

# Population genetics of *IFITM3* in Portugal and Central Africa reveals a potential modifier of influenza severity

Susana David<sup>1,2</sup> · Vanessa Correia<sup>3</sup> · Liliana Antunes<sup>4</sup> · Ricardo Faria<sup>1</sup> · José Ferrão<sup>1</sup> · Paula Faustino<sup>1,5</sup> · Baltazar Nunes<sup>4</sup> · Fernando Maltez<sup>6</sup> · João Lavinha<sup>1,7</sup> · Helena Rebelo de Andrade<sup>3,8</sup>

Received: 18 April 2017 / Accepted: 4 August 2017 / Published online: 25 August 2017  
© Springer-Verlag GmbH Germany 2017

**Abstract** Influenza epidemics are a serious global public health and economic problem. The *IFITM3* allele (rs12252-C) was suggested as a population-based genetic risk factor for severe influenza virus infection by A(H1N1)pdm09. We analyzed the population genetics of *IFITM3* variants in the Portuguese general population ( $n = 200$ ) and Central Africans (largely Angolan) ( $n = 148$ ) as well as its association to influenza severity in Portuguese patients ( $n = 41$ ). Seven SNPs, within the 352 bp *IFITM3* amplicon around rs12252, were identified. SNP distributions in the Portuguese appeared at an intermediate level between the Africans and other Europeans. According to HapMap, rs34481144 belongs to the same linkage disequilibrium (LD) block as rs12252 and is in strong LD with rs6421983. A negative association with severe relative to mild disease was observed for allele rs34481144-A, indicating a protective effect under the dominant model. Moreover, haplotype Hap4 with rs34481144-A,

not including rs12252-C, was significantly associated to mild influenza. Conversely, although with borderline significance, haplotype Hap1 with rs34481144-G, not including rs12252-C, was associated to severe disease. Moreover, in comparison to the general Portuguese population, statistical significant differences in the frequencies of the protective allele rs34481144-A in the severe disease group, the deleterious Hap1 in the mild disease group, and the protective Hap4 in the severe disease group were observed. The population attributable risk (PAR) for the targeted rs34481144 allele or genotype was of 55.91 and 64.44% in the general population and the mildly infected individuals, respectively. Implication of these variants in disease phenotype needs further validation, namely through functional analysis as is discussed.

**Keywords** *IFITM3* gene · Influenza virus · Population genetics · Portugal · Central Africa

**Electronic supplementary material** The online version of this article (doi:10.1007/s00251-017-1026-2) contains supplementary material, which is available to authorized users.

✉ Susana David  
suzana.david@insa.min-saude.pt

<sup>1</sup> Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Av. Padre Cruz, 1649-016 Lisbon, Portugal

<sup>2</sup> Centro de Estudos de Ciência Animal, Instituto de Ciências e Tecnologias Agrárias e Agro-Alimentares, Universidade do Porto, Porto, Portugal

<sup>3</sup> Departamento de Doenças Infeciosas, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisbon, Portugal

<sup>4</sup> Departamento de Epidemiologia, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisbon, Portugal

<sup>5</sup> ISAMB, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal

<sup>6</sup> Serviço de Doenças Infeciosas, Hospital de Curry Cabral, Centro Hospitalar de Lisboa Central, Lisbon, Portugal

<sup>7</sup> BioISI, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

<sup>8</sup> Centro de Patogénese Molecular—Retrovirus e Infecções Associadas (CPM-URIA), Faculdade de Farmácia, Universidade de Lisboa, Lisbon, Portugal

## Introduction

Annual epidemics of human influenza cause significant morbidity and mortality worldwide. Three to five million people experience severe illness and 0.25 to 0.5 million people die of influenza yearly (<http://www.who.int/mediacentre/factsheets/2003/fs211/en/>). Frequent complications include lower respiratory tract infections (e.g., pneumonia) that may result in hospitalization and death.

In industrialized countries, severe influenza is generally limited to high-risk groups (elderly, chronically ill). However, major changes in the viral genetic makeup, namely in the influenza A virus, have resulted in pandemics associated with severe outcomes also amongst healthy younger age groups. The global health and economic burdens represented by these pandemics have led to increased efforts to understand why some people develop severe disease while others do not. Susceptibility to severe seasonal or pandemic influenza in humans is likely to be polygenic and codetermined by pathogen characteristics, prior infection history, comorbidities, and environmental factors (Horbey et al. 2012). Animal studies suggest that genetic control of susceptibility to severe influenza is complex and not controlled by a single locus, but some of the host genetic determinants of susceptibility to severe disease may be common across influenza viral subtypes (Horbey et al. 2012).

Host type I interferons (IFNs) are cytokines essential to survive acute viral infection through stimulation of antiviral effectors (Trinchieri 2010). Amongst the effectors of type I IFNs, the interferon-induced transmembrane proteins, IFITM1, 2, and 3, are involved in (i) virus sequestration, (ii) virus receptor blockade, and (iii) endocytosis/fusion restriction, thus inhibiting the replication of several types of viruses including the influenza Orthomyxovirus (Schoggins et al. 2011).

The *IFITM3* SNP rs12252-C allele, which translates into a truncated protein missing the first 21 amino-terminal amino acid residues, has been associated to severe pandemic influenza A(H1N1)pdm09 virus infection in patients from the UK and China (Everitt et al. 2012; Zhang et al. 2013; Yang et al. 2015) as well as to mild influenza, but not severe A(H1N1)pdm09 infection, in a European cohort (Mills et al. 2013). Population genetics studies suggested that this allele, rare in European ancestry but very common in the Han Chinese and Japanese, is a potentially strong population-based risk factor (Zhang et al. 2013).

In the 1000 genomes allele frequency database (The 1000 Genomes Project Consortium 2012), the prevalence of SNP rs12252-C is heterogeneously distributed amongst populations worldwide. This is true for the Iberian populations and most African populations. The latter are particularly relevant, as in tropical regions of the developing world, viral transmission continues year round with high fatality rates (<http://www.who.int/mediacentre/factsheets/2003/fs211/en/>).

