Medical School, Huazhong University of Science and Technology.

## 2.2 | Sequencing and genotyping

Genomic deoxyribonucleic acid (gDNA) was extracted from peripheral blood samples by a Blood DNA kit (TIANGEN BIOTECH). The concentration of each DNA sample was measured using a NanoDrop 2000 Spectrophotometer (NanoDrop Technologies) and then diluted to 5 ng/μl. Then, gDNA samples were amplified and sequenced on the Ion Torrent platform (Thermo Fisher), and data processing and bioinformatics analysis were performed as we previously described.[12,13] Based on resequencing results, one significant single nucleotide polymorphism (SNP) rs117179004 was found to be statistically associated with IPF risk.

To fine-map the association signal, the genotype of rs117179004 was further evaluated by direct sequencing in stage two (in 162 IPF cases, 182 CTD-ILD cases, and 225 controls). gDNA was amplified using the following primers for rs117179004 (Forward, 5′-TCCTTATGCCACTATTCCAG-3′; reverse, 5′-AGACCCTGACTTACCCTTTC-3′). PCR conditions were as follows: 35 cycles of 94°C for 30 s, 58°C for 35 s, and 72°C for 30 s and 1 cycle of 72°C for 5 min to terminate the reaction. PCR products were sequenced using the Big Dye v.1.1 terminator cycle sequencing kit and an Applied Biosystems 3500xl capillary sequencer (Applied Biosystems).

## 2.3 | Statistical analysis

The chi-square test was used to compare the clinical categorical variables between cases and controls, and average age was evaluated by independent-sample *t*-test. Multiple logistic regression models (codominant1, codominant2, dominant, recessive, additive, and allele models) were performed to determine odds ratios (OR), 95% confidence intervals (CI), and *p* values using SPSS 19.0. $p < 0.05$ was considered statistically significant.

Hardy-Weinberg equilibrium (HWE) was calculated for controls using the goodness-of-fit chi-squared test. False-positive report probability (FPRP) analysis was conducted for the significant findings. We set 0.2 as FPRP threshold and adopt a prior probability of 0.1 to detect OR of 1.50/0.67 (risk/protective effects) as described previously.[14–16] The association that reached the FPRP threshold of <0.2 was considered noteworthy. Both HWE and FPRP analyses were performed with a R package for genetic analysis (gap 1.2.2).[17]

**TABLE 1** Clinical characteristics in IPF, CTD-ILD, and control subjects

| | Stage 1 | | | Stage 2 | | | | |
|---|---|---|---|---|---|---|---|---|
| Variables | IPF (*n* = 253) | Controls (*n* = 125) | p | IPF (*n* = 162) | p | CTD-ILD (*n* = 182) | p | Controls (*n* = 225) |
| Age (years) | 65.4 ± 11.1 | 65.3 ± 10.8 | 0.942 | 64.2 ± 10.9 | 0.002 | 56.2 ± 11.6 | 0.001 | 60.9 ±10.5 |
| Gender | | | | | | | | |
| Male | 169 (66.8%) | 84 (67.2%) | 0.938 | 99 (61.1%) | 0.758 | 109 (59.9%) | 0.945 | 134 (59.6%) |
| Female | 84 (33.2%) | 41 (32.8%) | | 63 (38.9%) | | 73 (40.1) | | 91 (40.4%) |
| Smoking status | | | | | | | | |
| Never | 155 (61.3%) | 78 (62.4%) | 0.831 | 104 (64.2%) | 0.435 | 122 (67.0%) | 0.836 | 153 (68.0%) |
| Ever | 98 (38.7%) | 47 (37.6%) | | 58 (35.8%) | | 60 (33.0%) | | 72 (32.0%) |
| Cough | 247 (97.6%) | 0 | | 152 (93.8%) | | 64 (35.2%) | | 0 |
| Chronic exertional dyspnea | 142 (56.1%) | 0 | | 90 (55.6%) | | 38 (20.9%) | | 0 |
| Finger clubbing | 74 (29.2%) | 0 | | 46 (28.4%) | | 18 (9.8%) | | 0 |
| Bibasilar inspiratory crackles | 137 (54.2%) | 0 | | 82 (50.6%) | | 51 (31.3%) | | 0 |
| Pulmonary function test | | | | | | | | |
| FVC% pred | 75.2 (28.5–122.6) | – | | 75.0 (31.2–112.6) | | 76.6 (38.2–123.8) | | – |
| DLCO% pred | 55.4 (15.2–85.6) | – | | 54.5 (15.2–89.5) | | 58.3 (19.3–91.5) | | – |

*Note:* Data are shown in mean ± SD, *n* (%) or mean (range); for IPF and CTD-ILD cases, age means onset age; FVC % pred: percent predicted forced vital capacity; DLCO% pred: percent predicted diffusion capacity for carbon monoxide.

