

# SCIENTIFIC REPORTS

OPEN

## Whole Exome Sequencing reveals new candidate genes in host genomic susceptibility to Respiratory Syncytial Virus Disease

Antonio Salas<sup>1,2,3,4</sup>, Jacobo Pardo-Seco<sup>1,2,3,4</sup>, Miriam Cebey-López<sup>3,4</sup>, Alberto Gómez-Carballa<sup>1,2,3,4</sup>, Pablo Obando-Pacheco<sup>3,4</sup>, Irene Rivero-Calle<sup>3,4</sup>, María-José Currás-Tuala<sup>1,2,3,4</sup>, Jorge Amigo<sup>1,2,3,4</sup>, José Gómez-Rial<sup>3,4</sup>, Federico Martínón-Torres<sup>3,4</sup> & GENDRES network\*

Respiratory syncytial virus (RSV) is an important cause of serious lower respiratory tract disease in infants. Several studies have shown evidence pointing to the genome of the host as an important factor determining susceptibility to respiratory disease caused by RSV. We sequenced the complete exomes of 54 patients infected by RSV that needed hospitalization due to development of severe bronchiolitis. The Iberian sample (IBS) from The 1000 Genomes Project (1000G) was used as control group; all the association results were pseudo-replicated using other 1000G-European controls and Spanish controls. The study points to SNP rs199665292 in the olfactory receptor (OR) gene *OR13C5* as the best candidate variant ( $P$ -value =  $1.16 \times 10^{-12}$ ; OR = 5.56). Genetic variants at HLA genes (*HLA-DQA1*, *HLA-DPB1*), and in the mucin 4 gene (*MUC4*) also emerge as susceptibility candidates. By collapsing rare variants in genes and weighing by pathogenicity, we obtained confirmatory signals of association in the OR gene *OR8U1/OR8U8*, the taste receptor *TAS2R19*, and another mucin gene (*MUC6*). Overall, we identified new predisposition variants and genes related to RSV infection. Of special interest is the association of RSV to olfactory and taste receptors; this finding is in line with recent evidence pointing to their role in viral infectious diseases.

Respiratory syncytial virus (RSV) was first isolated from chimpanzees in 1956<sup>1</sup>. Soon thereafter it was observed in young children and recognized as an important cause of serious lower respiratory tract disease in infants and young children, such as pneumonia and bronchiolitis. Today, RSV infection is the single most important cause of lower respiratory tract infections and hospital admissions during infancy and childhood. While most children are infected by the age of two years, only a minor proportion of them need hospitalization due to development of severe disease<sup>2</sup>. Furthermore, RSV also causes repeated infections and significant disease throughout life. Treatment is generally based on supportive care and includes oxygen therapy; after more than 60 years since its discovery, no effective vaccine is available yet<sup>3</sup>. There are several risk factors that predispose children to severe bronchiolitis and pneumonia, including premature birth, congenital heart disease, T-cell immunodeficiency, and chronic lung disease. However, most of the children suffering severe RSV disease are apparently healthy without known identifiable risk factors<sup>4</sup>. At the same time, age and sex are important independent risk factors for serious RSV disease<sup>5</sup>.

There is a growing body of evidence indicating association of RSV disease susceptibility with ethnicity, and with a family history of asthma<sup>6</sup>. Moreover, a few recent studies indicate that susceptibility to RSV infection in mice can be genetically determined<sup>7,8</sup>. The study by Stark *et al.*<sup>7</sup> suggested that susceptibility to RSV is likely to be a multigenetic trait (as with other infectious diseases<sup>9,10</sup>) while some authors have suggested that host genetic

<sup>1</sup>Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses, Instituto de Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela(USC), Galicia, Spain. <sup>2</sup>GenPoB Research Group, Instituto de Investigaciones Sanitarias (IDIS), Hospital Clínico Universitario de Santiago de Compostela (SERGAS), Galicia, Spain. <sup>3</sup>Translational Pediatrics and Infectious Diseases, Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain. <sup>4</sup>GENVIP Research Group (www.genvip.org), Instituto de Investigación Sanitaria de Santiago, Galicia, Spain. \*A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to A.S. (email: [antonio.salas@usc.es](mailto:antonio.salas@usc.es))

variation could play a role in the severity of RSV disease. For instance, Thomsen *et al.*<sup>11</sup> reported significantly higher concordance rates in the susceptibility to severe RSV infection in identical twin pairs, when compared to fraternal twin pairs.

Most genetic studies hitherto carried out in the context of RSV disease have focused on a candidate-gene based strategy. There are only a few case-control association and family-based studies performed on human cohorts reporting several candidate genes that could predispose to RSV infection<sup>12</sup>. Several genes seem to be related to host genetics on innate defense given that primary RSV disease occurs before the development of specific acquired immunity and in the presence of maternal antibody. The genes related to the hydrophilic surfactant proteins A1, A2 and D, namely, *SFPA1*, *SFPA2*, and *SFPD*, are considered to play a major role in the pulmonary defense mechanism and therefore in RSV infection<sup>13</sup>. The Toll-like receptor 4 gene (*TLR4*) regulates innate and adaptive immune response by recognizing pathogen-associate molecular patterns and could also play a crucial role in RSV infection<sup>14</sup>. Inoue *et al.*<sup>15</sup> reported the association of *CD14* variants with the development of RSV bronchiolitis in the Japanese population. Several authors suggested the involvement of cytokine genes in the association of RSV disease. For instance, Th1 and Th2 cytokine genes such as *IL4*, *IL13*, and *IL5*, have been found to be related to RSV disease severity in a Korean cohort of hospitalized children<sup>16</sup>. The study by Janssen *et al.*<sup>17</sup> signaled several innate immunity genes (*VDR*, *JUN*, *IFNA5* and *NOS2A*) as candidate genes in RSV infection<sup>17</sup>.

Genome-wide association studies (GWAs) have been very popular to disentangle the genetic component of complex multifactorial diseases; however, little effort has been devoted to genome-wide strategies for the study of the genetic susceptibility to RSV infection so far. Very recently, Pasanen *et al.*<sup>18</sup> carried out the first genome-wide analysis of bronchiolitis patients; the study indicated suggestive association signals for several SNPs, including rs269094 located in gene *KCND3*, which has previously been linked to asthma<sup>19</sup>. GWAs studies usually target common and noncoding genetic variants (MAF  $\geq$  5%) whose effects are small (relative risk  $\sim$  1.5), and therefore cannot cover rarer genetic factors that could be associated to RSV disease. It is therefore possible that SNPs usually genotyped in GWAs capture only a small proportion of disease heritability, while an important part of this heritability could reside in rare variants. In contrast, next generation sequencing (NGS) technologies applied to the study of exomes (whole exome sequencing; WES) allow exploring the majority of non-synonymous coding genetic variants in cohorts of patients. Sequencing the complete exome of patients enables the identification of all genetic variation existing in a gene (most of human variation consists of rare variants), and therefore there is no need to rely on linkage disequilibrium (LD) to tag uncovered causal SNPs. In addition, studies based on the selection of extreme phenotypes for the identification of relevant clinical variants have several advantages when compared to those studies that do not take into account clinical sub-phenotypes<sup>20</sup>. Thus, a modest sample size can achieve the necessary statistical power to identify new candidate genes, because the frequency of alleles that contribute to disease would be enriched in one or both phenotypic extremes and their genetic effect could be higher<sup>21</sup>.

