

AluYb8 insertion polymorphism in the *MUTYH* gene impairs mitochondrial DNA maintenance and affects the age of onset of IPF

Wei Zhou^{1,4}, Jiapeng Sun^{1,4}, Wenwen Guo³, Yi Zhuang^{1,2}, Lizhi Xu^{1,4}, Yaping Wang^{1,4}

¹Department of Medical Genetics, Nanjing University School of Medicine, Nanjing, China

²Department of Respiriology, Medical School Affiliated Drum Tower Hospital, Nanjing University, Nanjing, China

³Department of Pathology, The Second Affiliated Hospital of Nanjing Medical University, Nanjing, China

⁴Jiangsu Key Laboratory of Molecular Medicine, Nanjing University School of Medicine, Nanjing, Jiangsu, China

Correspondence to: Wenwen Guo, Yaping Wang; email: hguowen@163.com, wangyap@nju.edu.cn

Keywords: polymorphism, *MUTYH*, mitochondrial DNA stabilization, idiopathic pulmonary fibrosis, age of onset

Received: May 10, 2018

Accepted: January 23, 2019

Published: February 4, 2019

Copyright: Zhou et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: Idiopathic pulmonary fibrosis (IPF) is an age-related fatal disease with an unknown etiology. Increased oxidative stress and mitochondrial dysfunction are thought to be involved in its pathogenesis. However, the effect of the *AluYb8MUTYH* polymorphism on IPF is not known.

Results: The mean age of onset for IPF in patients homozygous for the *AluYb8MUTYH* variant (*P/P*) was 66.5 years old, which was significantly earlier than that in patients with the wild-type (*A/A*, 70.45 years old). For the 97 male IPF patients with lung function data, the FVC% of the *P/P* patients was lower than that of the wild-type (*A/A*) or heterozygous (*A/P*) patients. The laboratory analysis indicated that an increased mtDNA content and impaired mitochondrial quality control were associated with the *P/P* genotype. We also confirmed that *AluYb8* insertion into *MUTYH* caused decreased *MUTYH1* expression in lung tissues.

Methods: We compared the lung function of IPF patients and observed the mtDNA content, mtDNA integrity and molecular expression of mitochondrial quality control among subjects with different *AluYb8MUTYH* genotypes. Additionally, immunoblotting and a reporter gene system were used to test whether altered mitochondrial *MUTYH1* expression was linked to *AluYb8MUTYH*.

Conclusions: The *AluYb8* insertion polymorphism in *MUTYH* impairs mtDNA stability and affects the age of onset of IPF.

INTRODUCTION

As one of the most common interstitial lung diseases, idiopathic pulmonary fibrosis (IPF) is defined as a chronic progressive fibrotic lesion with unknown etiology and a histological pattern of usual interstitial pneumonia (UIP). Although the clinical state of IPF can last several years, its prognosis is quite poor, with median survival of 3–5 years from diagnosis. Several risk factors have been implicated in IPF pathogenesis, including smoking, environmental exposures, microbial infections and genetic susceptibility. However, epide-

miological data show that aging is the most important risk factor for IPF, suggesting that age-associated changes may predispose individuals to IPF development [1].

Accumulating evidence shows that mitochondria play a key role in the pathogenesis of aging and age-related diseases. Mitochondria produce ATP via oxidative phosphorylation and are the energy-producing organelles of cells. Mitochondria are also the major source of reactive oxygen species (ROS), which can damage biological macromolecules in cells, including lipids, DNA, RNA and

proteins. Among the types of ROS-mediated DNA damage, an oxidized form of the guanine base [8-oxo-7, 8-dihydroguanine (8-oxoG)] is considered mutagenic [2]. Mitochondrial DNA (mtDNA) is vulnerable to ROS due to its proximity to the respiratory chain and lack of protective histones. Aging is accompanied by accumulation of 8-oxoG and a reduced oxidative damage repair ability, which can be a driving mechanism for the accumulation of dysfunctional mitochondria and multiple organ injury. Increasing evidence has shown that ROS-mediated DNA damage is involved in a variety of age-related diseases, including sporadic tumors and neurodegenerative, cardiovascular and cerebrovascular diseases [3–5].

To maintain genomic integrity, several biochemical repair pathways have developed in cells. Base excision repair (BER) is the main mechanism that deals with oxidized bases and is initiated by recognition of DNA glycosylation [6]. MUTYH (the human MutY homolog) is a DNA glycosylase enzyme of the BER system that is responsible for removing an adenine that is incorrectly paired with 8-oxoG and leaves an apurinic/aprimidine site (AP-site) [7, 8]. Two MUTYH forms [MUTYH 1 (CCDS41320.1, p57) and MUTYH 2 (CCDS41322.1, p60)] are present in human cells. The two MUTYH isoforms are encoded by different transcripts that are grouped into three major categories (named the α , β and γ transcripts) with differing 5' exons. MUTYH 1 is a product of the α transcript and has a mitochondrial targeting signal (MTS) at its N-terminus, which mainly localizes the isoform to the mitochondria, whereas MUTYH 2 is generated from the β or γ transcript, lacks the MTS and localizes to the nucleus.

We previously described an *AluYb8* element (a transposable short repeat sequence) insertion/deletion polymorphism in the fifteenth intron of the *MUTYH* gene called *AluYb8MUTYH* [9]. According to the presence (P) or absence (A) of the insertion, three genotypes were observed in the human population: homozygous absent (*absence/absence*, *A/A*), homozygous present (*presence/presence*, *P/P*) and heterozygous (*absence/presence*, *A/P*). *AluYb8MUTYH* is a common variant with an allele frequency of 43.2%, and approximately 21% of investigated Chinese individuals have been shown to have the *P/P* genotype for this variation [9]. Additionally, a similar frequency distribution for this polymorphism was detected in healthy Germans by screening a cohort with a small sample size, suggesting that it may also be common variant in the Caucasian population [9]. Using human peripheral blood leukocytes, we found that the *AluYb8* insertion was associated with reduced MUTYH1 protein expression and that the protein was selectively localized in the mitochondria. [10].

Compared to individuals with *A/A* or *A/P*, individuals with the *P/P* genotype had an unstable mtDNA state and decreased mitochondrial activity in their cells, which could affect the occurrence and clinical phenotypes of age-related diseases [9, 11, 12].

In the present study, we extended our investigation of alterations in MUTYH protein expression to individuals carrying different genotypes, evaluated functional impairment of mtDNA maintenance in IPF patients, and examined whether this polymorphism was associated with the occurrence of IPF and affected the prognosis of age-related diseases.

RESULTS

The polymorphic distribution of *AluYb8MUTYH* in the IPF patients and healthy controls

The three genotypes were identified by agarose gel electrophoresis of PCR products. Supplementary Table 2 shows the frequencies of the three genotypes detected in the current subjects. The allele frequencies for variant “P” were 41.3% and 44.9% in 277 IPF patients and 810 healthy controls, respectively. No significant difference was found between the two groups.

The pulmonary function test results of the IPF patients

We assessed the correlation between the pulmonary function test results and the *AluYb8MUTYH* genotypes in IPF patients. Pulmonary function test data were collected for 115 hospitalized IPF patients, including 97 males and 18 females. We only compared data from male IPF patients due to the smaller sample size of women. The results showed that the FVC% in the *P/P* patients was significantly lower than that in patients with the *A/A* and *A/P* genotypes. No significant difference was found in FEV₁% and DLCO% among patients with the three genotypes (Table 1).

The onset and death ages of IPF patients with different *AluYb8MUTYH* genotypes

We recruited 277 sporadic IPF patients to investigate the relationship between *AluYb8MUTYH* and IPF development. The mean age of occurrence for IPF patients with the *P/P* genotype was 66.5 years old, which was significantly lower than that for patients (70.45 years old) with the *A/A* genotype. Among the 210 IPF patients for whom we obtained follow-up data, 95 patients (45%) died, and the mean survival time from the diagnosis of IPF was 24.6 months. A lower age of

Table 1. Comparison of pulmonary function test results from IPF patients with different *AluYb8MUTYH genotypes.**

	Total n=97	A/A n=27	A/P n=50	P/P n=20	P-value
Age, year	65.5±8.4	65.2±9.7	65.6±8.3	65.5±7.4	
FVC, % predicted	72.6±19.1	74.8±12.6	75.7±20.7	62.1±19.2	0.02
FEV ₁ , % predicted	76.1±20.0	77.9±17.5	78.7±21.6	67.0±17.3	0.07
DLOC, % predicted	53.3±21.1	51.8±18.2	56.1±21.9	48.5±22.9	0.36

*All of the tested IPF patients were male. FVC, forced vital capacity; FEV₁, forced expiratory volume in 1 s; DLOC, diffusion capacity for carbon monoxide.

death was also observed for the *P/P* patients compared with that of the *A/A* patients (Table 2). However, no significant difference in the survival time of the IPF patients was found among the three *AluYb8MUTYH* genotypes.