As part of the current global endeavor to understand how host genetic determinants modulate the response to influenza virus infection, this study aimed to contribute to the population genetics of *IFITM3* genetic variation around the rs12252 locus in the Portuguese general population and Central African (largely Angolan) populations and to investigate its association to influenza severity in Portuguese patients.

## Materials and methods

### Subjects

For the population genetics, the samples were convenience samples obtained from an anonymized collection of the Department of Human Genetics, Instituto Nacional de Saúde Doutor Ricardo Jorge, and included 200 Portuguese general population and 148 Central African individuals, for 57% (85/148) of which at least one of the parents was Angolan.

For genetic association studies with influenza severity, 41 Portuguese patients with laboratory confirmed influenza A(H1N1)pdm09 infection were selected and classified either as “mild cases” receiving ambulatory treatment only or as “severe cases” requiring admission to intensive care units (ICUs). Criteria for admission to ICUs included pneumonia or dyspnea. Chronic medical conditions, identified as potential A(H1N1)pdm09 risk factors (Van Kerkhove et al. 2011), in these patients included immunosuppression (cancer, HIV infection, Crohn’s disease, and organ transplant), diabetes, chronic respiratory conditions, e.g., asthma, and cardiovascular disease. For influenza laboratory diagnosis, viral nucleic acid extraction from oropharyngeal swabs was performed using the automated extractor EasyMag (bioMérieux, Lyon, France) according to the manufacturer’s recommendations. The presence of influenza virus was searched for by real-time RT-PCR, using a protocol kindly provided by the Health Protection Agency, UK. All the samples collected from the hospitalized and community patients and associated data were anonymized prior to investigation.

### Genotyping

Human genomic DNA was extracted from peripheral leukocytes using MagNA Pure LC (Roche Diagnostics GmbH, Mannheim, Germany) and from nasopharyngeal swabs using the QIAamp® Viral RNA Mini Kit (Qiagen, Hilden, Germany). PCR amplification of the 352 bp fragment in the first exon of the *IFITM3* gene (DNA position chr11: 320,556–320,907, GRCh38/hg38) was carried out using the forward and reverse primers: 5'-GGAAACTGTTGAGAAACCGA A-3' and 5'- CATA CGCACCTTCACGGAGT-3' (Zhang et al. 2013). The PCR products were purified and sequenced with BigDye Terminator v1.1 Sequencing Standard Kit

(Applied Biosystems, Portugal) and analyzed with the ABI 3130 XL Genetic Analyzer (Applied Biosystems, MA, USA). Sequences were analyzed using FinchLab™ (Geospiza). Haplotypes were reconstructed from the genotyping data using the PHASE, version 2.1, program (Stephens et al. 2001, Stephens and Donnelly 2003).

### Association testing

Association of genetic variants to disease severity was assessed by *IFITM3* genotyping of the individuals with severe influenza and those with mild influenza. The latter were considered to be a better control for the studies of association to severe influenza, rather than the general Portuguese population group, for which the status of influenza infection was unknown. However, the frequencies of the SNPs and haplotypes between each patients' group and the general Portuguese population were also compared. Hardy-Weinberg equilibrium (HWE) calculations were performed in the general population and influenza patient groups. Only SNPs under HWE and with minor allele frequency > 5% in both case and control groups were analyzed for association.

Statistical analysis of genetic data was performed using Rstats library in the R-Statistical package, R Development Core Team (<http://www.R-project.org>). The chi-square test was used for testing association under different genetic models. As the chi-square assumptions did not always hold, due to small sample size, the Fisher's Exact Probability Test (two-tailed) was also used. The odds ratio (OR) was calculated under different genetic models to measure the risk of severe influenza. To evaluate the possible confounding effects of age, gender, and chronic medical condition, the chi-square test was used to test for association of these factors to genotype as well as to mild and severe influenza. Estimates of the statistical powers for the effect size range of the chi-square test used in association testing in our patient population ( $n = 41$ ) were obtained. The chi-square test was also used for analyzing the frequencies of the SNPs and haplotypes between each patients' group and the general Portuguese population.

Results of the genotype-to-phenotype association testing were presented with the OR and confidence interval (CI) revealing the degree of precision in the size of the effect. In this study, the statistical significance must be interpreted with reserve due to the reduced sample size. Moreover, the expected size of the effect for common "low-penetrant" variants such as observed for SNP rs34481144 in the Portuguese population is small (Alcaïs et al. 2009), making it more difficult to detect. Therefore, we assumed, in the context of this preliminary study and in the light of the current interest in the exploration of the *IFITM3* promoter variants for their effect on viral influenza clinical phenotypes in populations of different ancestries, that it was justifiable to also report the 10% significance levels, designated "borderline" as being suggestive but not

conclusive. For these the 90% CI of the OR was also calculated. It can further be recalled that the choice of the significance level does not prove that the null hypothesis is either right or wrong (Beaglehole et al. 1993). Moreover, to increase stringency, the False Discovery Rate criterion (FDR) was performed for the genetic models tested per SNP and for all the haplotypes tested, although, given the non-random basis for the study of the *IFITM3* genetic loci, correction for multiple testing would not be mandatory.

The population attributable risk (PAR), the fraction of the population risk of severe influenza outcome attributable to the studied allele or genotype, was estimated as follows:  $PAR\% = 100 \times p(1 - OR) / [p(1 - OR) + 1]$ , where  $p$  is the prevalence of the studied allele in the general population (or mild influenza patient group) and OR represents an estimate of the relative risk of a severe infection outcome in the population with the studied allele (or genotype) compared to the population that does not carry the allele.