At the same time, owing to the characteristics of the patients suffering from pediatric infectious diseases, reaching sufficiently large sample sizes to carry out powerful studies is typically a challenging task. Therefore, sequencing exomes of selected extreme phenotypes emerged as the most promising strategy for the analysis of genetic risk factors underlying pediatric infectious diseases, allowing focusing on protein-altering variants. This variation is enriched for causal effects<sup>22</sup>.

The present study aims to explore genome-wide variation of severe phenotypes of RSV disease patients compared to healthy controls using a WES strategy of selected phenotypes. WES allows exploring common and rare variation involved in host susceptibility to RSV infection.

## Material and Methods

**Study design and inclusion criteria.** Patients were selected from a cohort prospectively recruited under an observational study run in Spain through a national hospital-based research network for pediatric respiratory research: GENDRES (Genetics, Vitamin D and Respiratory Infections Research Network – [www.gendres.org](http://www.gendres.org)). This network includes 13 Spanish tertiary hospitals. Eligible study participants in GENDRES were previously healthy children under 14 years of age admitted to a participating hospital with an acute respiratory infection diagnosis. From Dec 2009 to Jun 2014, a total of 551 subjects were recruited by this network. For the present study, hospitalized children from 0 to 3 years of age with confirmed diagnosis of RSV bronchiolitis were selected, provided that (i) written informed consent was available from a parent or legal guardian, (ii) at least a DNA sample was collected, and (iii) the minimum mandatory demographic and clinical data set was available.

RSV status was determined in all patients by routine diagnostic investigations as part of clinical care in the referring hospital. Only patients with immunofluorescence test-confirmed RSV infection and no features of co-existing bacterial infection at admission according to clinical data, inflammatory markers, radiological findings and/or appropriate cultures were included in this study. Patients included were diagnosed as acute bronchiolitis due to RSV according to the ICD-10 definition.

A blood sample was collected from each patient for DNA extraction and posterior sequencing analysis. DNA of these patients was extracted using the Wizard<sup>®</sup> Genomic DNA Purification Kit (Promega).

For the present study, 59 children with confirmed diagnosis of RSV were finally selected from the GENDRES cohort, although only 54 were used after applying different methodological filters. The patients' main characteristics are summarized in Table 1.

The study was approved by the Ethical Committee of Clinical Investigation of Galicia (CEIC ref. 2010/015). The study was conducted according to the principles of the Declaration of Helsinki and in accordance with all applicable Spanish normative, namely, Law for Biomedical Research (Law 14/2007-3 of July), Law 41/2002 of Autonomy of the Patient, Decree SAS/3470/2009 for Observational Studies and Law 15/1999 of Data Protection.

Control groups for genomic comparisons were selected from The 1000 Genomes Project (<http://www.1000genomes.org>; hereafter 1000G); see details below. The 1000G European population sets were used for association

Variables	RSV (n = 54)
Demographic characteristics	
Sex (male) <sup>a</sup>	61.1% (33/54)
Age (years) <sup>b</sup>	0.2 (0.1–0.7)
Past family history	
Asthma <sup>a</sup>	29.6% (16/54)
Respiratory problems <sup>a</sup>	33.3% (18/54)
Medical history	
Pneumococcal vaccination status <sup>a</sup>	
PVL 7	5.7% (3/53)
PVL 10	1.9% (1/53)
PVL 13	41.5% (22/53)
Clinical data	
Hospital length of stay (days) <sup>b</sup>	7 (5–10)
PICU admission <sup>a</sup>	50% (19/38)
PICU (days) <sup>b</sup>	6 (4–8.25)
Respiratory distress <sup>a</sup>	
Severe	27.8% (15/54)
Moderate	57.4% (31/54)
Mild	14.8% (8/54)
Respiratory support <sup>a</sup>	
Mechanical	3.7% (2/54)
Non-invasive	29.6% (16/54)
ReSVinet score <sup>b</sup>	11 (8–13.5)
Wood Downes score <sup>b</sup>	6 (5–7)
Oxygen <sup>a</sup>	79.6% (43/54)
Bacterial super-infection <sup>a</sup>	13.2% (7/53)

**Table 1.** Summary of the demographic and clinical characteristics of the study cohort. <sup>a</sup>Percentage and number of patients. <sup>b</sup>Median (interquartile range).

tests: the Iberian control group (IBS) was used as the main control group; while CEU, GBR and TSI were used as additional control groups (see more details below).

**Exome sequencing.** Samples were prepared according to Agilent's SureSelect Protocol Version 1.2. Enrichment was carried out according to Agilent SureSelect protocols. Concentration of each library was determined using Agilent's QPCR NGS Library Quantification Kit (G4880A). Samples were pooled prior to sequencing with each sample at a final concentration of 10nM.

Sequencing was performed on the Illumina HiSeq2000 platform using TruSeq v3 chemistry. Mapping and alignment were carried out as follows. First, read files (Fastq) were generated from the sequencing platform using the manufacturer's proprietary software. Reads were mapped to their location in the reference human genome (hg19/b37) using the Burrows-Wheeler Aligner (BWA) package, version 0.6.2. Local realignment of the mapped reads around potential insertion/deletion (indel) sites was carried out with the Genome Analysis Tool Kit (GATK) version 1.6. This algorithm ensures that the alignment has the minimum number of mismatching bases across the reads; the main effect of this step is to reduce false-positive single nucleotide variant calls around indels, and hence determine indel length more accurately. Duplicate reads were marked using Picard version 1.104. This removes reads likely to be the result of PCR bias. Such PCR artifacts can introduce false-positive SNP calls. Reads were not removed from the alignment but they were not considered further in the analysis.

Additional BAM file manipulations were performed with Samtools 0.1.18. Base quality (Phred scale) scores were recalibrated using GATK's covariance recalibration. This improves the accuracy of the base quality metrics, which in turn enhances the quality of variant calls.

Exome sequencing was carried out in Oxford Gene Technology (OGT; <http://www.ogt.co.uk>) using Agilent's SureSelectXT Human All Exon V5 exome design without UTRs. The raw data were processed entirely at the laboratory of Santiago de Compostela (Spain). The average on-target coverage found on all samples was 64.23×.

**Quality control.** One additional sample was processed following exactly the same steps as the rest of the samples, including WES. The two exome sequences were compared and the coincidence of the sequencing results was 99.999%.

In addition, we carried out an independent genotyping of the SNP variants that turned to be statistically associated in our case-control associated study between RSV and IBS controls, including both SNPs obtained from individual association tests and those SNPs in genes statistically significant under a burden association test. We therefore designed three SNP multiplex reactions for genotyping using the MassARRAY SNP genotyping system (Sequenom, San Diego, CA, USA) and using the iPLEX GOLD (Agena Bioscience) assay to increase pllexing

efficiency and flexibility. Since this genotyping effort aimed at independently evaluating the quality of exome sequencing results, only those SNPs meeting the highest standards required by the Assay Design 3.1 software were incorporated to the final plexes and successfully genotyped ( $n = 31$ ).