Relationship between the *AluYb8MUTYH* genotypes and the mtDNA content in IPF patients

The mtDNA content was examined in peripheral blood cells from 206 patients with IPF and 206 age-matched controls. First, we performed real-time PCR to test the fragments of two genes (*MT-ND1* and *MT-TL1*) as mtDNA markers, and using the *β-actin* as a reference. The results showed that the relative mtDNA content in the IPF patients was significantly higher than that in the healthy controls (Figure 1A and Supplementary Figure 1A). Then, we compared the relative mtDNA content among the IPF patients and the age-matched controls. In the healthy controls, the *P/P* individuals had a decreased mtDNA content compared to that of the *A/A* or heterozygous (*A/P*) individuals (Figure 1B). Unexpectedly, the relative mtDNA content was significantly higher in the IPF patients with *P/P* compared to that of the patients with *A/A*. We also compared the relative mtDNA content between the healthy individuals and IPF patients with the same genotypes and found a significant increase in the mtDNA content only in the *P/P* patients, whereas no significant difference was found between the patients and controls with the *A/A* or *A/P* genotype (Figure 1B). The *MT-TL1* results are consistent with those obtained for *MT-ND1* (Supplementary Figure 1B).

The droplet-based digital PCR (ddPCR) method was used for confirmation of the mtDNA copy number analysis with peripheral blood samples randomly selected from the 206 IPF patients and age-matched controls and to verify the correlation of the *AluYb8MUTYH* polymorphism with the mtDNA content. Consistent with the relative quantification of

the mtDNA content, we found that the cells from the IPF patients had a significantly higher mtDNA content than those from the age-matched healthy controls ($P < 0.01$) (Supplementary Figure 2A). The controls with the *P/P* genotype had decreased mtDNA contents compared to those of the wild-type (*A/A*) and heterozygote (*A/P*) controls (Supplementary Figure 2B). Additionally, an interesting phenomenon occurred in which the IPF patients with the *P/P* genotype had a significantly increased mtDNA content compared to that of the controls with the same genotype ($P < 0.05$, Supplementary Figure 2B) in this experiment.

Association of *AluYb8MUTYH* with the accumulation of damaged mtDNA in the IPF patients

To further investigate the relationship between the *AluYb8* insertion in *MUTYH* and DNA damage in IPF patients, we performed a long-range PCR to analyze the mtDNA and nuclear DNA integrity. As shown in Figure 1C and 1D, the relative amplification of nuclear DNA was significantly lower in the IPF patients when compared with that in the healthy controls. The decreased relative amplification of nuclear DNA was found in IPF patients with the *A/A*, *A/P* and *P/P* genotypes, comparing to the healthy controls with the same genotypes, respectively (Figure 1E), suggesting that an increased oxidative stress exist in the IPF patients. However, no significant difference was found for the relative amplification of mtDNA between IPF patients and healthy controls. Interestingly, the relative amplification of mtDNA was higher in the subjects with *P/P* genotype, either in IPF patients or healthy controls, when compared with the *A/A* or *A/P* genotype (Figure 1F). But, a lower relative amplification of mtDNA was observed in the *P/P* genotype for IPF patients, comparing to the healthy controls with the same genotype. Combining with the result of mtDNA content analysis, we considered that the *AluYb8MUTYH* variant could injure the repair of oxidative damage to mtDNA and cause the accumulation of damaged mtDNA

Table 2. Comparison of the ages of onset and death among IPF patients with different *AluYb8MUTYH* genotypes.

Genotypes	Age of onset*			Age of death*		
	Number	Mean age	<i>P</i> -value	Number	Mean age	<i>P</i> -value
<i>A/A</i>	65	70.5±1.0		32	74.6±1.4	
<i>A/P</i>	108	69.7±0.8	0.54	45	71.5±1.3	0.12
<i>P/P</i>	37	66.5±1.1	0.01	18	69.4±1.6	0.02

*A comparison was made between the *A/P* and *A/A* or between the *P/P* and *A/A* genotypes using one-way ANOVA, followed by post hoc analysis. The age is shown as the mean ± SD. *P*<0.05 was considered statistically significant.

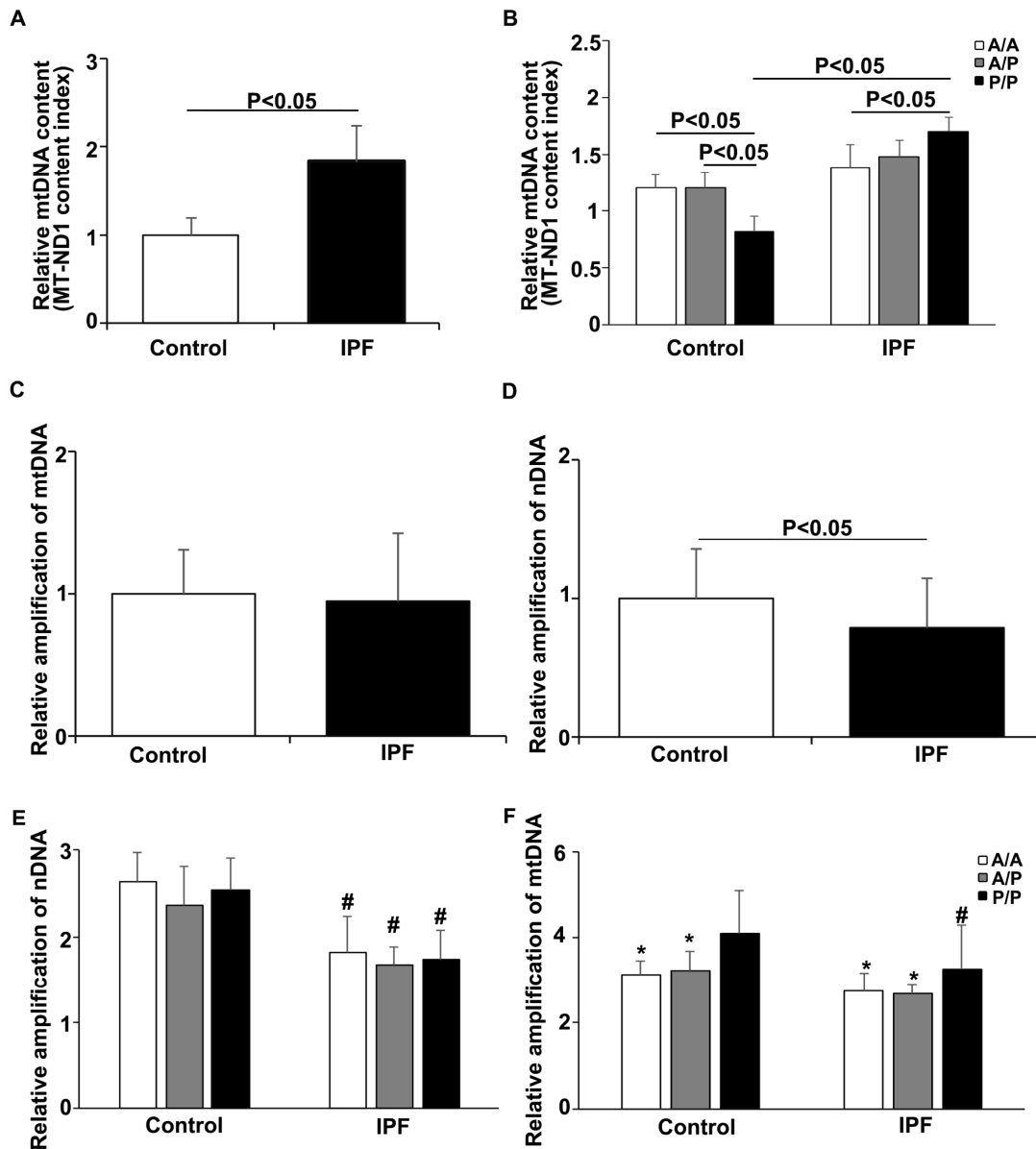


Figure 1. Relationship between the *AluYb8MUTYH* genotype and mtDNA state in the IPF patients and healthy controls. (A) The *MT-ND1* content index was increased in the IPF patients, *P*<0.05. **(B)** The *MT-ND1* content index of the *P/P* patients was significantly higher than that of the healthy controls with the same genotype, *P*<0.05. **(C)** Relative amplification of mtDNA in the IPF patients and healthy controls. **(D)** Relative amplification of nDNA in the IPF patients and healthy controls. **(E)** Relative amplification of nDNA in patients with different *AluYb8MUTYH* genotypes. **(F)** Relative amplification of mtDNA in patients with different *AluYb8MUTYH* genotypes. * A significant difference compared to *P/P*. # A significant difference compared to the healthy controls with the same genotype, *P*<0.05.