### Gene expression

The biological significance of rs34481144, in altering gene expression levels, was tested in a HeLa cell system using a protocol adapted from Shen et al. (2013). For this a 347-bp region of the *IFITM3* promoter was amplified, using the forward 5'-TATACTGCAGCTAGCGAGCCCTGAACCGGGACAGTG-3' and reverse 5'-TATACTGCACTCGA GTGGTGTCCAGCGAAGACCAGC-3' primers, and cloned into a pGL3 basic plasmid (Promega, Madison, WI, USA) between the *NheI* and *XhoI* restriction sites. Site-directed mutagenesis, of the c.-23G>A (rs34481144) site in the pGL3-*IFITM3* construct, was carried out using the forward and reverse primers, 5'-CCAGTAACCCGACCACCGCTGGTCTT CGC-3' and 5'-GCGAAGACCAGCGGTGGTCCGGTTAC TGG-3', respectively, in a reaction mix containing KOD Hot Start buffer (1×) (Novagen TOYOBO), KOD Hot Start polymerase (1 U), MgSO<sub>4</sub> 1.5 mM (Novagen TOYOBO), dNTPs 10 mM (Promega, Madison, WI, USA), primers (0.3 μM each) and 250 ng plasmid DNA. DNA was submitted to denaturation at 95 °C for 10 min, followed by 12 cycles at 95 °C for 1 min, 55 °C for 1 min and 68 °C for 16 min, and a final extension step at 68 °C for 10 min. After the mutagenesis reaction, the original templates were digested with the *DpnI*, 1.5 h at 37 °C. The mutation was confirmed by DNA sequencing.

For cell culture and luciferase assays, HeLa cells were maintained in Dulbecco's modified Eagle's medium (Gibco BRL, USA) supplemented with 10% fetal bovine serum, penicillin (100 U/ml), and streptomycin (100 U/ml). Twenty-four-hour cultured HeLa cells, seeded in 48-well plates, were transfected with the constructed pGL3-*IFITM3* and the pGL4.70hCMV plasmids, using Lipofectamine 2000 (Invitrogen, USA) in GIBCO™ in Opti-MEM I Reduced-

Serum Medium (Gibco BRL, USA). Assays were performed on cells treated at 20 h, or not, with human recombinant IFN- $\gamma$  or IFN- $\alpha$  (Alpha 2a) (final concentration of 100 pg/ml and 100 ng/ml). Firefly luciferase and renilla luciferase (via pGL3-basic and pGL4.70hCMV, respectively) luminescences were sequentially measured using a Dual-Luciferase Reporter Assay System (Promega, USA). Results were expressed as the ratio of light units of firefly luciferase activity over light units of renilla luciferase activity. All experiments were performed in triplicate and repeated three times.

## Results

A total of seven SNPs were detected in the *IFITM3* 352 bp amplicon. These included four previously described variations, namely SNP rs34481144, c.-23G>A (5' UTR), two synonymous SNPs rs12252, c.42T>C (p.Ser14Ser) and rs11553885, c.165C>T (p.Pro55Pro), and a missense SNP rs1136853, c.9C>A (p.His3Gln). In addition, three synonymous variations were identified, namely c.33T>C (p.Pro11Pro), c.51C>T (p.Pro17Pro), and c.60T>C (p.Tyr20Tyr) (Supplementary Fig. 1). The c.33T>C (p.Pro11Pro) variant is a novel, previously unreported SNP in DNA position 320781, and was only found in heterozygosity once in the Portuguese mild influenza patient group. The c.51C>T and c.60T>C variants, previously referred as rs56323507 and rs56020216, respectively (Everitt et al. 2012), were detected in heterozygosity simultaneously in only one individual in the Central African group also harboring a rs11553885c.165C>T variant. The allele distribution of these SNPs in the study populations and 1000 genomes project (when applicable) is summarized in Table 1. Hardy-Weinberg equilibrium was verified in all instances except in Central Africans who present a statistically significant homozygote excess for SNP rs11553885 (Table 1).

Our general population genetics study contributed to better delineate the *IFITM3* genetic variation profile around rs12252 reported in the 1000 genomes project database. In fact, the allele and genotype distributions of SNPs rs12252, rs1136853, and rs34481144 place the Portuguese in an intermediate position between the other Europeans and the sub-Saharan Africans, whereas the Central Africans closely resemble the other African populations as reported in the 1000 genomes project (Supplementary Fig. 2; Table 1).

The association studies of allele and genotype frequencies with severe influenza relative to mild influenza were carried out for the three SNPs with minor allele frequency (MAF) > 5% in the Portuguese influenza A(H1N1) patients with severe and mild disease, namely rs12252, rs1136853, and rs34481144, assuming different genetic models (Supplementary Table 2). Results showed a statistical significant protective effect of the rs34481144-A allele against

severe influenza under the dominant model (OR = 0.26; 95% CI 0.07–0.97) and a borderline statistical significant effect for the overdominant model (OR = 0.27; 90% CI 0.09–0.82). Moreover, possibly due to lack of power (the maximum power in our study was 50%), these results did not resist the FDR criterion.

Since, from our analysis, age and gender were not associated with genotype, also confirming the lack of evidence that the *IFITM3* allele or genotype frequencies vary with these factors, there was no evidence of need to adjust for a confounding effect by these factors. Moreover, there was no adjustment for confounding by chronic medical conditions, as this factor was not associated with the disease severity phenotypes in our study. Given that the absence of the allele rs34481144-A appeared to represent an increased risk of severe disease, we estimated the PAR, the proportion of severe influenza cases that would be preventable if all the individuals in the population had the protective allele. According to our results, 55.91 and 64.44% of the risk of severe infection are attributable to the absence of the allele rs34481144-A, respectively, in the general population and the mildly infected individuals. No significant association was found with the other two SNPs, although there seems to be a trend of the allele rs12252-C being associated to increased risk of severe influenza disease, as previously reported in other populations.

Haplotype reconstruction data is summarized in the Supplementary Table 1. Twenty-one haplotypes were identified in the various population groups analyzed. Haplotype-1 (Hap1) was the most frequent haplotype across all the study groups, ranging from 0.46 to 0.70 in frequency. A number of haplotypes (5, 6, and 21) not observed in the Portuguese general population were detected in the Portuguese mild influenza patients albeit at low frequencies (< 3%). Three haplotypes (2, 8, and 20) were detected in severe influenza, whereas they were not observed in the mild influenza cases. Genetic association studies to increased risk of severe influenza were carried out only for the three major haplotypes, with frequencies above 5%, in the population groups under study, Portuguese general population and A(H1N1) influenza patients. For Hap1, as a deleterious genetic factor against influenza severity, an association with influenza severity, relative to mild influenza, was observed with borderline statistical significance (OR = 2.41; 90% CI 1.14–5.08) (Table 2). For haplotype-4 (Hap4), as a protective genetic factor against influenza severity, a statistical significant association was observed (OR = 0.29; 95% CI 0.10–0.82) (Fig. 1; Table 2). For Hap4, the chi-square test resisted the FDR criterion.