The concordance between MALDI-TOFF assays genotyping and parallel sequencing was 98.0% for all the genotypes (and 98.1% for those carriers of the minor alleles at the individual associated variants found in the present study). The very low estimated sequencing error rate in our exomes (0.001%) coupled with the sequencing error rate of the Mass Array platform (2.3%;<sup>23</sup>), add solid support to the variation detected in the present study.

**Variant detection, annotation and *in silico* assessment of their pathogenicity.** Variants were detected using GATK v3.4 best practice for multi-sample calling, which increases the detection and genotyping power of this already very sensitive variant caller. Genomic VCF files were obtained for each sample using the HaplotypeCaller algorithm, and joint genotyping was performed using the GenotypeGVCFs algorithm. Variant quality scores were then recalculated using the VQSR algorithm to allow the filtering of outlying variants by improving the sensitivity vs. specificity balance. Variants were annotated using ANNOVAR<sup>24</sup>, and using gene and gene function data from Ensembl (<http://www.ensembl.org/index.html>). This procedure makes it possible to determine which genes and transcripts are affected by the variations, and whether these variants are likely to cause significant functional problems. Known variants from dbSNP (Release 135) were annotated within the dataset so that novel variants with serious predicted consequences may be rapidly identified.

Different scoring systems were applied to the annotated variants in order to measure the pathogenicity/deleteriousness of single nucleotide variants as well as insertion and deletion variants in the exomes (PolyPhen<sup>25</sup>, SIFT<sup>26</sup>, GERP<sup>27</sup>, etc); Table S1. Besides those methods that exploit a single information type (such as conservation) and/or are restricted in scope (e.g. to missense changes) we also obtained the Combined Annotation Dependent Depletion (CADD), a score that integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated mutations<sup>28</sup>. CADD can quantitatively prioritize functional, deleterious, and disease-causal variants across a wide range of functional categories, effect sizes and genetic architectures and can be used to prioritize causal variation in both research and clinical settings. The linear kernel support vector machine-based algorithm used in CADD analysis has been improved by using a deep neural network which also considers nonlinear effects; this algorithm is known as Deleterious Annotation of Genetic Variants (DANN)<sup>29</sup>, and it also provides a score.

**Population-based and statistical analysis.** Identity-by-state values were computed from SNP data using PLINK<sup>30</sup>. In order to allocate the genome variability observed in our cohort of patients, we carried out Multidimensional Scaling (MDS) analysis on a matrix of pairwise individual identity-by-state values, and in the context of main continental variation. MDS was performed using the function *cmdscale* (library stats) from R (<http://www.r-project.org>). Next, we additionally investigated admixture patterns in our patients using different source populations representing main continental regions (sub-Saharan Africa, East Asia, and Europe). Population data for comparison was retrieved from 1000G with the assistance of previous bioinformatics developments<sup>31,32</sup>. First, all variant files were downloaded from the 1000G public ftp (<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/>); next, we followed the procedures described previously for the processing of these data<sup>33</sup>.

Maximum likelihood estimation of individual ancestries from multi-locus SNP data was carried out using ADMIXTURE<sup>34</sup>.

We also explored the genome of all patients and controls for potential familial relationship as done previously<sup>35,36</sup>.

Single-point association analyses were carried out using the Fisher exact test. The 1000G European datasets for Iberia (IBS), Great Britain (GBR), Tuscany (TSI), and CEU (USA-European ancestry, CEU) were used as controls. The possible confounding effect of population stratification was estimated by calculating the genomic inflation factor (lambda or  $\lambda$ ) from the data<sup>37</sup>;  $\lambda$  is the ratio of the median of the empirically observed distribution of the chi-squared values to the expected median. As such, it provides a way to quantify the extent of the genome inflation and the excess false-positive rate. This uniform correction can over-adjust or under-adjust certain SNPs depending on the ancestral information of individual variants; however, as shown below, we first demonstrate that the samples analyzed, cases and controls, are genetically homogeneous and share a common European ancestral background.

In addition, SNP association analyses were carried out, collapsing variants by genes and taking into account several parameters that could indicate potential pathogenicity (weighted burden tests). The procedure was carried out as follows. For an individual  $j$  and a Region of Interest (ROI) with  $K$  biallelic positions (being  $K > 5$ ), we define  $x'_{ij}$  as the number of variants at position  $i$  for subject  $j$ . This parameter can take values 0, 1 or 2. Our approach, modified from<sup>38</sup>, follows the following steps:

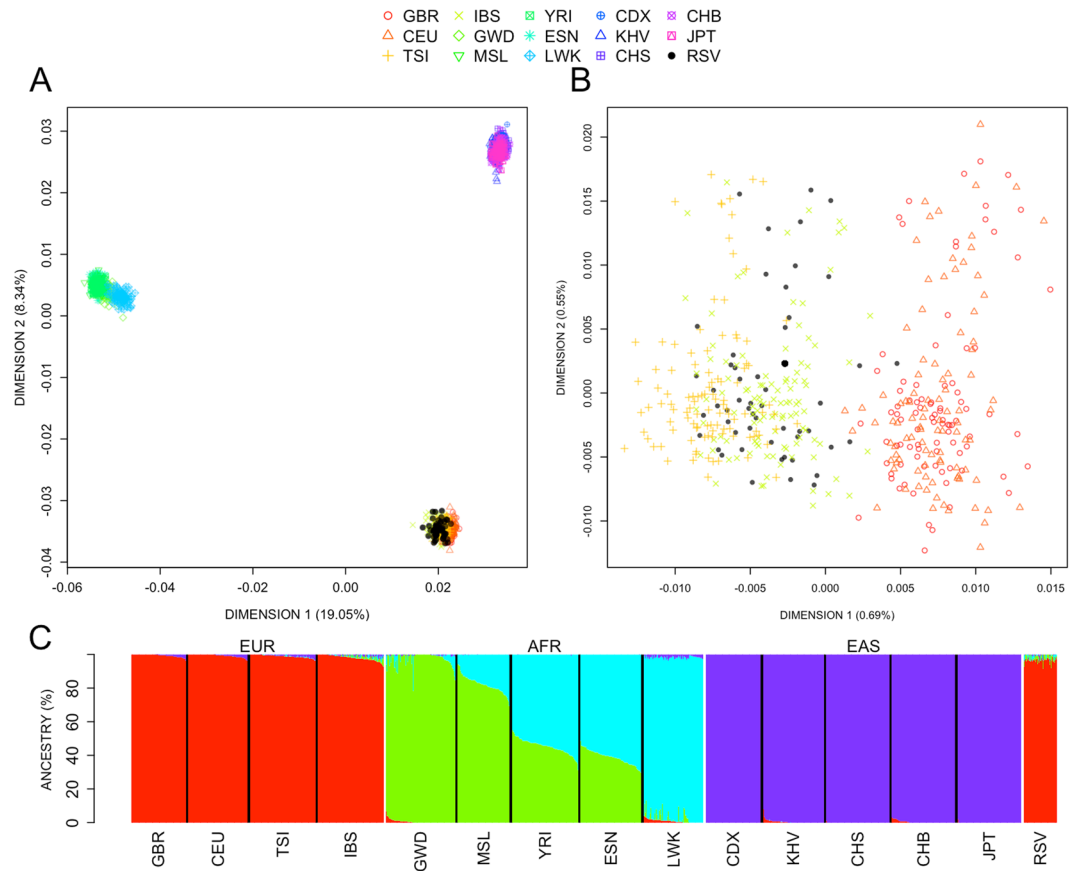
For each subject, we compute  $x'_j = \frac{1}{K} \sum_{i=1}^K \text{dann}_i x'_{ij}$ .