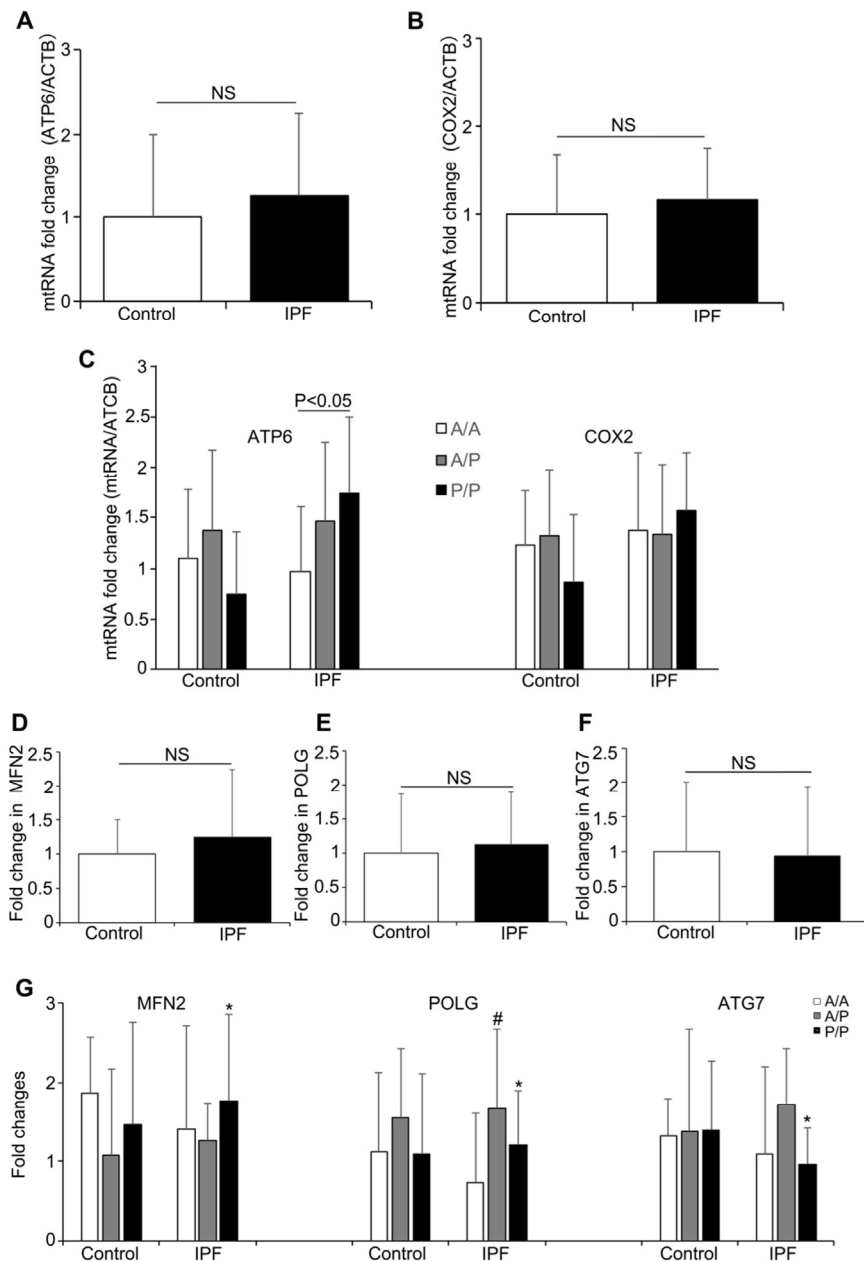


Figure 2. The mRNA levels of mitochondrial genes and mitochondrial regulation-related genes in the IPF patients and healthy controls. (A) *ATP6* and (B) *COX2* mRNA expression in the IPF patients and healthy controls. (C) The *ATP6* and *COX2* expression levels in the patients and healthy controls with different *AluYb8MUTYH* genotypes. (D–F) *MFN2*, *POLG* and *ATG7* expression in the patients and healthy controls. (G) The *MFN2*, *POLG* and *ATG7* mRNA levels among the IPF patients and healthy controls with different *AluYb8MUTYH* genotypes. * A significant difference between IPF patients with the *A/P* and *P/P* genotypes, $P < 0.05$. # A significant difference between IPF patients with the *A/A* and *A/P* genotypes, $P < 0.05$.

in cells, especially in IPF patients under the severe oxidative stress. To test this hypothesis, we chose *ATP6* and *COX2* in the mtDNA as targets to investigate whether the expression levels of mitochondrial genes exhibited a synchronizing with the variations of mtDNA content and the relative amplification of mtDNA. The results showed no significant differences in mRNA expression of either of the ana-

lyzed genes between the IPF patients and healthy controls, but the *ATP6* mRNA expression level was higher in the *P/P* patients than in the *A/A* patients (Figure 2A–2C). Furthermore, we analyzed the transcript levels of the *POLG*, *MFN2* and *ATG7* genes, which regulate mtDNA replication, mitochondrial fusion and mitophagy, respectively. No significant difference was found in *POLG*, *MFN2* and *ATG7*

mRNA expression between the IPF patients and healthy controls (Figure 2D–2F). In the IPF patients, the *MFN2* expression level in the homozygous *AluYb8MUTYH* group (*P/P*) was significantly higher than that in the heterozygous group (*A/P*), and the *POLG* and *ATG7* mRNA levels in the *P/P* patients were lower than those in the *A/P* patients (Figure 2G). However, no significant difference was found in the transcript levels of the three target genes among healthy individuals with the *A/A*, *A/P* and *P/P* genotypes (Figure 2G). These data suggested that the increased level of mtDNA content in the IPF patients with *P/P* genotype could be resulted from an accumulation of impaired mtDNA.

MUTYH 1 protein expression is highly associated with the *AluYb8MUTYH* genotype

To uncover whether the *MUTYH* expression pattern that we previously reported in peripheral blood cells and fibroblasts was also present in lung tissue cells, we obtained snap frozen non-cancerous lung tissues from the tissue snap frozen non-cancerous lung tissues from the tissue bank of the Pathology Department, the Second Affiliated Hospital of Nanjing Medical University, as per the approved guideline. The immunoblotting results showed that *MUTYH* 1 protein expression was significantly reduced in the mutant (*P/P*) compared to that in either the wild-type (*A/A*) or heterozygous (*A/P*) cells (Figure 3A–B). The decreased