Analysis of the frequencies of SNPs between each patients' group and the general Portuguese population, showed, in a dominant model of inheritance, a borderline statistical significant difference in the distribution of the SNP rs34481144-A allele, in the severe disease group compared to that observed in the general Portuguese population ( $p = 0.071$ ). Concerning



**Table 1** Allele and genotype frequencies of *IFITM3* variants in Portuguese and Central African general populations and in influenza A(H1N1)pdm09 Portuguese patients

Variant	General population							Influenza A(H1N1)pdm09 Portuguese patients <sup>b</sup>		
	Genotype	East Asian <sup>a</sup> (n = 286)	Europe <sup>a</sup> (n = 379)	IBS <sup>a</sup> (n = 14)	PGP <sup>b</sup> (n = 200)	Africa <sup>a</sup> (n = 246)	Central Africa <sup>b</sup> (n = 148)	Total (n = 41)	Mild influenza <sup>c</sup> (n = 19)	Severe influenza <sup>d</sup> (n = 22)
rs12252 c.42T>C (p.Ser14Ser)	CC	0.3007 (86/286)	0.00 (0/379)	0.00 (0/14)	0.00 (0/200)	0.0610 (15/246)	0.0676 (10/148)	0.0244 (1/41)	0.0526 (1/19)	0.00 (0/22)
	CT	0.3811 (109/286)	0.0686 (26/379)	0.00 (0/14)	0.1200 (24/200)	0.3618 (89/246)	0.3919 (58/148)	0.1463 (6/41)	0.1053 (2/19)	0.1818 (4/22)
rs34481144 c.-23G>A (5' UTR)	TT	0.3182 (91/286)	0.9314 (353/379)	1.00 (14/14)	0.8800 (176/200)	0.5772 (142/246)	0.5405 (80/148)	0.8293 (34/41)	0.8421 (16/19)	0.8182 (18/22)
	Allele C	0.0343 (281/572)	0.0343 (26/758)	0.00 (0/28)	0.0600 (24/400)	0.2419 (119/492)	0.2635 (78/296)	0.0976 (8/82)	0.1053 (4/38)	0.2727 (6/22)
	Allele T	0.5087 (291/572)	0.9657 (732/758)	1.00 (28/28)	0.9400 (376/400)	0.7581 (373/492)	0.14	0.9024 (74/82)	0.8947 (34/38)	0.0909 (4/44)
	HWE <sup>e</sup>		0.815		0.815		0.0041	1.170		0.9091 (40/44)
	AA	0.00 (0/286)	0.2216 (84/379)	0.1429 (2/14)	0.1100 (22/200)	0.0041 (1/246)	0.00 (0/148)	0.0976 (4/41)	0.1053 (2/19)	0.0909 (2/22)
rs1136853 c.9C>A (p.His3Gln)	AG	0.4908 (186/379)	0.4908 (186/379)	0.4286 (6/14)	0.4450 (89/200)	0.1016 (25/246)	0.1757 (26/148)	0.4146 (17/41)	0.5789 (11/19)	0.2727 (6/22)
	GG	0.9930 (284/286)	0.2876 (109/379)	0.4286 (6/14)	0.4450 (89/200)	0.8943 (220/246)	0.8243 (122/148)	0.4878 (20/41)	0.3158 (6/19)	0.6364 (14/22)
	Allele A	0.0035 (2/572)	0.4670 (374/758)	0.3571 (10/28)	0.3325 (133/400)	0.0549 (27/492)	0.0878 (26/296)	0.3049 (25/82)	0.3947 (15/38)	0.2273 (10/44)
	Allele G	0.9965 (570/572)	0.5330 (404/758)	0.6429 (18/28)	0.6675 (267/400)	0.9451 (465/492)	0.9122 (270/296)	0.6951 (57/82)	0.6053 (23/38)	0.7727 (34/44)
	HWE <sup>e</sup>		0.001		0.001		1.372	0.019		0.00 (0/22)
rs11553885 c.165C>T (p.Pro55Pro)	AA	0.00 (0/286)	0.00 (0/379)	0.00 (0/14)	0.00 (0/200)	0.0081 (2/246)	0.0135 (2/148)	0.0244 (1/41)	0.0526 (1/19)	0.00 (0/22)
	AC	0.00 (0/286)	0.0475 (18/379)	0.00 (0/14)	0.1050 (21/200)	0.1179 (22/246)	0.1824 (27/148)	0.1463 (6/41)	0.1053 (2/19)	0.1818 (4/22)
	CC	1.00 (286/286)	0.9525 (361/379)	1.00 (14/14)	0.8950 (179/200)	0.8740 (215/246)	0.8041 (119/148)	0.8293 (34/41)	0.8421 (16/19)	0.8182 (18/22)
	Allele A	0.00 (0/572)	0.0237 (18/758)	0.00 (0/28)	0.0525 (21/400)	0.0671 (33/492)	0.1047 (31/296)	0.0976 (8/82)	0.1053 (4/38)	0.0909 (4/44)
SNP cd11 c.33T>C (p.Pro11Pro)	Allele C	1.00 (572/572)	0.9763 (740/758)	1.00 (28/28)	0.9475 (379/400)	0.9329 (33/492)	0.8953 (265/296)	0.9024 (74/82)	0.8947 (34/38)	0.9091 (40/44)
	HWE <sup>e</sup>		0.614		0.614		0.109	1.170		0.00 (0/22)
	TT	NA	NA	NA	0.00 (0/200)	NA	0.0203 (3/148)	0.00 (0/41)	0.00 (0/19)	0.00 (0/22)
	TC	NA	NA	NA	0.00 (0/200)	NA	0.1149 (17/148)	0.0488 (2/41)	0.0526 (1/19)	0.0455 (1/22)
rs56323507 c.51C>T (p.Pro17Pro)	CC	NA	NA	NA	1.00 (200/200)	NA	0.8649 (128/148)	0.9512 (39/41)	0.9474 (18/19)	0.9545 (21/22)
	Allele T	NA	NA	NA	0.00 (0/400)	NA	0.0777 (23/296)	0.0244 (2/82)	0.0263 (1/38)	0.0227 (1/44)
	Allele C	NA	NA	NA	1.00 (400/400)	NA	0.9223 (273/296)	0.9756 (80/82)	0.9737 (37/38)	0.9773 (43/44)
	HWE <sup>e</sup>		NA		NA		5.837	0.026		0.00 (0/22)
rs56020216 c.60T>C (p.Tyr20Tyr)	CC	NPD	NPD	NPD	0.00 (0/200)	NPD	0.00 (0/148)	0.00 (0/41)	0.00 (0/19)	0.00 (0/22)
	Allele T	NPD	NPD	NPD	0.00 (0/200)	NPD	0.00 (0/148)	0.00 (0/41)	0.00 (0/19)	0.00 (0/22)
	Allele C	NPD	NPD	NPD	1.00 (200/200)	NPD	1.00 (148/148)	0.9756 (40/41)	0.9474 (18/19)	1.00 (22/22)
	HWE <sup>e</sup>		NA		NA		NA	0.0122 (1/82)	0.0263 (1/38)	0.00 (0/44)
rs56020216 c.60T>C (p.Tyr20Tyr)	CC	NPD	NPD	NPD	0.00 (0/200)	NPD	0.00 (0/148)	0.00 (0/41)	0.00 (0/19)	0.00 (0/22)
	Allele T	NPD	NPD	NPD	0.00 (0/200)	NPD	0.00 (0/148)	0.00 (0/41)	0.00 (0/19)	0.00 (0/22)
	Allele C	NPD	NPD	NPD	1.00 (200/200)	NPD	0.9932 (147/148)	1.00 (41/41)	1.00 (19/19)	1.00 (22/22)
	HWE <sup>e</sup>		NA		NA		0.0034 (1/296)	0.00 (0/82)	0.00 (0/38)	0.00 (0/44)
rs56020216 c.60T>C (p.Tyr20Tyr)	CC	NPD	NPD	NPD	0.00 (0/200)	NPD	0.00 (0/148)	0.00 (0/41)	0.00 (0/19)	0.00 (0/22)
	Allele T	NPD	NPD	NPD	0.00 (0/200)	NPD	0.00 (0/148)	0.00 (0/41)	0.00 (0/19)	0.00 (0/22)
	Allele C	NPD	NPD	NPD	1.00 (200/200)	NPD	0.9932 (147/148)	1.00 (41/41)	1.00 (19/19)	1.00 (22/22)
	HWE <sup>e</sup>		NA		NA		0.9966 (295/296)	0.00 (0/82)	0.00 (0/38)	0.00 (0/44)