Abbreviations: IPF, idiopathic pulmonary fibrosis, CTD-ILD, connective tissue disease associated interstitial lung disease; n, number of subjects.

**TABLE 2** Identified SNPs by targeted sequencing *HYAL1* in 253 IPF patients and 125 control subjects

| SNP | N (IPF/ Control) | Allele Ref/ Alt | Function | Frequency ESP | 1000g_All | 1000g_ Eas | ExAC_All | ExAC_Eas |
|---|---|---|---|---|---|---|---|---|
| rs117179004 | 59/15 | G/A | Nonsynonymous | $3.000 \times 10^{-4}$ | 0.013 | 0.065 | 0.006 | 0.073 |
| rs782313024 | 1/0 | T/C | Nonsynonymous | – | – | – | $3.300 \times 10^{-5}$ | $3.000 \times 10^{-4}$ |
| rs150255984 | 1/0 | C/T | Nonsynonymous | $2.000 \times 10^{-4}$ | $1.997 \times 10^{-4}$ | 0.001 | $2.476 \times 10^{-5}$ | 0 |
| rs116482870 | 2/0 | C/T | Nonsynonymous | 0.048 | 0.027 | 0.002 | 0.051 | 0.001 |
| rs139187462 | 1/0 | G/A | Synonymous | $7.700 \times 10^{-5}$ | – | – | $1.650 \times 10^{-5}$ | 0 |
| rs138951582 | 6/3 | G/A | Nonsynonymous | $2.000 \times 10^{-4}$ | 0.003 | 0.013 | 0.002 | 0.017 |
| rs587709776 | 1/0 | A/G | Synonymous | – | $5.990 \times 10^{-4}$ | 0.003 | $4.956 \times 10^{-5}$ | $7.000 \times 10^{-4}$ |
| rs781878519 | 0/1 | C/A | Synonymous | – | -s | – | $6.633 \times 10^{-5}$ | $1.000 \times 10^{-4}$ |
| rs202067357 | 0/1 | G/A | Nonsynonymous | – | $3.994 \times 10^{-4}$ | – | $2.000 \times 10^{-4}$ | 0.001 |
| rs141770421 | 0/1 | G/A | Nonsynonymous | $7.700 \times 10^{-5}$ | – | – | $5.011 \times 10^{-5}$ | 0 |
| rs782142144 | 1/0 | G/A | Intronic | – | – | – | $3.508 \times 10^{-5}$ | $2.000 \times 10^{-4}$ |
| rs587672526 | 1/0 | C/T | Intronic | – | $9.984 \times 10^{-4}$ | – | $7.549 \times 10^{-5}$ | 0 |
| rs782454520 | 1/0 | C/T | Intronic | – | – | – | $2.033 \times 10^{-5}$ | 0 |
| rs782360909 | 0/1 | T/A | Intronic | – | – | – | $2.134 \times 10^{-5}$ | $3.000 \times 10^{-4}$ |
| rs587620179 | 0/1 | C/T | Intronic | – | $1.997 \times 10^{-4}$ | 0.001 | $8.505 \times 10^{-5}$ | 0.001 |

Abbreviations: -, data were unavailable; 1000g_All, 1000 Genomes Project for all population groups; 1000g_Eas, 1000 Genomes Project for East Asian; Alt, alternative allele; ESP, exome sequencing project; ExAC_All, The Exome Aggregation Consortium for all population groups; ExAC_Eas, The Exome Aggregation Consortium for East Asian; IPF, idiopathic pulmonary fibrosis; Ref, reference allele.

**TABLE 3** Genotype and allele frequencies of rs117179004 in IPF and control subjects in Stage 1

| Type | IPF *n* (%) | Control *n* (%) | Models | IPF vs. Control OR (95%CI) | *p* | OR (95%CI)[a] | *p*[a] |
|---|---|---|---|---|---|---|---|
| Genotype | | | | | | | |
| GG | 194 (76.7%) | 110 (88.0%) | Codominant 1 | **2.041 (1.100–3.787)** | **0.024** | **2.047 (1.103–3.801)** | **0.023** |
| GA | 54 (21.3%) | 15 (12.0%) | Codominant 2 | – | – | – | – |
| AA | 5 (2.0%) | 0 (0.0%) | Dominant | **2.230 (1.208–4.117)** | **0.010** | **2.239 (1.212–4.137)** | **0.010** |
| | | | Recessive | – | – | – | – |
| $P_{HWE}$ | | 0.475 | Additive | – | – | – | – |
| Allele | | | | | | | |
| G | 442 (87.4) | 235 (94.0%) | | 1 | | 1 | |
| A | 64 (12.6%) | 15 (6.0%) | | **2.268 (1.265–4.068)** | **0.006** | **2.277 (1.269–4.085)** | **0.003** |