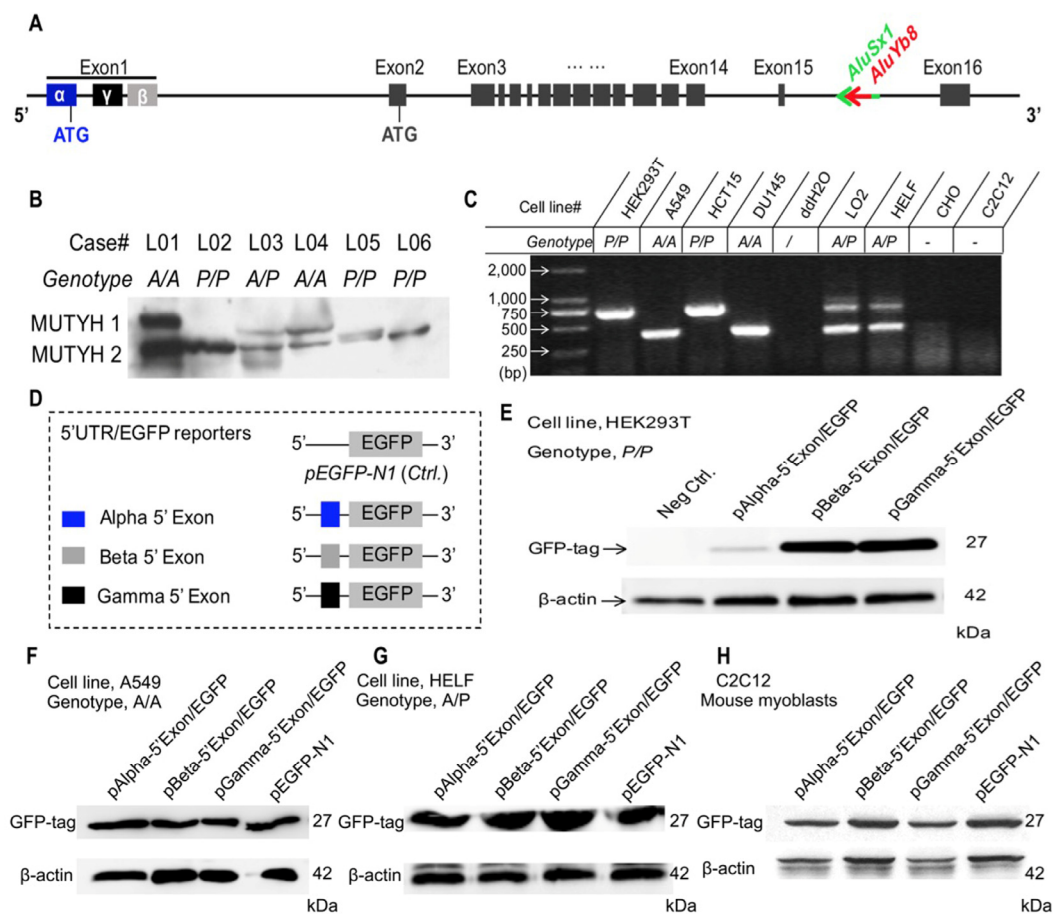


Figure 3. Association of the *AluYb8* insertion with altered protein expression patterns. (A) Schematic representation of the genomic structure of the *MUTYH* gene. The *AluYb8* element is inserted into existing *AluSx1* sequences in the *MUTYH* 15th intron and is clearly marked by a red arrow. *AluSx1* sequences are depicted with a bright green arrow. Exons are shown with black boxes. (B) Representative immunoblotting result showing the altered *MUTYH* protein expression pattern in human lung tissue cells with the mutant genotype (*P/P*). The two major *MUTYH* isoforms (*MUTYH* 1 and *MUTYH* 2) are indicated. β-actin was used as a protein loading control. The case IDs and their genotypes are shown. (C) *AluYb8MUTYH* genotyping of experimental cultured cell lines. (D) Schematic representation of the *pEGFP* cloned constructs with different 5' exon (1st exon) sequences from the *MUTYH* gene. The 5' exon sequences of the α, β and γ *MUTYH* transcripts are depicted as blue, gray and black boxes, respectively. (E–H) Representative immunoblotting results for GFP expression in the reporter gene system. The GFP reporter protein of the recombinant protein with the 5' exon from the α *MUTYH* transcript was observed in the A549 (*A/A* genotype) and HELFL (*A/P* genotype) human cell lines and the C2C12 (mouse myoblasts) cell line but not in human cells (HEK293T) with the mutant (*P/P*) genotype. GFP reporters from the recombinant vector with the 5' exon of the β and γ *MUTYH* transcripts were expressed in all cultured cells. β-actin was used as a protein loading control.

MUTYH 1 expression pattern in the lung tissues from patients with the *P/P* genotype was consistent with previous reports in blood cells [10]. Then, we used a reporter gene system to construct recombinant *pEGFP-N1* vectors for verification of the regulatory role of the polymorphism in the 5'-untranslated sequence of the α -type transcript in MUTYH1 expression (Figure 3C-D). We observed that the recombinant *pEGFP-N1* vector with the 5'-untranslated exon sequences of the α -type transcript (*pAlpha-5'Exon/EGFP*) expressed the GFP reporter protein in the wild-type (*A/A*) and heterozygous (*A/P*) human cell lines and rodent cells but exhibited decreased expression in human cells with the *P/P* genotype (Figure 3E). However, the GFP reporters of the recombinant *pEGFP-N1* vector with the 5' exons of the β - and γ -type MUTYH transcripts (*pBeta-5'Exon/EGFP* and *pGamma-5'Exon/EGFP*, respectively) were highly expressed in all of the cultured cells in this study (Figure 3F-3H). Meanwhile we adopted the reverse transcription quantitative PCR (RT-qPCR) method to determine the mRNA expression levels of the GFP reporter gene in all experimental cells transfected with the recombinant *pEGFP-N1* vectors, with an embedded *neomycin*-encoded gene within the *pEGFP-N1* vector and a housekeeping gene (*β -actin*) used as the references. The mRNA expression analysis results showed no significant differences in transfection efficiency among the three recombinant vectors (comparison of the *neomycin* mRNA levels (Supplementary Figure 4A), under standard transfection conditions, and no significant difference was observed in the GFP mRNA expression levels among these transfected cells (Supplementary Figure 4B). These results indicate that the translation of α -type MUTYH transcripts can be selectively suppressed in cells with the *P/P* genotype and that the *MUTYH* 5'-untranslated exon sequence may participate in regulation of host genes at the translational level.

DISCUSSION

IPF is a terminal disease that is characterized by extracellular matrix deposition with limited therapeutic choices. The morbidity and mortality of IPF are increasing worldwide and show a positive correlation with aging. Accumulated evidence indicates that oxidative stress and mitochondrial dysfunction are involved in IPF pathogenesis [13]. Furthering understanding of these alterations could provide novel therapeutic targets for IPF and improve quality of life for IPF patients.

The human *MUTYH* gene is located on chromosome 1 and contains 16 exons and primary transcripts (i.e., α , β and γ) [14]. MUTYH 1, which is translated by the α mRNA transcript, initiates the mitochondrial BER pathway by identifying and removing adenines opposite

8-oxoG and 2-hydroxyadenines opposite guanine, thereby preventing the base substitution G:C to T:A [15]. After excision of the mismatched A, an AP site left on the DNA strand generates a single-strand break (SSB), which can be further repaired by endonuclease, DNA polymerase and DNA ligase [16]. Buildup of SSBs in the mtDNA can cause mtDNA depletion, affect DNA polymerase in long-range PCR and activate calpain, thereby executing caspase-independent cell death [17, 18]. This mechanism can be used to explain why the decreased mtDNA content of cells and reduced efficiency of long fragment DNA amplification are associated with an increase in oxidative stress during aging. Conversely, a defect in MUTYH could fail to remove the mismatched adenine and 2-hydroxyadenine during oxidative stress. The AP sites and subsequent SSBs would not be generated, leading to accumulation of oxidative DNA damage in cells. In our study, amplification of long fragment mtDNA was significantly increased in *P/P* patients. This result could be due to loss of mitochondrial MUTYH 1 resulting from the *AluYb8* insertion and a consequent reduction in SSBs. This result also indicated that the increased mtDNA content in *P/P* patients should be associated with accumulation of oxidative damaged mtDNA due to MUTYH 1 dysfunction. In the present study, we also examined MUTYH expression in human lung tissues, and the results verified our previous reports in peripheral blood cells and cultured fibroblasts [10]. In addition, we constructed a recombinant genetic reporter system to test translation regulation of the 5' exon sequence of the α -type *MUTYH* transcript in cultured cells with the *A/A*, *A/P* and *P/P* genotypes. We confirmed that MUTYH1 expression was greatly inhibited at the translational level in cells with homozygous *AluYb8MUTYH*, which impaired MUTYH1-mediated mtDNA maintenance in *P/P* cells.