*IBS* Iberian Population in Spain (1000 genomes project), *PGP* Portuguese general population (this study), *fTWE* Hardy-Weinberg equilibrium, *NA* observed variant without frequency or population (1000 genomes project), *NPD* not previously described, *SNP* single nucleotide polymorphism, *cd* codon

<sup>a</sup> Allele frequency and genotypes obtained from the 1000 genomes project

<sup>b</sup> Allele frequency and genotypes obtained from the present study

<sup>c</sup> Mild infection corresponded to patients seen only in ambulatory care

<sup>d</sup> Severe infection corresponded to patients requiring admission to intensive care units

<sup>e</sup> Hardy-Weinberg equilibrium chi-square values obtained using data from the present study

**Table 2** Association of major haplotypes to increased risk of severe influenza

Contingency table		P value chi-square ( $\chi^2$ )	P value Fisher's exact test <sup>a</sup>	OR 95% CI
Haplotype-1	Haplotype-3			
Severe influenza	Present	0.051	0.075	2.41 (0.99–5.86) <sup>b</sup>
Mild influenza	28			
Severe influenza	Present	NA	0.463	0.00 (0.00–NaN)
Mild influenza	16			
Severe influenza	1	0.016	0.024	0.29 (0.10–0.82)
Mild influenza	7			
Severe influenza	15			
Mild influenza	23			

CI confidence interval, OR odds ratio, NA not applicable

Major haplotypes are those with frequencies above 5% in the population groups under study: Portuguese general population and A(H1N1) influenza patients. For haplotype definitions, refer to Table 1 of the supplementary materials

<sup>a</sup> The Fisher's exact test refers to the two-tailed Fisher's exact test

<sup>b</sup> The 90% confidence interval for 10% statistical significance (borderline): OR = 2.41; 90% CI 1.14–5.08