*Note:* Codominant 1: GA vs. GG, codominant 2: AA vs. GG, dominant: GA +AA vs. GG, recessive: AA vs. GG +GA, additive: GG vs. GA vs. AA, allele: A vs. G.

Abbreviations: CI, confidence interval; HWE, Hardy-Weinberg equilibrium; IPF, idiopathic pulmonary fibrosis; n, number of subjects; OR, odd ratio; vs., versus.

[a]Adjusted for gender, age, and smoking status; the significant results were in bold.

## 3 | RESULTS

### 3.1 | Baseline characteristics

A total of 253 IPF patients and 125 matched controls were included in stage one (Table 1). No significant difference was found in age (65.4 vs. 65.3 years), sexual proportion (66.8% vs. 67.2%), or smoking history (38.7% vs. 37.6%) between the cases and controls. In stage two, 162 IPF cases, 182 CTD-ILD cases, and 225 controls were included. Similarly, except for age, no significant difference was found in sexual proportion or smoking history between IPF cases or CTD-ILD cases and controls (Table 1), for CTD-ILD cases had relatively earlier age of onset.

### 3.2 | Variants detected in the *HYAL1* gene by targeted sequencing

Variant screening of *HYAL1* gene in IPF patients and healthy controls was sequenced using Ion Torrent semiconductor sequencing as previously described. Seven nonsynonymous variants, three synonymous variants. and five intronic variants were detected (Table 2). One distinct nonsynonymous variant (rs117179004) was observed in 59 IPF patients and in 15 healthy controls, respectively. Most of the variants were heterozygous, with fairly low frequency in multiple databases (Table 2).

### 3.3 | Association analysis of rs117179004 with the risk of IPF or CTD-ILD

In stage one, genotype frequencies of rs117179004 (GG: GA: AA) in the IPF group and control group were 76.7:21.3:2.0 and 88.0:12.0:0.0, respectively (Table 3). In codominant1 model (GA vs. GG), GA genotype frequencies were significantly different between the IPF and control groups (OR = 2.041, 95% CI = 1.100–3.787, $p$ = 0.024). In dominant model (GA + AA vs. GG), frequencies of GA and AA genotypes were significantly different between the IPF and control groups (OR = 2.230, 95% CI = 1.208–4.117, $p$ = 0.010). After adjusting for gender, age, and smoking status, the results remained significance (OR = 2.239, 95% CI = 1.212–4.137, $p$ = 0.010). The A allele frequency of rs117179004 was also showed significant association (crude: OR = 2.268, 95% CI = 1.265–4.068, $p$ = 0.006; adjusted: OR = 2.277, 95% CI = 1.269–4.085, $p$ = 0.003) (Table 3).

To further verify the association of rs117179004 and IPF risk, and explore whether there is a correlation between rs117179004 and CTD-ILD, direct sequencing was performed in 162 IPF cases, 182 CTD-ILD cases, and 225 controls (Figure 1). The results showed replicated association of rs117179004 and IPF risk in codominant 1 model (crude: OR = 2.291, 95% CI = 1.322–3.969, $p$ = 0.003; adjusted: OR = 2.348, 95% CI = 1.342–4.110, $p$ = 0.003), dominant model (crude: OR = 2.325, 95% CI = 1.355–3.989, $p$ = 0.002; adjusted: OR = 2.383, 95% CI = 1.376–4.128, $p$ = 0.002), additive model (crude: OR = 2.208, 95% CI = 1.323–3.685, $p$ = 0.002; adjusted: OR = 2.255, 95% CI = 1.339–3.797, $p$ = 0.002), and allele model (crude: OR = 2.183, 95% CI = 1.320–3.613, $p$ = 0.002; adjusted: OR = 2.228, 95% CI = 1.336–3.716, $p$ = 0.002) (Table 4, Figure 1). However, genotype and allele distributions of rs117179004 were not statistically different between the CTD-ILD and control groups (Table 5, Figure 1). No deviations from Hardy-Weinberg equilibrium were found in the control group ($P_{HWE}$ = 0.475 in stage 1, $P_{HWE}$ = 0.883 in stage 2). Genetic analysis was performed by logistic regression analysis.