IPF is an age-related disease. All of the enrolled patients were ≥ 50 years old, and the oldest was 85 years old. The healthy individuals selected as controls in this work were age-matched. Our results showed that the mtDNA content in the cells from the healthy controls with the *P/P* genotype was lower than that of those with the *A/A* and *A/P* genotypes. This result is in line with our previous reports and reflects the impairment of mtDNA maintenance in the aged healthy controls homozygous for *AluYb8MUTYH* [11]. However, the *P/P* patients showed an increased mtDNA content because transcriptional expression of mtDNA-encoded genes was not upregulated synchronously. We previously showed that *P/P* cells with increased 8-oxoG lesion levels in the genomic DNA had reduced mitochondrial MUTYH 1 protein expression [10]. Sustained low-level oxidative stress can develop with aging and increase the burden of DNA oxidative damage, including 8-oxoG. Cells from individuals with the *P/P* genotype fail to accomplish

mtDNA oxidative damage repair because of the mitochondrial *MUTYH* 1 defect. Thus, mtDNA oxidative damages builds up in the mitochondria and can activate mitochondrial quality control (mtQC), which induces mtDNA degradation and mitophagy. Mitochondria are dynamic organelles, and mtQC is intimately linked to the dynamic behavior of mitochondria and maintenance of cellular homeostasis; thus, mtQC is also an efficient mechanism for preventing the accumulation of impaired mtDNA and maintaining intact mtDNA [19, 20]. The *MFN2*, *POLG* and *ATG7* genes are involved in mtQC and are responsible for regulation of mitochondrial fusion, mtDNA replication and mitophagy, respectively. Compared with the *MFN2*, *POLG* and *ATG7* mRNA levels in the cells of *A/A* patients, the expression levels of the three genes in *A/P* patients could reflect a role for the mtQC mechanism under oxidative stress. The increased *MFN2* expression and decreased *POLG* and *ATG7* expression suggested impairment of mtQC in the *P/P* patients. These results suggested a barrier for mtDNA oxidative damage repair in IPF patients with the *P/P* genotype, which would cause dysregulation of mtQC mechanisms; as a result, enhanced mitochondrial fusion and reduced mitophagy in the cells could lead to an increased mtDNA content and accumulation of mtDNA damage. However, it is still poorly understood the pathogenesis of this disease, further intensive study should be continued to elucidate the mechanism underlying the increase of mtDNA copy number in IPF.

The functions of mitochondria extend well beyond ensuring cellular energy demands. They are also involved in many cellular pathways, including cell proliferation, differentiation, autophagy, and regulation of apoptosis [21]. Dysfunctional mitochondria are even more likely to play a central role in aging and age-related diseases than was previously known. However, the change in the mtDNA level and the mechanisms underlying IPF remain to be elucidated. Ryu C et al. recently reported that the mtDNA concentration was increased in the broncho-alveolar lavage (BAL) and plasma of IPF patients [22]. These patients with high plasma mtDNA concentrations trended towards lower FVC% and reduced event-free survival. However, the origin and characteristics of the extracellular mtDNA in IPF patients have not been described. We observed that *P/P* patients showed lower FVC% than patients with the *A/A* and *A/P* genotypes. Moreover, the *P/P* patients had earlier ages of onset and death for IPF, although different survival times were not observed among IPF patients with the three genotypes.

The main novelty of our work is the relationship between mtDNA maintenance and the *AluYb8MUTYH*

genotypes, which has a phenotypic effect on the age-related disease IPF. The *AluYb8MUTYH* polymorphism is a common variant in Chinese and Western populations. IPF patients with the homozygous variant (*P/P*) are associated with an increased mtDNA content and accumulation of mtDNA damage in cells and show earlier onset and death ages for IPF.

MATERIALS AND METHODS

Subjects

We recruited 277 patients from Nanjing Drum Tower Hospital Affiliated to Nanjing University School of Medicine with a clinical diagnosis of sporadic IPF from 2002 to 2016. A total of 810 unrelated healthy individuals were enrolled from the same hospital. The details of the IPF diagnosis, enrollment criteria, clinical data acquisition and follow-up information are provided in the Supplementary Methods.

Peripheral venous blood samples were collected from all subjects in EDTA-containing anticoagulant tubes. The demographic characteristics of the study subjects are shown in Supplementary Table 1.

PCR-based *AluYb8MUTYH* genotyping

Genomic DNA was extracted from the sample cells using the TIANamp Genomic DNA Kit (TIANGEN) according to the manufacturer's protocol. The *AluYb8* insertion in intron 15 of the *MUTYH* gene was detected by PCR as described in a previous study [9].

Analysis of the mtDNA content, mtDNA integrity and target gene expression

We used the real-time PCR method to determine the mtDNA content of 206 IPF patients (71 with the *A/A* genotype, 72 with the *A/P* genotype and 63 with the *P/P* genotype) and 206 matched healthy controls (see the Supplementary Methods). A confirmatory experiment was completed using droplet-based digital PCR (ddPCR) (see the Supplementary Methods) to verify the correlation of the *AluYb8MUTYH* polymorphism with the mtDNA content in a proportion of the peripheral blood samples randomly selected from the IPF patients and age-matched controls.

To examine the mtDNA integrity, we performed long-range PCR amplification (see Supplementary Methods) of a specific mtDNA fragment (10 kb) and nDNA fragment (8.7 kb) from 105 IPF patients (35 with the *A/A* genotype, 35 with the *A/P* genotype and 35 with the *P/P* genotype) and 105 matched healthy controls.

We recruited 87 IPF patients (30 with the *A/A* genotype, 30 with the *A/P* genotype and 27 with the *P/P* genotype) and 87 matched healthy controls for analysis of mtDNA coding gene and mitochondrial-related gene expression using the RT-qPCR technique. The mRNA expression of the GFP reporter gene in the transfected cells was detected by the RT-qPCR method. The mRNA levels of the embedded neomycin-encoded gene within the *pEGFP-N1* vector and the housekeeping gene (β -actin) were used as references for the transfection experiments in this study (see the Supplementary Methods).

Cell culture, plasmid construction and transfection

Human cells (A549, HELF, HEK293T, LO2, DU145 and HCT15) and rodent cells (CHO and C2C12) were grown in suitable complete growth medium (see the Supplementary Methods) and maintained in a humidified incubator at 37°C with 5% (v/v) CO₂. To investigate whether polymorphic *AluYb8MUTYH* played a regulatory role in the *MUTYH* 5'-untranslated sequence at the translational level, the three conventional 5' exon sequences of *MUTYH* transcripts were cloned, including the 5'-untranslated exon sequences from the α -, β - and γ -type *MUTYH* transcripts (named α -5'Exon, β -5'Exon and γ -5'Exon, respectively). The three fragments were initially amplified with the designed forward and reverse primers (Supplementary Table 3) and then inserted into the *pEGFP-N1* vector using *Bgl* II and *Pst* I restriction sites. The three *pEGFP-N1*-based recombinant plasmid constructs (named *pAlpha-5'Exon/EGFP*, *pBeta-5'Exon/EGFP*, and *pGamma-5'Exon/EGFP*) were confirmed by sequencing. The cultured cells were grown to 60–80% confluence and transfected with the recombinant and control plasmids (*pEGFP-N1*) using an effective transfection reagent (Lipofectamine 3000, Invitrogen). The cells were harvested for analysis after 48 hours in culture.

Immunoblotting

Tissue or cell lysates were prepared with RIPA lysis buffer (Thermo Scientific). The supernatant was collected and denatured by heating at 95°C for 10 min. Expression of *MUTYH* and the *EGFP* reporter gene was measured by immunoblotting with an anti-*MUTYH* (BS2535, Bioworld) antibody and GFP-tag (7G9) mouse mAb (Abmart), respectively. The blot was developed with an enhanced chemiluminescent (ECL) detection system (Millipore). Endogenous β -actin expression was used as an internal control.

Statistical analysis

The statistical calculations were performed using the SPSS Statistics 19.0 software package. The data are

expressed as the mean \pm SD where indicated. Separate variable comparisons among subjects with different *AluYb8MUTYH* genotypes were conducted with a nonparametric Kruskal-Wallis test or one-way ANOVA, followed by a post hoc analysis. The categorical variables were analyzed using the Chi-square test or Fisher's exact test. Significance was assumed when $P < 0.05$.

Abbreviations

IPF: idiopathic pulmonary fibrosis; mtDNA: mitochondrial DNA; UIP: usual interstitial pneumonia; ROS: reactive oxygen species; 8-oxoG: 8-oxo-7, 8-dihydroguanine; BER: base excision repair; MTS: mitochondrial targeting signal; FVC: forced vital capacity; FEV1: forced expiratory volume in 1 s; DLCO: diffusion capacity for carbon monoxide; SSB: single-strand break; mtQC: mitochondrial quality control; BAL: bronchoalveolar lavage; ECL: enhanced chemiluminescence.

AUTHOR CONTRIBUTIONS

WZ and WG designed the study, recruited the patients, performed most of the laboratory procedures, analyzed the data and drafted the manuscript. JS performed some of the laboratory procedures and analyzed the data. YZ and LX collected the cases, evaluated the patients and analyzed the data. YW designed the study and is responsible for the entire manuscript overall. All authors reviewed the manuscript.