The sum of ranks of  $x'_j$  is computed for affected subjects  $r = \frac{1}{K} \sum_{j \in \text{affected}} x'_j$ . To deal with the possible ties, we assign the average rank position for the tied individuals.

The affection status is permuted  $B$  times, and  $r^*_b$  values are computed to simulate the distribution under the null hypothesis.

The  $P$ -value is calculated considering the normal distribution with the permuted mean ( $\widehat{\mu}_B = \frac{1}{B} \sum_{i=1}^B r_i^*$ ) and permuted standard deviation ( $\widehat{\sigma}_B = \sqrt{\frac{1}{B} \sum_{i=1}^B (r_i^* - \widehat{\mu}_B)^2}$ ).

Collapsing variant analysis was carried out on all variants observed in exomes, and separately on common, rare, and non-synonymous variants.



**Figure 1.** (A) MDS analysis carried out on a matrix of pair-wise individual identity-by-state values between RSV patients vs. different 1000G population reference sets (see legend inset). (B) Enlarged detail of the European cluster observed in (A). (C) Admixture analysis of the samples analyzed in (A).

A Bonferroni correction was used in order to account for multiple hypotheses. Most of the association test analyses were carried out using in-house R and Perl (<http://www.perl.org>) scripts.

Haploview was used to display LD patterns between SNPs within genes<sup>39</sup>. Finally, Gene Ontology (<http://www.geneontology.org>) analysis of statistically significant genes was carried out on PANTHER<sup>40</sup>.

## Results

**Clinical features of patients.** A total of 54 patients of European ancestry were sequenced in the present study. The cohort characteristics are summarized in Table 1. From this group, 61.1% of the subjects were male with a median age of 0.2 years (0.1–0.7). Family history showed respiratory problems in 18 (33.3%) of the subjects, 16 (29.6%) of them being asthma. Prematurity was part of the medical history of two of the subjects. The main reason for hospital admission was respiratory distress ( $n = 53$ ). We classified subjects clinically with regards to the level of respiratory affection: 8 (14.8%) were mild, 31 (57.4%) were moderate, and 15 (27.8%) were severe. Wood-Downes and ReSVinet scores were also applied to assess patient severity<sup>41</sup>. The Wood-Downes score divided the patients in 11 (20.4%) clinically mild, 30 (55.6%) moderate and 13 (24.1%) severe subjects, with an average score of 6 (5–7). The ReSVinet score distinguished between severe ( $\geq 13$ ) and moderate-mild, with an average score of 11 (8–13.5). During admission, supplementation with oxygen was needed in 43 (79.6%) cases. Corticoids (systemic or inhaled) were used in eight (14.8%) subjects. 19 (35.2%) of the subjects were admitted to PICU, with a median stay in days of 6 (4–8.3). Ventilatory support was applied in 18 (33.3%) patients, 16 (29.6%) were non-invasive and two (3.7%) invasive ventilation. Secondary to hemodynamic instability inotropic drugs were deemed necessary in 21 (38.9%) subjects. The median length of hospital stay was 7 days (5–10). No major sequelae or mortality was present at discharge.

**Ancestry analysis of RSV patients.** Analysis of the population-genetic characteristics of patients is important in order to detect possible genomic outliers. We first carried out a MDS to a continental scale using 1000G populations as references. As shown in Fig. 1A and B, our cohort of patients shows patterns of variation that fit perfectly in the European cluster in both the first dimension (accounting for ~19% of the variation) and the second dimension (~8%) of the MDS, clustering together with other samples of European ancestry. A good fit exists between RSV cases and IBS controls when other dimensions of the MDS are examined (Figure S1). In addition, we carried out an admixture analysis in order to corroborate the patterns observed in the MDS. As expected, ancestry of our cohort of patients is virtually 100% European (Fig. 1C).

Variation	N
Downstream	6
Exonic	81461
Exonic/splicing	54
Intergenic	38
Intronic	365
ncRNA_exonic	5845
ncRNA_exonic; splicing	5
ncRNA_intronic	497
ncRNA_splicing	3
Splicing	52
Upstream	18
Upstream; downstream	3
UTR3	10958
UTR5	7390
UTR5/UTR3	16
Non-synonymous SNP	41596
Stopgain	416
Stoploss	43
Synonymous SNP	38401
Unknown	1059

**Table 2.** Description of variants found in our patients.

We finally investigated potential relationships in our cohort of patients in order to avoid potential artifacts and interferences with the association analysis. The data show that there are no close relationships among them.

**Annotated variant and SNP association analyses.** A total of 108,800 variants (passing the filters indicated above) were annotated from the exome data, of which 106,711 remained after filtering by Hardy-Weinberg equilibrium ( $P$ -value  $> 0.005$  in all cases, control groups individually, and control groups merged in a single group). 81,461 of the filtered variants were exonic, of which 41,596 are non-synonymous (Table 2). All these variants have an average genotyping rate  $> 99.99\%$ .

Although MDS and admixture analysis do not show evidence of population stratification between cases and controls, we computed the inflation factor on  $P$ -values existing between cases and the different European controls used for the association analysis. Given that  $\lambda$  was below 1, it was not necessary to control association tests for inflation.