### 3.4 | False-positive report probability analysis

We preset 0.2 as the threshold of false-positive report probability (FPRP).

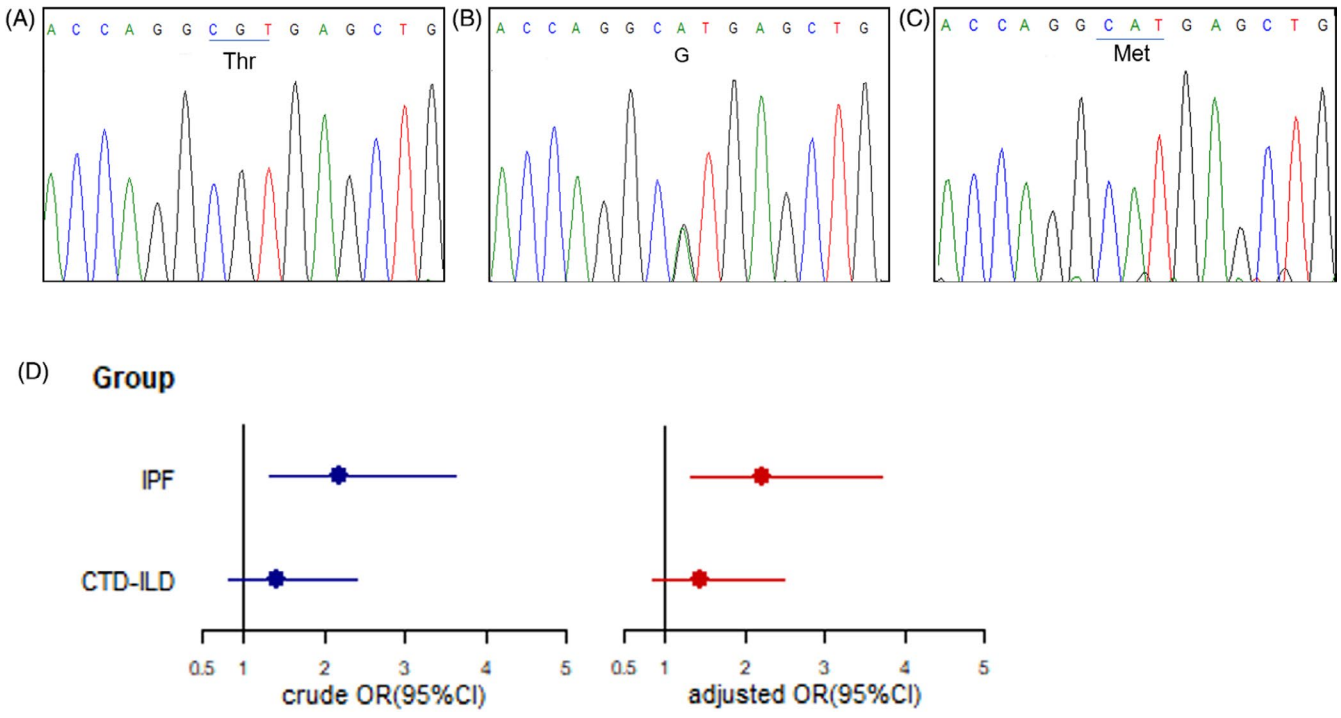


**FIGURE 1** Sanger sequencing results of rs117179004, and allele-based association of rs117179004 and idiopathic pulmonary fibrosis (IPF) or connective tissue diseases related interstitial lung disease (CTD-ILD). A, wild type (GG); B, mutant heterozygote (GA); C, mutant homozygote (AA); D, forest plot of rs117179004 in IPF or CTD-ILD, dot and error bars show OR (odds ratio) and 95% CI (confidence interval), respectively