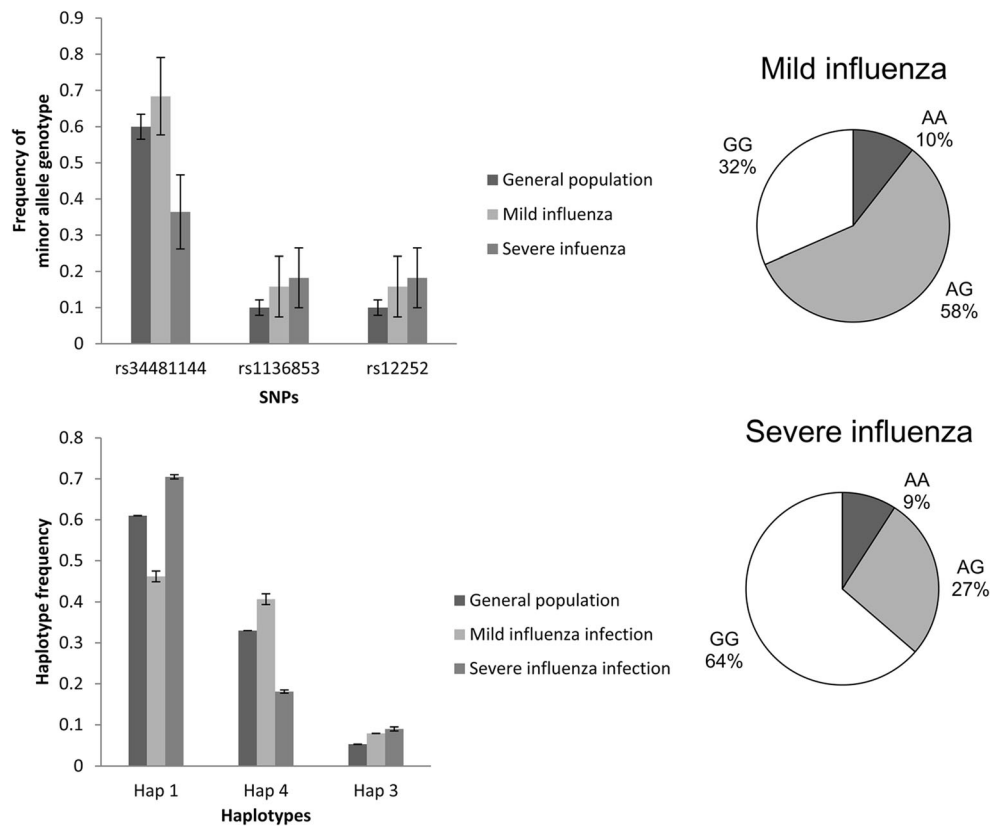
the haplotypes, the distribution of the deleterious Hap1 in the mild disease group was significantly different from that observed in the general Portuguese population ( $p = 0.017$ ) whereas the distribution of the protective Hap4 in the severe disease group was significantly different than that observed in the general Portuguese population ( $p = 0.016$ ).

Linkage disequilibrium (LD) patterns of genetic variants in the region around rs34481144 according to HapMap were visualized in the population with northern and western European ancestry (CEU) using Haploview (Barret et al. 2005). The plot showed this SNP as part of a LD block that included the SNP rs12252 (Fig. 2). This latter SNP was in perfect LD with rs1136853 ( $r^2 = 1$ ), as confirmed in our dataset. The targeted SNP rs34481144 was in strong LD with rs6421983 ( $r^2 = 0.88$ ). We first analyzed in silico the transcription factor binding preferences for the 5'UTR SNP rs34481144 using the JASPAR open-access database (Mathelier et al. 2014). The existence of a possible binding site (Zinc Finger Protein 354C), where there would be increased transcription with the allele rs34481144-A, justified the *IFITM3* in vitro transcription analysis of the major haplotypes, with and without this allele (Hap1 and Hap4, respectively). Under the conditions tested, no significant difference of activity was observed with this assay, either with or without interferon stimulation, IFN- $\alpha$  (Supplementary Fig. 3) or IFN- $\gamma$  (not shown). In silico analysis of the intronic rs6421983 c.249+171G>A and c.249+171G>C variants, using the Human Splicing Finder software (Desmet et al. 2009), did not reveal potential alterations of splicing.

## Discussion

The IFITM proteins are viral restriction factors affecting several stages of the virus replication cycle, namely viral-host receptor binding, endocytosis, or late endosome fusion and acidification (Brass et al. 2009; Weidner et al. 2010). In humans, these small proteins, of approximately 130 amino acids, are encoded by five genes: *IFITM1*, *IFITM2*, *IFITM3*, *IFITM5*, and *IFITM10*, located as a cluster on chromosome 11p15.5. The first three are considered type I IFN effectors (Lewin et al. 1991). However, *IFITM3* expression is also induced by multiple cytokine signals including type II IFNs (Bailey et al. 2012) which is suggestive of it playing a central role in protection against both viral (Schoggins et al. 2011; Brass et al. 2009; Weidner et al. 2010; Bailey et al. 2012; Jia et al. 2012) and bacterial infections (Shen et al. 2013). Several studies on human influenza have associated the host's IFITM proteins to protection against viral infection and subsequent disease severity (Everitt et al. 2012; Zhang et al. 2013; Yang et al. 2015; Mills et al. 2013; Brass et al. 2009; Weidner et al. 2010; Bailey et al. 2012; Jia et al. 2012).

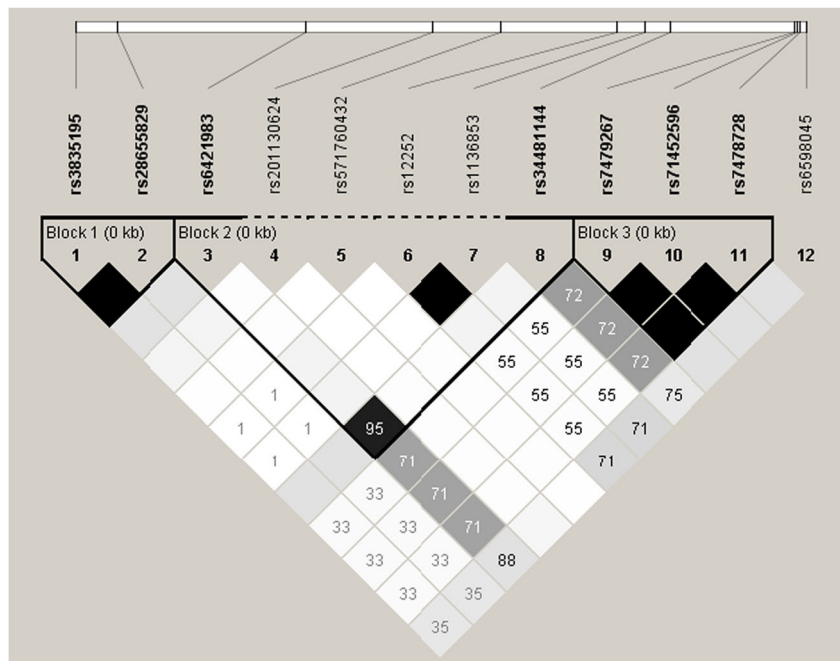
**Fig. 1** Histograms display allele and haplotype frequencies amongst Portuguese A(H1N1)pdm09 influenza patients. Error bars correspond to standard errors. The pie charts on the right represent the genotype distribution at SNP rs34481144 in the two influenza severity groups