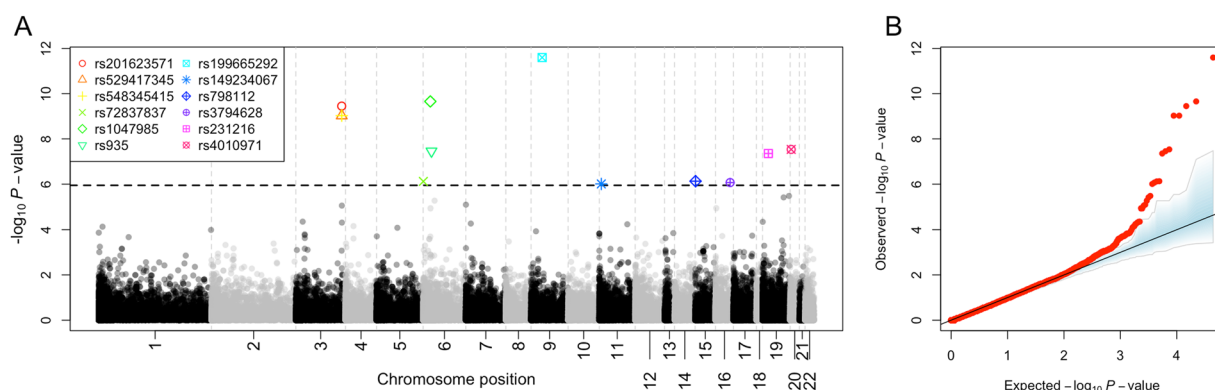
Single association tests were carried out for common variants with minor allele frequencies (MAF)  $\geq 0.05$  (Table 3; Table S1). The QQ-plot in Fig. 2B shows a few SNPs with very low  $P$ -values that differ from the expected distribution of  $P$ -values under a uniform distribution. To further test for the robustness of the statistical significance of these associated SNPs, a simulation approach was conceived that computes the  $P$ -values for each SNP marker after permuting the case-control status of all samples. Under 1,000 permutations, we observe a very good fit between the theoretical expected  $P$ -values and the permuted ones (Fig. 2B). A total of 12 SNPs have  $P$ -values that are below the conservative Bonferroni adjustment (Fig. 2A) in IBS. The signal of association was explored in the other three control groups (CEU, GBR, and TSI). All the markers surpassed the multiple test threshold (Bonferroni adjustment considering 12 variants) in all control groups. Moreover, four of them (rs199665292, rs201623571, rs529417345, and rs548345415; Table 3) yielded statistical  $P$ -values below the genome Bonferroni threshold (Fig. 3; see also Figure S2); only two of them (rs798112, and rs3794628) were below the genome-wide Bonferroni adjustment when merging all the control groups (including IBS) in a single one (hereafter “ALL”); Fig. 3A. To test for the suitability of CEU, GBR, and TSI as control groups of our RSV cohort, we carried out an association test between SNP variations in IBS vs. these other three control groups. Only 18, 10, and 7 SNPs out of 108,800 variants were statistically significant under a Bonferroni adjustment, and none of these few SNPs coincided with the 12 SNPs observed to be statistically significant between RSV and the control groups (Figure S3). This shows that the four cohorts are not different from the point of view of an association test.

We furthermore carried out an association test between RSV patients and the exome data of the Spanish control group (EC) in Dopazo *et al.*<sup>42</sup>. Only five out of the 12 candidate SNPs are present in this dataset. The five SNPs received  $P$ -values below the Bonferroni threshold (Table 3).

The lead SNP was rs199665292 ( $P$ -value<sub>IBS</sub> =  $2.53 \times 10^{-12}$ , OR<sub>IBS</sub> = 9.09;  $P$ -value<sub>ALL</sub> =  $1.16 \times 10^{-12}$ , OR<sub>ALL</sub> = 5.56), an exonic synonymous SNP located in gene *OR13C5* (Table 2). This SNP appears in ExAC browser as tetra-allelic, although two alleles appear only twice in more than  $> 103,400$  alleles; however, we only detected the two common main alleles in our cohort of patients and in all reference populations (including 1000G). The second most significant SNP, rs1047985 ( $P$ -value<sub>IBS</sub> =  $2.21 \times 10^{-10}$ , OR<sub>IBS</sub> = 0.05;  $P$ -value<sub>ALL</sub> =  $1.43 \times 10^{-09}$ , OR<sub>ALL</sub> = 0.06), falls within the *HLA-DQA1* gene and it is located in an UTR5 region. The next most significant variants are three SNPs, all within gene *MUC4*: rs201623571 ( $P$ -value<sub>IBS</sub> =  $3.5 \times 10^{-10}$ , OR<sub>IBS</sub> = 0.08;  $P$ -value<sub>ALL</sub> =  $2.48 \times 10^{-10}$ , OR<sub>ALL</sub> = 0.10), rs529417345 and rs548345415 ( $P$ -value<sub>IBS</sub> =  $9.40 \times 10^{-10}$ , OR<sub>IBS</sub> = 0.03;

SNP	Chr	Position	Gene	Gene region	Exonic function	OR <sub>IBS</sub>	P-value <sub>IBS</sub>	OR <sub>ALL</sub>	P-value <sub>ALL</sub>	P-value <sub>EC</sub>
rs199665292	9	107361452	<i>OR13C5</i>	exonic	Synonymous	9.09	$2.53 \times 10^{-12}$	5.56	$1.16 \times 10^{-12}$	$3.53 \times 10^{-15}$
rs1047985	6	32605216	<i>HLA-DQA1</i>	UTR5	—	0.05	$2.21 \times 10^{-10}$	0.06	$1.43 \times 10^{-09}$	$2.50 \times 10^{-06}$
rs201623571	3	195508393	<i>MUC4</i>	exonic	Non-synonymous	0.08	$3.55 \times 10^{-10}$	0.10	$2.48 \times 10^{-10}$	—
rs529417345	3	195508416	<i>MUC4</i>	exonic	Synonymous	0.03	$9.40 \times 10^{-10}$	0.03	$5.31 \times 10^{-10}$	—
rs548345415	3	195508418	<i>MUC4</i>	exonic	Non-synonymous	0.03	$9.40 \times 10^{-10}$	0.03	$5.31 \times 10^{-10}$	—
rs4010971	20	29637822	<i>MLLT10P1</i>	ncRNA_exonic	—	0.19	$2.92 \times 10^{-08}$	0.20	$1.16 \times 10^{-09}$	$3.77 \times 10^{-06}$
rs935	6	33054579	<i>HLA-DPB1</i>	UTR3	—	0.14	$3.51 \times 10^{-08}$	0.15	$5.09 \times 10^{-09}$	—
rs231216	19	36259494	<i>PROSER3</i>	UTR3	—	5.35	$4.43 \times 10^{-08}$	3.72	$5.70 \times 10^{-08}$	$9.27 \times 10^{-10}$
rs798112	15	31110279	<i>HERC2P10</i>	ncRNA_exonic	—	0.10	$7.43 \times 10^{-07}$	0.13	$8.15 \times 10^{-06}$	$4.30 \times 10^{-04}$
rs72837837	6	15520916	<i>JARID2</i>	UTR3	—	—	$7.59 \times 10^{-07}$	—	$9.09 \times 10^{-08}$	—
rs3794628	16	88729788	<i>SNAI3-AS1</i>	ncRNA_exonic	—	0.14	$8.52 \times 10^{-07}$	0.13	$6.29 \times 10^{-09}$	—
rs149234067	11	33722084	<i>C11orf91</i>	exonic	Non-synonymous	0.23	$9.81 \times 10^{-07}$	0.30	$1.67 \times 10^{-05}$	—

**Table 3.** Association test of common variants. The SNPs showing the lowest  $P$ -value against the IBS control group were further tested using CEU, GRB, and TSI as control groups and merging the three sample sets into a single one (“ALL” = IBS + CEU + GBR + TSI). EC refers to the exome sequencing data in the Spanish control group ( $n = 267$ ) of Dopazo *et al.*<sup>42</sup>. The SNP rs199665292 has merged into rs75081605 (<https://www.ncbi.nlm.nih.gov>); a hyphen in column “ $P$ -value<sub>EC</sub>” indicates that there is not genetic information available in the Spanish control groups concerning this variant. Chr = chromosome;  $P$ -values: Fisher’s Exact Test; OR = odds ratio.