**TABLE 4** Genotype and allele frequencies of rs117179004 in IPF and control subjects in Stage 2

| Type | IPF *n* (%) | Control *n* (%) | Models | IPF vs. Control | | | |
|---|---|---|---|---|---|---|---|
| | | | | OR (95%CI) | *p* | OR (95%CI)[a] | *p*[a] |
| Genotype | | | | | | | |
| GG | 123 (75.9%) | 198 (88.0%) | Codominant 1 | **2.291 (1.322–3.969)** | **0.003** | **2.348 (1.342–4.110)** | **0.003** |
| GA | 37 (22.8%) | 26 (11.6%) | Codominant 2 | 3.220 (0.289–35.81) | 0.342 | 3.259 (0.287–36.988) | 0.341 |
| AA | 2 (1.2%) | 1 (0.4%) | Dominant | **2.325 (1.355–3.989)** | **0.002** | **2383 (1.376–4.128)** | **0.002** |
| | | | Recessive | 2.800 (0.252–31.144) | 0.402 | 2.830 (0.250–32.080) | 0.401 |
| $P_{HWE}$ | 0.883 | | Additive | **2.208 (1.323–3.685)** | **0.002** | **2.255 (1.339–3.797)** | **0.002** |
| Allele | | | | | | | |
| G | 283 (87.3%) | 422 (93.7%) | | 1 | | 1 | |
| A | 41 (12.7%) | 28 (6.2%) | | **2.183 (1.320–3.613)** | **0.002** | **2.228 (1.336–3.716)** | **0.002** |

*Note*: Codominant 1: GA vs. GG, codominant 2: AA vs. GG, dominant: GA +AA vs. GG, recessive: AA vs. GG +GA, additive: GG vs. GA vs. AA, allele: A vs. G.

Abbreviations: CI, confidence interval; HWE, Hardy-Weinberg equilibrium; IPF, idiopathic pulmonary fibrosis; n, number of subjects; OR, odd ratio; vs., versus.

[a]Adjusted for gender, age, and smoking status; the significant results were in bold.

As shown in Table 6, at the prior probability level of 0.1 and FPRP threshold of 0.2, the significant findings became unnoteworthy either in stage one or in stage two. However, when we combined the two stages, the observed FPRP values under the prior probability level of 0.1 were all less than 0.20 (Table 6, FPRP = 0.039, GA vs. GG; FPRP = 0.021, GA +AA vs. GG; FPRP = 0.013, GG vs. GA vs. AA; FPRP = 0.012, A vs. G), confirming notable associations.

## 4 | DISCUSSION

The present investigation aimed to explore the correlation between the *HYAL1* gene and IPF or CTD-ILD in Southern Han Chinese population. Our findings suggested that rs117179004 of *HYAL1* might be significantly correlated with IPF susceptibility but not CTD-ILD.

Next-generation sequencing (NGS) of *HYAL1* detected seven nonsynonymous variants in stage one, among which only rs117179004 showed significantly difference between the IPF and control groups. Considering the relatively higher cost of NGS in stage one, fewer controls were included. To further confirm the preliminary results of stage one, we explored the probably associations in stage two and included more controls, further verified the results. The minor A allele frequencies of rs117179004 were 0.00035, 0.0005, 0.0005, and 0.0292 in European, American, African, and Asian, respectively (dbSNP BUILD152 https://www.ncbi.nlm.nih.gov/snp/rs117179004#frequency_tab). However, the frequency is much higher in Chinese population, 0.07885 in Chinese millionome database (CMDB) (https://db.cngb.org/cmdb/). In our study, the A allele frequency of this SNP in healthy control group is 0.061 (0.060 in stage one, 0.062 in stage two), and the GG, GA, and AA frequencies were 0.880, 0.120, and 0.00 in stage one, and 0.880, 0.116, and 0.04 in stage two, respectively. Therefore, ethnic differences are shown in this nonsynonymous SNP.

The conservative incidence range of IPF was 3~9 cases per 100,000 per year in Whites, 1~4 in east Asians.[18] Relatively lower incidence of IPF in East Asians suggests heterogeneity of this disease. A gain-of-function promoter variant (*MUC5B*, rs35705950) has been reported to be associated with IPF. The minor T allele of this SNP presents in approximately 30%–40% of patients with IPF, compared with only 9%–10% in healthy controls.[19,20] However, T allele frequencies in Chinese and Japanese patients with IPF were 3.7% and 3.4%, respectively, and in healthy controls were 0.8% consistently.[12,21] This distinct difference indicates the genetic heterogeneity of IPF across different ethnicities. Considering the A allele frequency of rs117179004 is much higher in Asians than in Europeans, we propose that the A allele of rs117179004 may be a risk factor for the development of IPF in Asians or Southern Han Chinese.