ACKNOWLEDGMENTS

We are very grateful to Dr. Y Liu and Dr. X Yan (Nanjing Drum Tower Hospital) for their assistance with the clinical diagnoses, to JH Dai and HY Huang (Nanjing Drum Tower Hospital) for their assistance with the collection of the samples and with data management, to XX Li and CL Wang (Nanjing University School of Medicine) for technical assistance with DNA isolation and genotyping, and to Dr. YM Fan (Nanjing University School of Medicine) for statistical discussion.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This study was supported by the National Natural Science Foundation of China (grant number: 81270152, 81501205, 81771504).

REFERENCES

1. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, Swigris JJ, Taniguchi H, Wells AU. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers*. 2017; 3:17074. <https://doi.org/10.1038/nrdp.2017.74>
2. Nakabeppu Y, Sakumi K, Sakamoto K, Tsuchimoto D, Tsuzuki T, Nakatsu Y. Mutagenesis and carcinogenesis caused by the oxidation of nucleic acids. *Biol Chem*. 2006; 387:373–79. <https://doi.org/10.1515/BC.2006.050>
3. Maynard S, Schurman SH, Harboe C, de Souza-Pinto NC, Bohr VA. Base excision repair of oxidative DNA damage and association with cancer and aging. *Carcinogenesis*. 2009; 30:2–10. <https://doi.org/10.1093/carcin/bgn250>
4. Ayala-Peña S. Role of oxidative DNA damage in mitochondrial dysfunction and Huntington's disease pathogenesis. *Free Radic Biol Med*. 2013; 62:102–10. <https://doi.org/10.1016/j.freeradbiomed.2013.04.017>
5. Marzetti E, Csiszar A, Dutta D, Balagopal G, Calvani R, Leeuwenburgh C. Role of mitochondrial dysfunction and altered autophagy in cardiovascular aging and disease: from mechanisms to therapeutics. *Am J Physiol Heart Circ Physiol*. 2013; 305:H459–76. <https://doi.org/10.1152/ajpheart.00936.2012>
6. Fortini P, Pascucci B, Parlanti E, D'Errico M, Simonelli V, Dogliotti E. The base excision repair: mechanisms and its relevance for cancer susceptibility. *Biochimie*. 2003; 85:1053–71. <https://doi.org/10.1016/j.biochi.2003.11.003>
7. Parker AR, Eshleman JR. Human MutY: gene structure, protein functions and interactions, and role in carcinogenesis. *Cell Mol Life Sci*. 2003; 60:2064–83. <https://doi.org/10.1007/s00018-003-3053-4>
8. Banda DM, Nuñez NN, Burnside MA, Bradshaw KM, David SS. Repair of 8-oxoG:A mismatches by the MUTYH glycosylase: Mechanism, metals and medicine. *Free Radic Biol Med*. 2017; 107:202–15. <https://doi.org/10.1016/j.freeradbiomed.2017.01.008>
9. Sun C, Chen H, Guo W, Zhang K, Qi Q, Gu X, Zhu D, Wang Y. A common mutation of the MYH gene is associated with increased DNA oxidation and age-related diseases. *Free Radic Biol Med*. 2010; 48:430–36. <https://doi.org/10.1016/j.freeradbiomed.2009.11.015>
10. Guo W, Zheng B, Cai Z, Xu L, Guo D, Cao L, Wang Y. The polymorphic AluYb8 insertion in the MUTYH gene is associated with reduced type 1 protein expression and reduced mitochondrial DNA content. *PLoS One*. 2013; 8:e70718. <https://doi.org/10.1371/journal.pone.0070718>
11. Guo W, Zheng B, Guo D, Cai Z, Wang Y. Association of AluYb8 insertion/deletion polymorphism in the MUTYH gene with mtDNA maintain in the type 2 diabetes mellitus patients. *Mol Cell Endocrinol*. 2015; 409:33–40. <https://doi.org/10.1016/j.mce.2015.03.019>
12. Chen H, Sun C, Guo W, Meng R, Du H, Qi Q, Gu X, Li L, Zhang K, Zhu D, Wang Y. AluYb8 insertion in the MUTYH gene is related to increased 8-OHdG in genomic DNA and could be a risk factor for type 2 diabetes in a Chinese population. *Mol Cell Endocrinol*. 2011; 332:301–05. <https://doi.org/10.1016/j.mce.2010.11.021>
13. Faner R, Rojas M, Macnee W, Agustí A. Abnormal lung aging in chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2012; 186:306–13. <https://doi.org/10.1164/rccm.201202-0282PP>
14. Takao M, Zhang QM, Yonei S, Yasui A. Differential subcellular localization of human MutY homolog (hMYH) and the functional activity of adenine:8-oxoguanine DNA glycosylase. *Nucleic Acids Res*. 1999; 27:3638–44. <https://doi.org/10.1093/nar/27.18.3638>
15. Kazak L, Reyes A, Holt IJ. Minimizing the damage: repair pathways keep mitochondrial DNA intact. *Nat Rev Mol Cell Biol*. 2012; 13:659–71. <https://doi.org/10.1038/nrm3439>
16. Prakash A, Doublé S. Base Excision Repair in the Mitochondria. *J Cell Biochem*. 2015; 116:1490–99. <https://doi.org/10.1002/jcb.25103>
17. Oka S, Ohno M, Tsuchimoto D, Sakumi K, Furuichi M, Nakabeppu Y. Two distinct pathways of cell death triggered by oxidative damage to nuclear and mitochondrial DNAs. *EMBO J*. 2008; 27:421–32. <https://doi.org/10.1038/sj.emboj.7601975>
18. Oka S, Nakabeppu Y. DNA glycosylase encoded by MUTYH functions as a molecular switch for programmed cell death under oxidative stress to suppress tumorigenesis. *Cancer Sci*. 2011; 102:677–82. <https://doi.org/10.1111/j.1349-7006.2011.01869.x>
19. Ni HM, Williams JA, Ding WX. Mitochondrial dynamics and mitochondrial quality control. *Redox Biol*. 2015; 4:6–13. <https://doi.org/10.1016/j.redox.2014.11.006>
20. Kotiadis VN, Duchon MR, Osellame LD. Mitochondrial quality control and communications with the nucleus are important in maintaining mitochondrial function and cell health. *Biochim Biophys Acta*. 2014; 1840:1254–65. <https://doi.org/10.1016/j.bbagen.2013.10.041>
21. Raimundo N. Mitochondrial pathology: stress signals from the energy factory. *Trends Mol Med*. 2014;

20:282–92.

<https://doi.org/10.1016/j.molmed.2014.01.005>

22. Ryu C, Sun H, Gulati M, Herazo-Maya JD, Chen Y, Osafo-Addo A, Brandsdorfer C, Winkler J, Blaul C, Faunce J, Pan H, Woolard T, Tzouveleki A, et al. Extracellular Mitochondrial DNA Is Generated by Fibroblasts and Predicts Death in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2017; 196:1571–81. <https://doi.org/10.1164/rccm.201612-2480OC>

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY METHODS

Subjects

The enrollment criteria for the IPF patients were based on the guidelines of the American Thoracic Society (ATS)/European Respiratory Society (ERS)/Japanese Respiratory Society (JRS)/Association of Latin American Thorax (ALAT) in 2011 [1]. We collected the following clinical data at our institution: age, gender, past medical history, smoking history, occupational exposure history, physical examination findings, laboratory results, and high-resolution computed tomography (HRCT) scans. Respiratory specialists diagnosed the recruited IPF patients based on clinical features and HRCT evaluation. Patients with usual interstitial pneumonia of known causes, including collagen vascular disease and environmentally induced pneumonia, were excluded from this study. We also collected pulmonary function test data for the hospitalized IPF patients. The physiological assessment included measurements of the forced vital capacity (FVC), forced expiratory volume in the first second (FEV1) and single breath diffusing capacity for carbon monoxide according to the ATS recommendations. We conducted a continuous follow-up of the IPF patients. Overall survival was defined as the time from the first diagnosis until death from any cause. However, 67 of the 277 patients were lost to follow up (non-response).

All recruited healthy controls were free from other clinical diseases, including acute inflammation, cancer, cardiovascular diseases, diabetes and any neurological or psychiatric disorders, as assessed by their medical history, physical examination and blood chemistry.