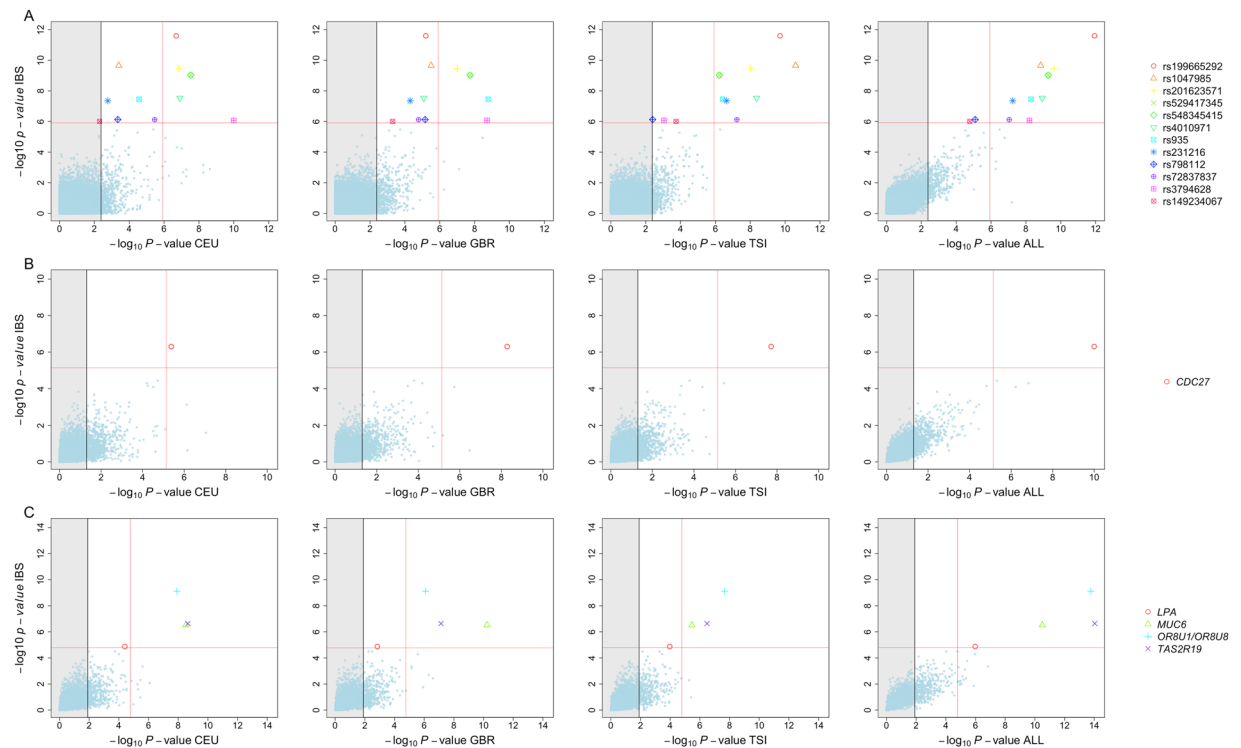
**Figure 2.** (A) Manhattan plot of common variants observed in RSV patients. (B) QQ-plot of  $P$ -values for common variation observed in IBS controls.  $P$ -values obtained under a permutation approach (1,000 permutations) are shown in blue.

$P$ -value<sub>ALL</sub> =  $5.31 \times 10^{-10}$ ; OR<sub>ALL</sub> = 0.03); interestingly, rs201623571 and rs548345415 are non-synonymous (Table 3). Another SNP out of the 12 statistically significant ones, rs149234067 in gene *C11orf91*, is also non-synonymous.

**Weighted burden test by genes.** By collapsing all variants observed in genes, a clear signal of association was observed for *CDC27*, with statistical significance below the genome-wide Bonferroni threshold for all European control groups individually, as well as for the merged controls ( $P$ -value<sub>IBS</sub> =  $4.97 \times 10^{-7}$ ,  $P$ -value<sub>ALL</sub> =  $1.01 \times 10^{-10}$ ); Table 4, and Fig. 3B.

Many deleterious and disease-related variants exist at low frequency in the human genome<sup>43</sup>. We partitioned the variants observed in our patients into separate classes according to their MAF: rare ( $<0.05$ ) and common variants ( $\geq 0.05$ ). By collapsing common variants in genes, we did not identify any candidate gene. However, burden tests on rare variants yielded significant signatures for five genes in IBS (Table 4; Fig. 3C). The signal of association was investigated in three other control groups (CEU, GBR, and TSI). All the genes surpassed Bonferroni correction (considering four independent genes). Moreover, all these genes yielded statistical  $P$ -values below the genome Bonferroni threshold in the European control category ALL, and most of them (with the exception of *LPA* gene) in the European control group individually (Fig. 3). The top association signal was found for the olfactory receptor *OR8U1/OR8U8* ( $P$ -value<sub>IBS</sub> =  $7.60 \times 10^{-10}$ ,  $P$ -value<sub>ALL</sub> =  $1.69 \times 10^{-14}$ ). The other genes were, in order of significance: *TAS2R19* ( $P$ -value<sub>IBS</sub> =  $2.30 \times 10^{-7}$ ,  $P$ -value<sub>ALL</sub> =  $9.21 \times 10^{-15}$ ), *MUC6* ( $P$ -value<sub>IBS</sub> =  $3.00 \times 10^{-7}$ ,  $P$ -value<sub>ALL</sub> =  $3.12 \times 10^{-11}$ ), and *LPA* ( $P$ -value<sub>IBS</sub> =  $1.34 \times 10^{-5}$ ,  $P$ -value<sub>ALL</sub> =  $1.07 \times 10^{-6}$ ).

In addition, we carried out an association test between RSV patients and the exome data of the Spanish control group in Dopazo *et al.*<sup>42</sup>. The analysis could only be carried out on three out of the five genes found to be associated in the discovery cohort (namely those genes with a minimum number of variants per gene according to the procedures described in Material and Methods). The three genes received  $P$ -values below the Bonferroni



**Figure 3.** (A)  $P$ -values of SNP association between RSV patients and controls in different control groups and in the merged control group “ALL” (IBS + CEU + GBR + TSI). (B)  $P$ -values of gene burden association test between RSV patients and controls in different control groups and in the merged control group “ALL” considering all variants in genes. (C)  $P$ -values of gene burden association test between RSV patients and controls in different control groups and in the merged control group “ALL” considering only rare variants in genes. The vertical grey shadow in the figures indicates the threshold for the adjusted Bonferroni  $P$ -value according to the number of independent tests considering CEU, GBR, TSI, and ALL as additional control groups of the best SNPs ( $n = 12$ ) and gene candidates (one gene when collapsing common variants and four genes when collapsing rare variants). The red vertical line indicates the genome Bonferroni threshold considering genes and all the common SNPs that are shared between cohorts.