CTD-ILD is a collection of various systemic autoimmune disorders that result in lung fibrosis.[2] Patients with CTD-ILD differ from IPF in terms of pathogenesis, demographics of affected cases, and clinical course.[1,2,22] IPF associated genomic markers, such as *MUC5B* rs35705950 and *TOLLIP* rs5743890, are less prevalent

**TABLE 5** Genotype and allele frequencies of rs117179004 in CTD-ILD and control subjects in Stage 2

| Type | CTD-ILD *n* (%) | Control *n* (%) | Models | CTD-ILD vs. Control | | | |
|---|---|---|---|---|---|---|---|
| | | | | OR (95%CI) | *p* | OR (95%CI)[a] | *p*[a] |
| Genotype | | | | | | | |
| GG | 153 (84.1) | 198 (88.0%) | Codominant 1 | 1.344 (0.754–2.396) | 0.317 | 1.326 (0.737–2.383) | 0.346 |
| GA | 27 (14.8) | 26 (11.6%) | Codominant 2 | 2.588 (0.233–28.809) | 0.439 | 3.818 (0.335–43.478) | 0.280 |
| AA | 2 (1.1%) | 1 (0.4%) | Dominant | 1.390 (0.790–2.446) | 0.253 | 1.401 (0.790–2.485) | 0.248 |
| | | | Recessive | 2.489 (0.224–27.669) | 0.458 | 3.681 (0.324–41.883) | 0.293 |
| $P_{HWE}$ | | 0.883 | Additive | 1.390 (0.823–2.348) | 0.219 | 1.428 (0.841–2.425) | 0.187 |
| Allele | | | | | | | |
| G | 333 (91.5) | 422 (93.7%) | | – | | – | |
| A | 31 (8.5%) | 28 (6.2%) | | 1.403 (0.825–2.386) | 0.211 | 1.444 (0.843–2.474) | 0.181 |

*Note:* Codominant 1: GA vs. GG, codominant 2: AA vs. GG, dominant: GA + AA vs. GG, recessive: AA vs. GG + GA, additive: GG vs. GA vs. AA, allele: A vs. G.

Abbreviations: CI, confidence interval; CTD-ILD, Connective Tissue Disease associated Interstitial Lung Disease; HWE, Hardy-Weinberg equilibrium; n, number of subjects; OR, odd ratio; vs., versus.

[a]Adjusted for gender, age, and smoking status.

**TABLE 6** FPRP analysis for the significant associations between IPF and rs117179004

| Stage | Model | OR (95% CI) | *p* | Statistical power | Prior probability | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | 0.25 | 0.1 | 0.01 | 0.001 | 0.0001 |
| Stage 1 | Codominant 1 | 2.041 (1.100–3.787) | 0.024 | 0.164 | 0.302 | 0.565 | 0.934 | 0.993 | 0.999 |
| | Dominant | 2.230 (1.208–4.117) | 0.010 | 0.102 | 0.233 | 0.476 | 0.909 | 0.990 | 0.999 |
| | Allele | 2.268 (1.265–4.068) | 0.006 | 0.083 | **0.179** | 0.396 | 0.878 | 0.986 | 0.998 |
| Stage 2 | Codominant 1 | 2.291 (1.322–3.969) | 0.003 | 0.065 | **0.125** | 0.299 | 0.825 | 0.979 | 0.998 |
| | Dominant | 2.325 (1.355–3.989) | 0.002 | 0.056 | **0.105** | 0.261 | 0.795 | 0.975 | 0.997 |
| | Additive | 2.208 (1.323–3.685) | 0.002 | 0.070 | **0.095** | 0.240 | 0.776 | 0.972 | 0.997 |
| | Allele | 2.183 (1.320–3.613) | 0.002 | 0.072 | **0.090** | 0.230 | 0.766 | 0.971 | 0.997 |
| Combined | Codominant 1 | 2.156 (1.445–3.219) | $1.701 \times 10^{-4}$ | 0.038 | **0.013** | **0.039** | 0.309 | 0.819 | 0.978 |
| | Dominant | 2.267 (1.529–3.362) | $4.673 \times 10^{-5}$ | 0.020 | **0.007** | **0.021** | **0.188** | 0.701 | 0.959 |
| | Additive | 2.212 (1.522–3.214) | $3.125 \times 10^{-5}$ | 0.021 | **0.004** | **0.013** | **0.129** | 0.600 | 0.937 |
| | Allele | 2.213 (1.528–3.204) | $2.612 \times 10^{-5}$ | 0.020 | **0.004** | **0.012** | **0.115** | 0.567 | 0.929 |

*Note:* Codominant 1: GA vs. GG, dominant: GA +AA vs. GG, additive: GG vs. GA vs. AA, allele: A vs. G. statistical power was calculated using the number of observations in the subgroup and the OR and P values in this table; the level of FPRP threshold was set at 0.2, and the significant results were in bold.

Abbreviations: CI, confidence interval; FPRP, false-positive report probability; OR, odd ratio.

in CTD-ILD.[23–25] In present study, the frequency of rs117179004 showed similar trend in CTD-ILD, suggested different genetic features between IPF and CTD-ILD.