Quantification of the mtDNA content by real-time quantitative PCR

We used real-time quantitative PCR (RT-qPCR) to determine the mtDNA content relative to that of the nDNA by measuring the ratio of the encoding mitochondrial genes *MT-TL1* and *MT-ND1* and the nuclear gene β -actin. The primers used for the mtDNA content assay are listed in the Supplementary Data (Supplementary Table 3). The SYBR kit (Vazyme) and ABI StepOne Real-Time System (Applied Biosystems) were used to measure the mtDNA content. The RT-qPCR was carried out in a 20 μ l total reaction volume containing 2 μ l (50 ng) of DNA, 10 μ l of AceQ[®] qPCR SYBR[®] Green Master Mix, 0.4 μ l of each primer and 7.2 μ l of sterile purified water. All mixtures were subjected to 5 min of pre-denaturation at 95°C, follow-

ed by 40 cycles of 10 seconds at 95°C and 30 seconds at 60°C. All samples were analyzed in triplicate. The average threshold cycle number (Ct) values of the nDNA and mtDNA were obtained from each case to determine the relative mtDNA quantities in each blood sample. We used the delta Ct ($\Delta C_t = C_{tmtDNA} - C_{tnDNA}$) in the same well as an exponent of 2 ($2^{\Delta C_t}$) to calculate the mtDNA content.

Analysis of the mtDNA copy number by droplet-based digital PCR (ddPCR)

To determine the mtDNA copy number, ddPCR was performed in the present study using samples obtained from the IPF patients and healthy controls. The ddPCR reaction mixtures were prepared in a 20 μ l volume and contained 0.1 ng of genomic DNA, 1 \times ddPCR supermix, 0.5 μ M of each primer (Supplementary Table 3), and 0.25 μ M of each probe (5'- FAM CCA TCA CCC TCT ACA TCA CCG CCC TAMRA-3' for *ND-1*) in water-in-oil emulsions using the droplet generator (Bio-Rad). Then, the generated droplets were transferred to a 96-well plate and amplified by a PCR thermocycler instrument. The amplification conditions were 1 cycle for 10 min at 95°C, followed by 40 cycles of 30 s at 94°C and 60 s at 60°C, a final step for 10 min at 98°C and cooling to 10°C. Subsequently, the PCR products were detected by the QX200 droplet reader (Bio-Rad), and data analysis was performed using the Quanta Soft analysis software. One nanogram of human genomic DNA is roughly equal to 300 copies of the haploid genome.

Evaluation of the mtDNA integrity

The primers used for the assay are listed in Supplementary Table 3. Using the PicoGreen fluorescent dye (Invitrogen), the long-range PCR product and the relative amounts of mtDNA/nDNA could be quantified as described previously [2]. Each PCR analysis was performed in duplicate.

Analysis of mitochondrial genes and mitochondrial-related gene expression

Peripheral blood mononuclear cells were isolated from whole blood samples from the IPF patients and healthy controls with lymphocyte separation medium. Total RNA was extracted from the isolated PBMCs and harvested cells with RNAiso Plus (Takara) and reverse transcribed to cDNA using the HiScript[®] II One Step RT-PCR Kit (Vazyme) according to the manufacturer's instructions. The AceQ[®] qPCR SYBR[®] Green Kit and ABI Viia 7 Real-Time System (Applied Biosystems) were used to amplify and assess the mRNA expression levels of target genes. The mRNA quantification

results were normalized to the reference gene expression level calculated with the $2^{\Delta Ct}$ method. *ATP6* and *COX2* mRNA expression was used to assess the mtDNA transcriptional activity. *POLG*, *MFN2* and *ATG7* expression was used to evaluate mtDNA replication, mitochondrial fusion and mitophagy, respectively. *GFP* and *neomycin* gene expression was used to detect the reporter gene transcription and transfection efficiencies in the transfected cells, respectively. The primer sequences for these assays are listed in Supplementary Table 3.

Cell culture

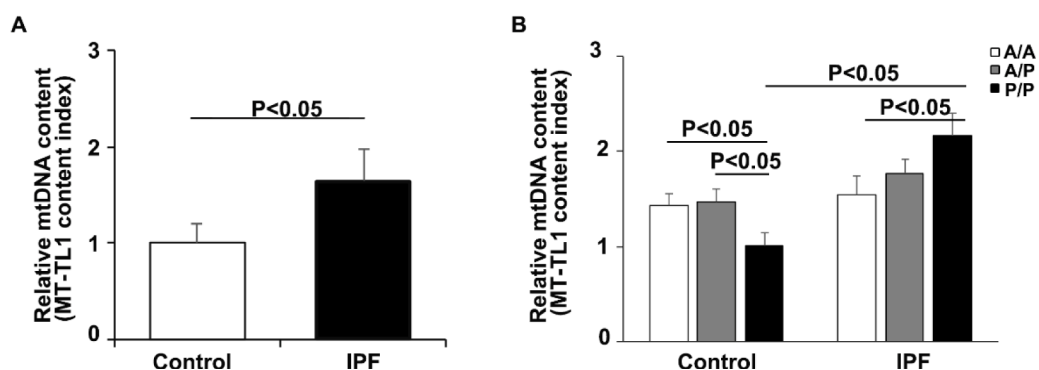
The human lung carcinoma epithelial A549 and Chinese hamster ovary (CHO) cell lines were cultured in F12K medium supplemented with 10% (v/v) fetal bovine serum (FBS). The human embryonic lung fibroblast (HELFL), human embryonic kidney 293T (HEK293T), human hepatocyte LO2, human prostate cancer DU145 and mouse myoblast (C2C12) cell lines were grown in Dulbecco's Modified Eagle's

Medium (DMEM) containing 10% (v/v) FBS. HCT15 on adenocarcinoma cells were propagated in RPMI-1640 medium with 10% (v/v) FBS. All cultures were supplemented with antibiotics (100 U/ml of penicillin and 100 µg/ml of streptomycin), and the cells were seeded in cell culture flasks and maintained in a humidified incubator at 37°C with 5% (v/v) CO₂.

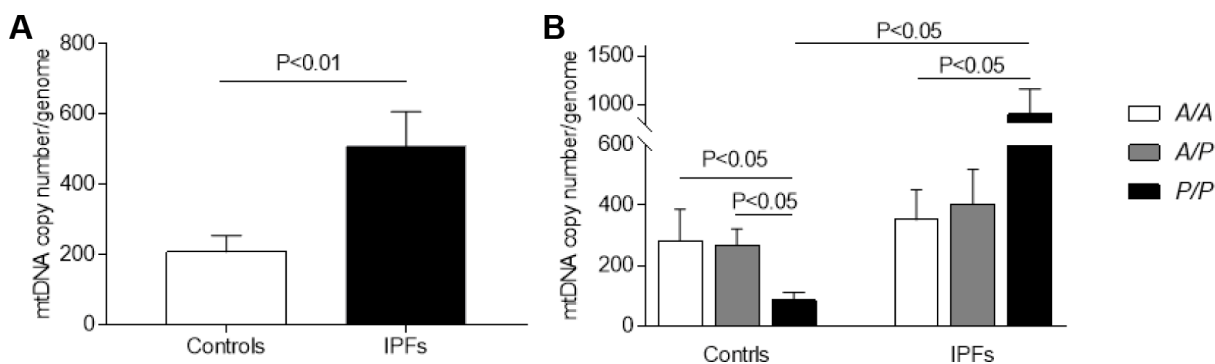
SUPPLEMENTARY REFERENCES

1. 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; 183:788–824.
<https://doi.org/10.1164/rccm.2009-040GL>
2. Guo W, Zheng B, Guo D, et al. Association of AluYb8 insertion/deletion polymorphism in the MUTYH gene with mtDNA maintain in the type 2 diabetes mellitus patients. *Mol Cell Endocrinol.* 2015;409:33–40.
<https://doi.org/10.1016/j.mce.2015.03.019>