The population genetics study we carried out on Portuguese general population, a Central African general population and Portuguese influenza A(H1N1)pdm09 patients, identified seven SNPs within a 352-bp *IFITM3* amplicon around rs12252. The general population study corroborated

the SNP distributions reported in the 1000 genomes project for all the SNPs described therein. Namely, the rs12252-C variant appears less common in the European and sub-Saharan populations than that reported in the Han Chinese, whereas the rs34481144-A variant appears to be more

**Fig. 2** LD patterns in the genomic region around rs34481144 according to HapMap, visualized using the Haploview software (Barret et al. 2005) in the population with ancestry from northern and western Europe (CEU). The second LD block contained the targeted SNP rs34481144 correlated with rs6421983 ( $r^2 = 0.88$ ) and two correlated SNPs rs12252 and rs1136853 ( $r^2 = 1.00$ ). The minor allele frequencies for these SNPs are as follows: 0.495 for rs34481144, 0.485 for rs6421983, and 0.045 for both rs12252 and rs1136853



common in populations of European descent and in sub-Saharan populations, having low prevalence in the Asian populations. Concerning the Portuguese population, it appears as an intermediate population between the Africans and the other Europeans reported in the 1000 genomes project. The departure from Hardy-Weinberg equilibrium of SNP rs11553885 in Central Africans, in which an excess of homozygotes was observed, may reflect one of several hypotheses whose relative likelihood would need to be assessed through power increment as well as demographic data on population distribution. These hypotheses include that the locus may be under selective pressure, that “null alleles” may be present, that inbreeding may be common in the population, or that the existence of a population substructure.

The onset of the influenza A(H1N1)pdm09 pandemic in 2009 led to a subsequent report on the association of an *IFITM3* variant at SNP rs12252 with severe influenza disease (Everitt et al. 2012). In their report, based on the analysis of 53 cases of severe A(H1N1)pdm09 or seasonal influenza virus infection in patients from the UK, these authors suggested an additive genetic model of association to the rs12252 minor allele C. However, a recessive model of association was reported in two ensuing studies; the first involving 32 cases of severe infection compared to 51 mild A(H1N1)pdm09 influenza cases in the Han Chinese (Zhang et al. 2013), and, the second, in an extensive population genetics study of the *IFITM3* gene, in a cohort of 34 severe A(H1N1) influenza cases (requiring intensive care unit admission for pneumonia) and over 5000 community acquired mild lower respiratory tract infection and matched controls of European ancestry (Mills et al. 2013).

In our genetic association study to influenza severity in the Portuguese, the SNP rs12252 minor frequency allele C was not significantly associated with severe influenza as previously reported (Everitt et al. 2012; Zhang et al. 2013; Yang et al. 2015; Mills et al. 2013). This could be due to lack of power in our study. Indeed from our calculations based on published data from a meta-analysis for rs12252 (Yang et al. 2015) and on our population association studies revealing an expected frequency of exposure of 6%, for a 95% confidence interval and an 80% probability of detection of the risk of severe vs. mild disease, we would require a sample of 586 individuals. The targeted SNP rs34481144 was in the same linkage disequilibrium (LD) block as rs12252 and showed strong LD with rs6421983. In reports from other groups, no reference was made to rs34481144, which we suggest to be protective against severe forms of influenza. Moreover, the novel polymorphism c.33T>C (p.Pro11Pro) reported in this study was not detected. This may have possibly been the result of methodological differences, e.g., the selected primers not amplifying the genome region where the SNP is located (Everitt et al. 2012) or because restriction fragment length polymorphism (RFLP) SNP detection used instead of sequencing, as in our

study, would not detect the polymorphisms herein reported nor novel variants (Mills et al. 2013). Also, the rs34481144 variant was probably not reported by Zhang and collaborators in the Han Chinese study because this SNP appears to be more common in populations of European descent, having low prevalence in the Asian populations (Zhang et al. 2013), which was also the case in the African populations as described in the 1000 genomes project and observed in our Central African population sample.

The genotype-to-phenotype association analysis, carried out for the most frequent haplotypes in the Portuguese general population in the severe relative to the mild influenza patient groups, revealed a statistical significant association to severe influenza (Hap1) or to mild influenza (Hap4), where this latter association resisted the FDR criterion (OR = 0.29; 95% CI 0.10–0.82). Noteworthy is the fact that this latter haplotype includes the protective allele rs34481144-A but not the susceptibility allele rs12252-C. Moreover, the results were suggestive of an association of Hap1, which includes the rs34481144-G allele but not the rs12252-C allele, to severe influenza at a borderline significance level (OR = 2.41; 90% CI 1.14–5.08). The association with influenza severity was also observed for the SNP rs34481144 where the minor frequency allele A indicated a protective effect against severe influenza under the dominant model (OR = 0.26; 95% CI 0.07–0.97). Moreover, also corroborating these results, when comparing the frequencies of the SNPs and haplotypes between each patients' group and the general Portuguese population, statistical significant differences were also observed, namely the frequency of the protective SNP rs34481144-A allele in the severe disease group was lower than expected from analysis of the general population, the frequency of the deleterious Hap1 in the mild disease group was lower than in the general population, and the frequency of the protective Hap4 in the severe disease group was lower than that observed in the general Portuguese population.

The opposite tendencies, observed in the prevalence of the minor alleles of the two SNPs associated on one hand with severe influenza A(H1N1) infection (rs12252-C) and on the other hand with protection against severe influenza A(H1N1) infection (rs34481144-A) in two of the major geographic regions studied in this context, namely Asia and Europe, respectively, suggest that influenza exposure may have at one time influenced these populations' allele frequencies. The proportion of severe cases that would be preventable if all the individuals in the population had the protective rs34481144-A allele was estimated as 55.91 and 64.44% in the general population and the mildly infected individuals, respectively.