The exact biologic function of SNP rs117179004 is still unknown. *HYAL1* encodes a lysosomal hyaluronidase, which intracellularly degrade hyaluronan, one of the major glycosaminoglycans

of the extracellular matrix.[26] Mutations in this gene are associated with mucopolysaccharidosis type IX, or hyaluronidase deficiency (OMIM: 601492).[27] A change in sequence (from G to A) of rs117179004 results in a shift of the amino acid chain (from T to M). The affected amino acid is located in the βII-hairpin unit of EGF-like domain, which is elongated and has extensive solvent accessibility, consistent with mediation of protein-protein interactions.[28] Functional studies indicated that hyaluronidase treatment blocked bleomycin-induced lung fibrosis while decreasing transforming growth factor (TGF)-β production and collagen deposition.[8] Intranasal immobilized hyaluronidase prevented connective tissue growth in the lungs exposed to bleomycin, indicating antifibrotic effect of hyaluronidase.[29] Alternative mRNA splicing exhibited different cellular expression of enzymatically active of hyaluronidase and may explain the elevated hyaluronidase levels in tumor bladder/prostate cancer.[30] Hence, we suppose this variant may affect the development of IPF by potentially altering gene expression or weakening enzyme activity. Public databases including expression data provide the possibility to perform genotype-based mRNA expression analysis and explore the effects of SNPs.[31,32] However, no significant eQTLs were found for rs117179004 in the GTEx portal (https://www.gtexportal.org/), probably because the relatively lower frequency in European (0.00035) and American (0.0005). Databases including more Chinese or functional study in the future maybe useful to further explore the potential role of this significant polymorphism.

Several limitations should be considered in this study. First, the relatively small sample size may result in limited statistical power. Second, we only included Southern Han Chinese participants from single geographical location, and selection bias may exist. Third, we were unable to further analyze the potential gene-environment interactions influencing the risk of IPF for the lack of environmental exposure information. Fourth, the mechanism of how rs117179004 involved in IPF risk has not been investigated in this study.

To the best of our knowledge, this is a first study to examine the association between the *HYAL1* gene and the risk of IPF. In aggregate, we found that the nonsynonymous polymorphism rs117179004 of *HYAL1* was associated with IPF, but not with CTD-ILD, in Southern Han Chinese population. The A allele of rs117179004 was related to the development of IPF. Further studies with more subjects and diverse populations, as well as functional studies, are needed to confirm our findings.

## ACKNOWLEDGEMENT

We would like to particularly acknowledge all the participants who took part in this study.

## CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## DATA AVAILABILITY STATEMENT

Data used in this research are available from the corresponding author on request.

## ORCID

*Juan Liu* https://orcid.org/0000-0003-2048-9512

## REFERENCES

1. Devine MS, Garcia CK. Genetic interstitial lung disease. *Clin Chest Med*. 2012;33(1):95-110.
2. Demoruelle MK, Mittoo S, Solomon JJ. Connective tissue disease-related interstitial lung disease. *Best Pract Res Clin Rheumatol*. 2016;30(1):39-52.
3. 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.
4. Kropski JA, Blackwell TS, Loyd JE. The genetic basis of idiopathic pulmonary fibrosis. *Eur Respir J*. 2015;45(6):1717-1727.
5. Fernandez BA, Fox G, Bhatia R, et al. A Newfoundland cohort of familial and sporadic idiopathic pulmonary fibrosis patients: clinical and genetic features. *Respir Res*. 2012;13:64.
6. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet*. 2013;45(6):613-620.
7. Calve S, Isaac J, Gumucio JP, Mendias CL. Hyaluronic acid, HAS1, and HAS2 are significantly upregulated during muscle hypertrophy. *Am J Physiol Cell Physiol*. 2012;303(5):C577-C588.
8. Bitencourt CS, Pereira PA, Ramos SG, et al. Hyaluronidase recruits mesenchymal-like cells to the lung and ameliorates fibrosis. *Fibrogenesis Tissue Repair*. 2011;4(1):3.
9. Jiang D, Liang J, Noble PW. Hyaluronan in tissue injury and repair. *Annu Rev Cell Dev Biol*. 2007;23:435-461.
10. Leng D, Huang X, Yi J, Zhao H, Zhang Y. HYAL1 is downregulated in idiopathic pulmonary fibrosis and inhibits HFL-1 fibroblast proliferation when upregulated. *Biomed Res Int*. 2020;2020:3659451.
11. 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.
12. Deng Y, Li Z, Liu J, et al. Targeted resequencing reveals genetic risks in patients with sporadic idiopathic pulmonary fibrosis. *Hum Mutat*. 2018;39(9):1238-1245.
13. Li Z, Zhou C, Tan L, et al. Variants of genes encoding collagens and matrix metalloproteinase system increased the risk of aortic dissection. *Sci China Life Sci*. 2017;60(1):57-65.
14. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst*. 2004;96(6):434-442.
15. He J, Zou Y, Liu X, et al. Association of common genetic variants in pre-microRNAs and neuroblastoma susceptibility: a two-center study in Chinese children. *Mol Ther Nucleic Acids*. 2018;11:1-8.
16. Shi X, Ma Y, Li H, Yu H. Association between FCGR2A rs1801274 and MUC5B rs35705950 variations and pneumonia susceptibility. *BMC Med Genet*. 2020;21(1):71.
17. Zhao JH. Gap: genetic analysis package. *J Stat Softw*. 2007;23(8):1-18.
18. Hutchinson J, Fogarty A, Hubbard R, McKeever T. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur Respir J*. 2015;46(3):795-806.
19. Seibold MA, Wise AL, Speer MC, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med*. 2011;364(16):1503-1512.