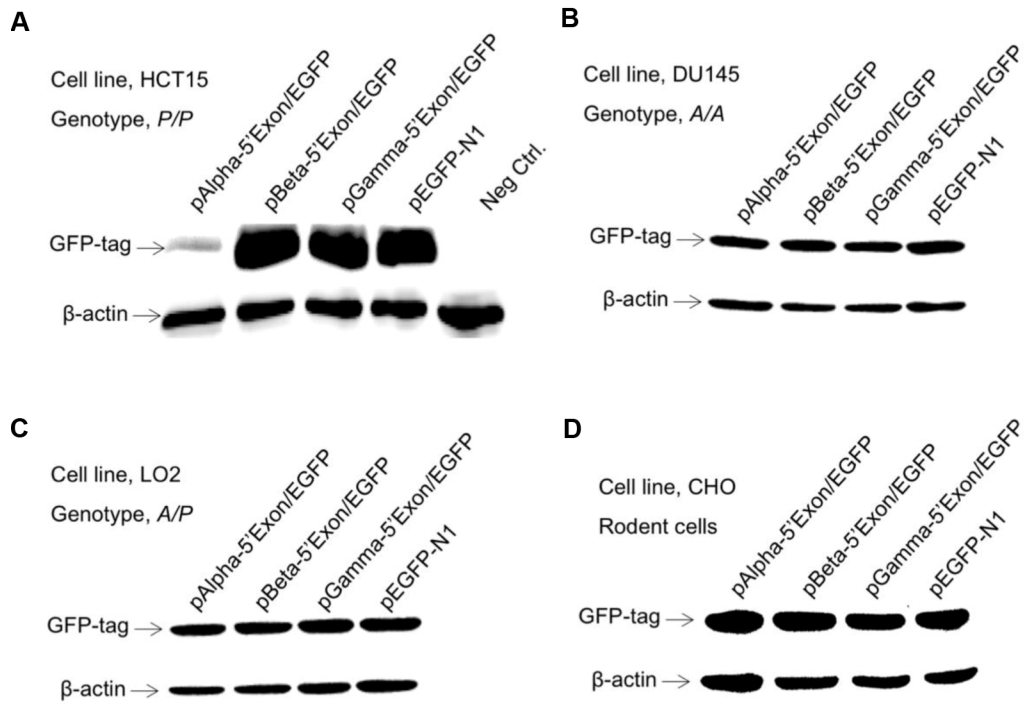
SUPPLEMENTARY FIGURES



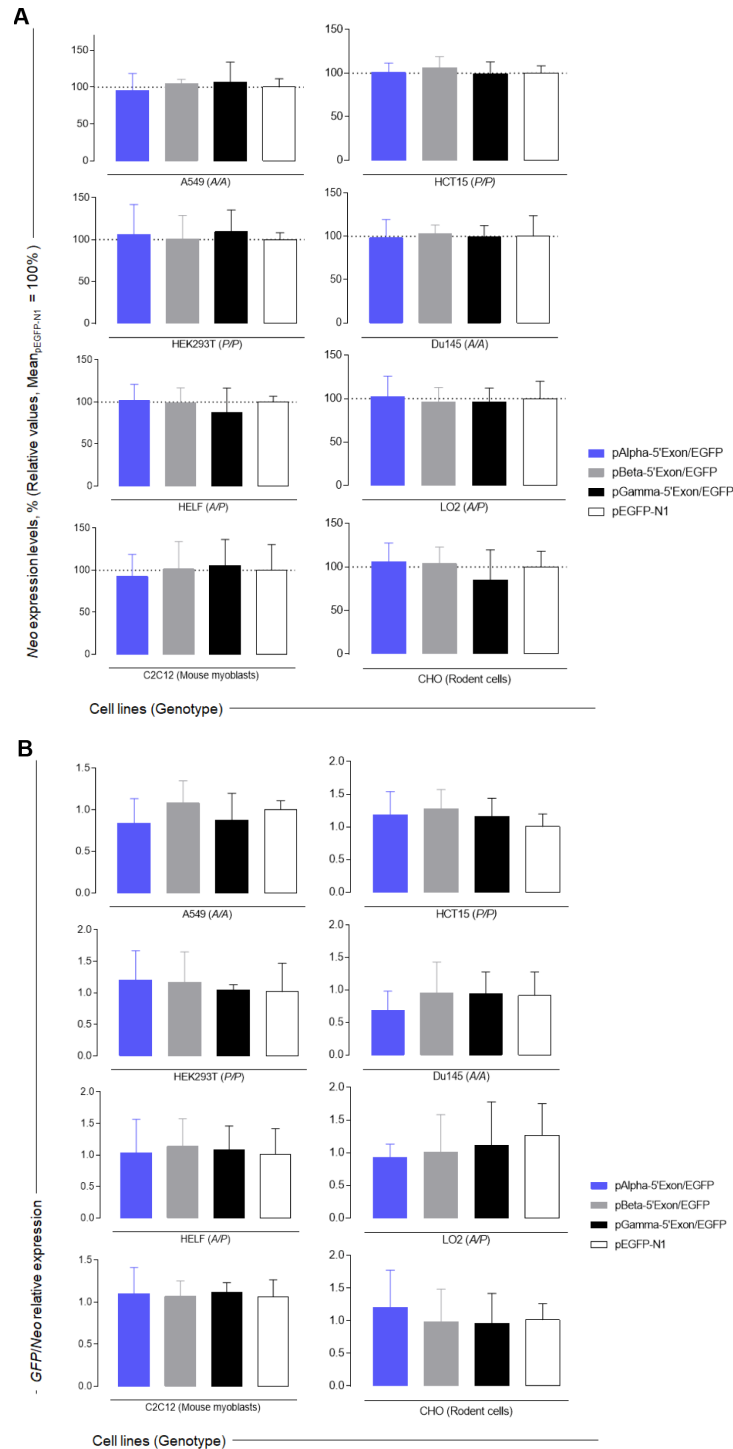
Supplementary Figure 1. Relationship between the *AluYb8MUTYH* genotype and mtDNA content in the IPF patients and healthy controls. (A) The *MT-TL1* content index was increased in IPF patients, $P < 0.05$. (B) In patients with *P/P*, the *MT-TL1* content index was significantly higher than that in healthy controls with *P/P*, $P < 0.05$.



Supplementary Figure 2. Result of the relationship between *AluYb8MUTYH* polymorphism and mtDNA copy number in the healthy control and IPF patient subjects by droplet digital PCR assay. (A) The cells of IPF patients have a significantly higher level of mtDNA than that of age-matched controls, $P < 0.01$. (B) The controls with *P/P* genotype had decreased levels of mtDNA content when compared to wild-type (*A/A*) and heterozygote (*A/P*) controls, and the IPF patients with *P/P* genotype are just the opposite, $P < 0.05$. The result from droplet digital PCR assay is in accordance with the result of relative quantifying mtDNA contents by SYBR green determination. One nanogram of human genomic DNA is roughly equal to 300 copies of the haploid genome.



Supplementary Figure 3. Immunoblotting results of GFP expression in reporter gene system. Highly expressed GFP reporter protein of the recombinant pAlpha-5' Exon/EGFP were observed in DU145 (A/A genotype), LO2 (A/P genotype) human cell lines and CHO (Chinese hamster cell line). However, the recombinant pAlpha-5' Exon/EGFP vector expressed GFP reporter protein at a lower level in the human cells (HCT15) with the mutant (P/P) genotype. Additionally, the GFP reporter of the recombinant vectors, pBeta-5'Exon/EGFP and pGamma-5'Exon/EGFP, were highly expressed in all cultured cells. β -actin was kept as protein loading control.



Supplementary Figure 4. The mRNA expression analysis for *GFP* and *Neo* genes of the recombinant vectors in all cultured cells. (A) No statistically significant difference in the mRNA expression levels of the embedded *Neo* gene among the cultured cells with different recombinant vectors. The relative expressions of *neomycin* and a host housekeeping gene (β -*actin*) reflect plasmid transfection efficiency of the recombinant vectors in the experimental cell lines. The results indicate there is no significant difference in the transfection efficiency for the different recombinant vectors in all the experimental cell lines under the standard transfection condition. The average value of the *Neo* relative expression (*Neo*/ β -*actin*) of cultured cells with the *pEGFP-N1* vector was set to 100. Data are mean \pm s.d. (B) Among the experimental cell lines with the different recombinant vectors, the difference was not statistically significant in the mRNA expression levels of *GFP* reporter gene ($P > 0.05$). An embedded *neomycin* encoded gene within the *pEGFP-N1* vector was used as the reference for the *GFP* reporter gene. Data are mean \pm s.d.

SUPPLEMENTARY TABLES

Supplementary Table 1. The demographic characteristics of the study subjects.

	Healthy controls	IPF patients
Number	810	277
Age*	53.96±14.08	61.38±12.60
Male	421	240

* The data represents the mean ± SD.

Supplementary Table 2. Allele frequency of *AluYb8MUTYH* in IPF and healthy control.

	Allele frequency		<i>P</i> value	OR(95%CI)
	IPF	Healthy control		
A	325 (58.7%)	892 (55.1%)	0.15	0.86 (0.71-1.05)
P	229 (41.3%)	728 (44.9%)		