The genetic and molecular mechanisms of the *IFITM3* that interfere with viral and bacterial cycles are not yet fully understood (Tartour and Cimarelli 2015; Ranjbar et al. 2015). Interestingly, Shen et al. (2013) identified a SNP in the *IFITM3* promoter (rs3888188) that responded to IFN- $\gamma$



stimulation, although not in the vicinity of the IFN- $\gamma$  activation site (GAS) element, and was associated with the risk for pediatric tuberculosis in Han Chinese population. In our study, the *IFITM3* promoter SNP rs34481144 was associated with severe influenza virus infection. While this SNP does not affect the consensus GAS or the IFN-stimulated response element (ISRE) sequences (Darnell et al. 1994), an in silico study of transcription factor binding preferences justified the in vitro transcription analysis of the major *IFITM3* haplotypes. However, no significant difference of activity was observed, either with or without IFN- $\alpha$  or IFN- $\gamma$  stimulation.

These findings point to important differences in the risk of developing severe disease due to A(H1N1) influenza virus infection in populations of different genetic background. Confirmation in larger and different population sets is necessary. Further studies of the SNP and loci in LD with rs34481144 (rs6421983) are needed to determine possible functional alterations, such as expression levels of IFITM3 in individuals with different genotypes, as this understanding can help to identify novel targets for health interventions aiming at influenza control.

**Acknowledgements** This study was carried out with financial support from FCT/MEC through national funds and co-financed by FEDER, under the Partnership Agreement PT2020, in the project with reference UIDMULTI/00211/2013. We acknowledge the clinical assistance of Drs. Madalena Almeida Santos and Maria José Silvestre from the Hospital Curry Cabral. We thank INSA's Unidade de Tecnologia e Inovação for technical assistance. We thank Patricia Barros, Rafaela Lacerda, and Juliane Menezes, Departamento de Genética Humana, INSA, and Luís Santos, Departamento de Doenças Infecciosas, INSA, as well as Stephanie David Bosne for fruitful discussions. Plasmid pGL3-control was kindly provided by Peter Jordan, Departamento de Genética Humana, INSA.

#### Compliance with ethical standards

**Conflict of interests** The authors declare that they have no conflict of interest.

**Ethical approval** For this type of study, formal consent is not required.

## References

- Alcaïs A, Abel L, Casanova JL (2009) Human genetics of infectious diseases: between proof of principle and paradigm. *J Clin Invest* 119:2506–2514. doi:10.1172/JCI38111
- Bailey CC, Huang I-C, Kam C, Farzan M (2012) Ifitm3 limits the severity of acute influenza in mice. *PLoS Pathog* 8:e1002909. doi:10.1371/journal.ppat.1002909
- Barret JC, Fry B, Maller J, Daly MJ (2005) Haplotype analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- Beaglehole R, Bonita R, Kjellstrom T (1993) Measuring health and disease. *Basic epidemiology*. World Health Organization, Geneva, pp 13–20
- Brass AL, Huang IC, Benita Y et al (2009) The IFITM proteins mediate cellular resistance to influenza A(H1N1) virus, west nile virus, and dengue virus. *Cell* 139:1243–1254
- Darnell JE Jr, Kerr IM, Stark GR (1994) Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* 264:1415–1421
- Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C (2009) Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37(9):e67. doi:10.1093/nar/gkp215
- Everitt AR, Clare S, Pertel T et al (2012) IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* 484:519–523
- Horby P, Nguyen NY, Dunstan SJ, Baillie JK (2012) The role of host genetics in susceptibility to influenza: a systematic review. *PLoS One* 7:e33180
- Jia R, Pan Q, Ding S, Rong L, Liu SL, Geng Y, Qiao W, Liang C (2012) The N-terminal region of IFITM3 modulates its antiviral activity by regulating IFITM3 cellular localization. *J Virol* 86:13697–13707
- Lewin AR, Reid LE, McMahon M, Stark GR, Kerr IM (1991) Molecular analysis of a human interferon-inducible gene family. *Eur J Biochem* 199:417–423
- Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen C-Y, Chou A, Lenasescu H, Lim J, Shyr C, Tan G, Zhou M, Lenhard B, Sandelin A, Wasserman WW (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 42(Database issue):D142–D147
- Mills TC, Rautanen A, Elliott KS et al (2013) IFITM3 and susceptibility to respiratory viral infections in the community. *J Infect Dis* 209:1028–1031
- Ranjbar S, Haridas V, Jasenosky LD et al (2015) A role for IFITM proteins in restriction of *Mycobacterium tuberculosis* infection. *Cell Rep* 13:874–883
- Schoggins JW, Wilson SJ, Panis M et al (2011) A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* 472:481–485
- Shen C, Wu XR, Jiao WW et al (2013) A functional promoter polymorphism of IFITM3 is associated with susceptibility to pediatric tuberculosis in Han Chinese population. *PLoS One* 8:e67816
- Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169
- Stephens M, Smith N, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- Tartour K, Cimarelli A (2015) Les IFITM, un obstacle commun à de nombreux virus. *Med Sci (Paris)* 31:377–382
- The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65
- Trinchieri G (2010) Type I interferon: friend or foe? *J. Exp Med* 207:2053–2063
- Van Kerkhove MD, Vandemaële KAH, Shinde V, Jaramillo-Gutierrez G, Koukounari A et al (2011) Risk factors for severe outcomes following 2009 influenza A (H1N1) infection: a global pooled analysis. *PLoS Med* 8:e1001053
- Weidner JM, Jiang D, Pan XB, Chang J, Block TM et al (2010) Interferon induced cell membrane proteins, IFITM3 and tetherin, inhibit vesicular stomatitis virus infection via distinct mechanisms. *J Virol* 84:12646–12657
- Yang X, Tan B, Zhou X, Xue J, Zhang X, Wang P et al (2015) Interferon-inducible transmembrane protein 3 genetic variant rs12252 and influenza susceptibility and severity: a meta-analysis. *PLoS One* 10:e0124985. doi:10.1371/journal.pone.0124985
- Zhang Y-H, Zhao Y, Li N et al (2013) Interferon-induced transmembrane protein-3 genetic variant rs12252-C is associated with severe influenza in Chinese individuals. *Nat Commun* 4:1418