Gene	Chr.	n° SNP	$P$ -value <sub>IBS</sub>	$P$ -value <sub>CEU</sub>	$P$ -value <sub>GBR</sub>	$P$ -value <sub>TSI</sub>	$P$ -value <sub>ALL</sub>	$P$ -value <sub>EC</sub>
<i>CDC27</i>	17	6	$4.97 \times 10^{-07}$	$4.20 \times 10^{-06}$	$5.29 \times 10^{-09}$	$1.92 \times 10^{-08}$	$1.01 \times 10^{-10}$	$4.3 \times 10^{-11}$
<i>OR8U1/OR8U8</i>	3	9	$7.60 \times 10^{-10}$	$1.22 \times 10^{-08}$	$8.10 \times 10^{-07}$	$2.14 \times 10^{-08}$	$1.69 \times 10^{-14}$	—
<i>TAS2R19</i>	4	6	$2.30 \times 10^{-07}$	$2.25 \times 10^{-09}$	$7.41 \times 10^{-08}$	$3.30 \times 10^{-07}$	$9.21 \times 10^{-15}$	—
<i>MUC6</i>	3	18	$3.00 \times 10^{-07}$	$3.10 \times 10^{-09}$	$5.85 \times 10^{-11}$	$3.46 \times 10^{-06}$	$3.12 \times 10^{-11}$	$7.84 \times 10^{-02}$
<i>LPA</i>	19	12	$1.34 \times 10^{-05}$	$3.96 \times 10^{-05}$	$1.44 \times 10^{-03}$	$1.07 \times 10^{-04}$	$1.07 \times 10^{-06}$	$1.07 \times 10^{-03}$

**Table 4.** Association test of SNPs collapsed by genes and using average CADD per gene as a covariant for the association test. *CDC27* appeared as statistically significant when analyzing common variants ( $MAF \geq 0.05$ ), while the other genes appeared as statically significant when analyzing rare variants ( $MAF < 0.05$ ). The genes showing the lowest  $P$ -value against the IBS control group were further tested using CEU, GRB, and TSI as control groups and merging the three sample sets into a single one (“ALL” = IBS + CEU + GBR + TSI); a hyphen in column “ $P$ -value<sub>EC</sub>” indicates that there is not enough genetic information available in the Spanish control groups.

threshold (Table 3), and the *CDC27* was highly significant ( $P$ -value<sub>EC</sub> =  $4.3 \times 10^{-11}$ ), surpassing the genome-wide Bonferroni threshold.

For all statistically significant genes observed in these analyses, the accumulated pathogenicity was increased in cases with respect to controls.

Non-statistically significant results were obtained when computing the weighted burden test only on non-synonymous variants.

**Gene ontology: enrichment analysis of genes associated to RSV infection.** To better understand the molecular function of the best candidate genes observed in the association analysis, we examined their Gene Ontology (GO) classification<sup>44</sup>. All genes signaled by the single base association analysis together with those



captured by collapsing variants, were used to infer whether any biological process, cellular component, or molecular function is over- or under-represented with respect to expectations.

The biological process “antigen processing and presentation of peptide or polysaccharide antigen *via* MHC class II” was the most statistically significant under Bonferroni correction ( $P$ -value =  $5.99 \times 10^{-4}$ ), with a Fold enrichment > 100. When analyzing GO by cellular component, the category “MHC protein complex” was also found to be overrepresented ( $P$ -value =  $1.37 \times 10^{-5}$ ), again with a Fold enrichment > 100. By analyzing by protein class, the “MHC antigen” class was also significant ( $P$ -value =  $7.58 \times 10^{-4}$ ; Fold enrichment: 97.7). Finally, when analyzing the “GO biological process complete”, the class “immune response-activating cell surface receptor signaling pathway” was also found to be enriched ( $P$ -value =  $2.09 \times 10^{-2}$ ; Fold enrichment: 20.9).

## Discussion

The present study adds convincing support to our hypothesis that the genome of the hosts carries variation predisposing to severe infection by RSV. Among the new predisposition variants and genes identified, olfactory and taste receptors associations are of special interest.

By analyzing the exome of RSV infected patients we aimed to capture functional variation that could explain the association between RSV infection and the host. In order to prevent the confounding effect of population stratification, we initially carried out several population-based analyses; these analyses indicated a good match between cases and controls in terms of their genomic ancestry.

A total of 12 SNPs show genome-wide statistical significance. The top SNP falls within *OR13C5* gene, known as OR family 13 gene. This is a risk variant and was consistently significant in all control groups, including the merged European controls. Although this is a synonymous SNP, its main interest derives from the fact that another OR gene (*OR8U1/OR8U8*), which is a member of the OR gene family 8 and subfamily U, was found to be statistically significant when collapsing rare variants and controlling for pathogenicity. Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals<sup>45</sup>. At the same time, the nasal cavity is the natural entrance and replication of RSV host invasion. Several studies have investigated the olfactory nerve, which connects the nasal cavity directly with the central nervous system (CNS), in regard to many virus infections, including influenza A virus, herpesviruses, adenoviruses, etc. Thus, several observations have shown that neurons in the olfactory epithelium can be targeted by influenza A virus in mice<sup>46</sup>, ferrets<sup>46</sup>, and humans<sup>47</sup>, and that highly neurovirulent H5N1 avian influenza strains in ferrets can spread through olfactory nerves to the olfactory bulb after intranasal infection<sup>48</sup>. In humans, influenza virus antigens in neurons and glia in the olfactory bulbs and tracts have recently been found in an immune-compromised child<sup>49</sup>. RSV is not clinically feared as a neurovirulent virus, with neurological symptoms present in less than 2% of patients admitted to hospital<sup>50</sup>. However, Riel *et al.*<sup>51</sup> suggested the interest of the nerve route not only for neurotropic viruses, but also as a shortcut for common respiratory infections, including RSV. Several viral and host factors contribute to the pathogenesis of the disease, and both acquired and genetic host factors will probably affect the pathogenesis and outcome of RSV infections, similarly to severe influenza infection in mouse models<sup>52</sup>. Herpes simplex virus 1 has been shown in mice model to efficiently use the olfactory neuroepithelium to infect, spread and reemerge peripherally in the skin without significant neurological disease associated<sup>53</sup>. Our findings of human genetic variations related to RSV infection suggest either a subclinical involvement of olfactory pathways during RSV infection, or at least a new specific pathway involved in the pathogenesis of the disease that deserves further exploration. According to Riel *et al.*<sup>51</sup>, although RSV and other non-neurotropic viruses do not cause severe CNS complications, they could cause subclinical or mild disease.

Kalbe *et al.*<sup>54</sup> suggested that OR modulate physiological processes in human airways smooth muscle cells, which significantly contributes to the progression of chronic inflammatory airway disease, including asthma and chronic obstructive pulmonary disease (COPD); see also<sup>55</sup>. There is also emerging evidence suggesting that OR can express in non-olfactory tissues; for instance, Li *et al.*<sup>56</sup> indicated that OR genes express in pulmonary macrophages in mice, and these genes may regulate macrophage function by regulating MCP-1 production and cell migration. According to these authors, stimulation of pulmonary macrophages with the host defense cytokine IFN- $\gamma$  up regulates the expression of ORs in this murine model.