20. Borie R, Crestani B, Dieude P, et al. The MUC5B variant is associated with idiopathic pulmonary fibrosis but not with systemic sclerosis interstitial lung disease in the European Caucasian population. *PLoS One*. 2013;8(8):e70621.
21. Horimasu Y, Ohshimo S, Bonella F, et al. MUC5B promoter polymorphism in Japanese patients with idiopathic pulmonary fibrosis. *Respirology*. 2015;20(3):439-444.
22. Meyer KC. Pulmonary fibrosis, part I: epidemiology, pathogenesis, and diagnosis. *Expert Rev Respir Med*. 2017;11(5):343-359.
23. Volkmann ER, Tashkin DP, Roth MD, et al. The MUC5B promoter variant does not predict progression of interstitial lung disease in systemic sclerosis. *Semin Arthritis Rheum*. 2020;50(5):963-967.
24. Wang N, Zhang Q, Jing X, Guo J, Huang H, Xu Z. The association between MUC5B mutations and clinical outcome in patients with rheumatoid arthritis-associated interstitial lung disease: a retrospective exploratory study in China. *Med Sci Monit*. 2020;26:e920137.
25. Newton CA, Oldham JM, Ley B, et al. Telomere length and genetic variant associations with interstitial lung disease progression and survival. *Eur Respir J*. 2019;53(4):1801641.
26. Weber GC, Buhren BA, Schrumpf H, Wohlrab J, Gerber PA. Clinical applications of hyaluronidase. *Adv Exp Med Biol*. 2019;1148:255-277.
27. Triggs-Raine B, Salo TJ, Zhang H, Wicklow BA, Natowicz MR. Mutations in HYAL1, a member of a tandemly distributed multigene family encoding disparate hyaluronidase activities, cause a newly described lysosomal disorder, mucopolysaccharidosis IX. *Proc Natl Acad Sci U S A*. 1999;96(11):6296-6300.
28. Chao KL, Muthukumar L, Herzberg O. Structure of human hyaluronidase-1, a hyaluronan hydrolyzing enzyme involved in tumor growth and angiogenesis. *Biochemistry*. 2007;46(23):6911-6920.
29. Dygai AM, Skurikhin EG, Ermakova NN, et al. Antifibrotic effects of immobilized hyaluronidase in repeated bleomycin-induced lesions of the alveolar epithelium. *Bull Exp Biol Med*. 2013;155(4):501-506.
30. Lokeshwar VB, Schroeder GL, Carey RI, Soloway MS, Iida N. Regulation of hyaluronidase activity by alternative mRNA splicing. *J Biol Chem*. 2002;277(37):33654-33663.
31. Zhu J, Fu W, Jia W, Xia H, Liu GC, He J. Association between NER pathway gene polymorphisms and wilms tumor risk. *Mol Ther Nucleic Acids*. 2018;12:854-860.
32. Zhuo ZJ, Liu W, Zhang J, et al. Functional polymorphisms at ERCC1/XPF genes confer neuroblastoma risk in Chinese children. *EBioMedicine*. 2018;30:113-119.

**How to cite this article:** Liu J, Deng Y, Wang Z, et al. A nonsynonymous polymorphism (rs117179004, T392M) of hyaluronidase 1 (*HYAL1*) is associated with increased risk of idiopathic pulmonary fibrosis in Southern Han Chinese. *J Clin Lab Anal*. 2021;35:e23782. https://doi.org/10.1002/jcla.23782