Two of the top SNPs, rs1047985 and rs935, fall within *HLA-DQA1* and *HLA-DPB1* genes, respectively, and have a protective effect. There are many studies indicating a role for HLA genes in infectious diseases in general<sup>57,58</sup>, and specifically with bronchiolitis<sup>18</sup>, as well as vaccine response<sup>59</sup>. The study of the transcriptome in RSV bronchiolitis has also shown a role for HLA genes<sup>60–62</sup>.

Three out of the 12 lead SNPs (Table 3) fall within the same gene, *MUC4*; these three variants do not constitute a strong block of linkage disequilibrium (Supplementary Data Figure S5). Another mucin gene, *MUC6*, appears as statistically significant when considering accumulated pathogenicity at rare variants. There are about 20 human mucin genes. These genes encode mucin monomers that are synthesized as rod-shape apomucin cores that are post-translationally modified by exceptionally abundant glycosylation. The dense sugar coating of mucins is very resistant to proteolysis, which may be important in maintaining mucosal barriers. There is an increased mucin production in many cancers, but mucins are also overexpressed in lung diseases such as asthma, bronchitis, COPD or cystic fibrosis. Mucins constitute an important component of the inflammatory and innate immune response although the expression of these molecules by respiratory viral infections is still largely unknown<sup>63</sup>. Polymorphisms in mucin genes have been identified in susceptibility to *H. pylori*-induced pathology in human populations<sup>64</sup>. A recent study of mucin expression in a human epithelial carcinoma cell line (A549) showed a strong induction of *MUC8*, *MUC15*, *MUC20*, *MUC21*, and *MUC22* mediated by RSV infection<sup>65</sup>. Furthermore, RSV fusion protein interacts with epidermal growth factor receptor (EGFR) in a strain-specific manner, suggesting that EGFR is a co-factor for infection and that EGFR plays a role in RSV-induced mucin expression<sup>66</sup>. In

general, there is abundant literature relating RSV infection to upregulation of mucin genes<sup>67–70</sup>. Finally, another gene that could be interesting for future investigations is *TAS2R19*, it is involved in the perception of salty and bitter tastes. A recent review by Lee *et al.*<sup>71</sup> pointed to an important role for taste receptors in airway innate immunity as sentinel detection systems. Also, Malki *et al.*<sup>72</sup> have very recently shown that taste and olfactory receptors show differential mRNA expression in five types of blood leukocytes. Polymorphisms in another taste gene receptor (*TAS2R38*) were related to nitric oxide production and responsiveness to gram-negative bacteria<sup>73</sup>.

As a whole, the candidate genes signaled by the present study seem to play a role in the immune response. In this sense, the major histocompatibility complex is signaled by GO as overrepresented in our cohort of cases.

There are two main limitations of the present study. First, the sample size is limited, although this work is the first of this kind in the field of RSV infection. We aimed at improving its power by selecting extreme phenotypes, assuming a larger effect of causal variants. Second, in lieu of an independent replication cohort we used alternative European control groups. The association findings could be consistently replicated in the different control groups, and the candidate genes observed in our study are also in line with previous evidence in the literature. Future genotyping and sequencing efforts would be necessary to obtain confirmatory evidence for the associations detected in the present study. It is also important to mention that allele frequencies for some of the candidate SNPs observed in the present study show differences between 1000G data and the known repository for exome sequencing data ExAC (<http://exac.broadinstitute.org>). For instance, for European non-Finish data set in ExAC, SNP rs199665292 has similar allele frequencies to our RSV patients; however, the exome data from Spanish controls in Dopazo *et al.*<sup>42</sup> have allele frequencies that are close to IBS (Table S1).

Future studies based on larger samples could also address the genetic variation underlying sub-phenotypes of RSV infection (e.g. severity, length of hospital stay, asthma, etc)<sup>74–76</sup>. In addition, the strategy employed in the present study based on WES could be also of interest at investigating the presumed genetic link existing between RSV and asthma disease<sup>75,77</sup>.

## Conclusion

The present study is pioneering in analyzing exome sequence variation in RSV patients, and in identifying common and rare variants related to host susceptibility to RSV infection. Single-base association analyses suggest new candidate variants related to RSV infection. In addition, burden association tests indicate that rare variants with a high predicted pathogenicity are enriched in a number of genes in carriers when compared to controls. Some of the variants detected fall in regions that have previously been reported as candidate genes in RSV infection (HLA genes), while others, such as olfactory and taste receptor genes are newly reported here, and therefore open new horizons in the search for genetic factors related to RSV infection. To the best of our knowledge this is the first attempt at using WES in the context of the genetic predisposition to infectious disease in childhood by targeting severe phenotypes. This work demonstrates that WES can be particularly useful for the detection of signals of association related to rare variants, in contrast with traditional studies that focused on common disease variation.

## References

- Blount, R. E. Jr, Morris, J. A. & Savage, R. E. Recovery of cytopathogenic agent from chimpanzees with coryza. *Proc. Soc. Exp. Biol. Med.* **92**, 544–549 (1956).
- Collins, P. L., Chanock, R. M. & Murphy, B. R. *Respiratory syncytial virus*. 1443–1486 (Lippincott Williams & Wilkins, 2001).
- Mazur, N. I. *et al.* Lower respiratory tract infection caused by respiratory syncytial virus: current management and new therapeutics. *Lancet Respir. Med.* **3**, 888–900 (2015).
- Hall, C. B. *et al.* The burden of respiratory syncytial virus infection in young children. *N. Engl. J. Med.* **360**, 588–598 (2009).
- Sommer, C., Resch, B. & Simoes, E. A. Risk factors for severe respiratory syncytial virus lower respiratory tract infection. *Open Microbiol. J.* **5**, 144–154 (2011).
- Hull, J. *Genetic susceptibility to RSV disease*. In: (ed) *Respiratory syncytial virus*. **14**, 115–140 (2007).
- Stark, J. M. *et al.* Genetic susceptibility to respiratory syncytial virus infection in inbred mice. *J. Med. Virol.* **67**, 92–100 (2002).
- Jessen, B., Faller, S., Krempl, C. D. & Ehl, S. Major histocompatibility complex-dependent cytotoxic T lymphocyte repertoire and functional avidity contribute to strain-specific disease susceptibility after murine respiratory syncytial virus infection. *J. Virol.* **85**, 10135–10143 (2011).
- Davila, S. *et al.* Genome-wide association study identifies variants in the *CFH* region associated with host susceptibility to meningococcal disease. *Nat. Genet.* **42**, 772–776 (2010).
- Martinon-Torres, F. *et al.* Natural resistance to Meningococcal Disease related to *CFH* loci: Meta-analysis of genome-wide association studies. *Sci. Rep.* **6**, 35842 (2016).
- Thomsen, S. F. *et al.* Increased concordance of severe respiratory syncytial virus infection in identical twins. *Pediatrics* **121**, 493–496 (2008).
- Choi, E. H., Lee, H. J. & Chanock, S. J. Human genetics and respiratory syncytial virus disease: current findings and future approaches. In: *Current Topics in Microbiology and Immunology. Challenges and opportunities for Respiratory Syncytial Virus Vaccines*. 121–137 (Springer-Verlag Berlin Heidelberg, 2013).
- Ghildyal, R. *et al.* Surfactant protein A binds to the fusion glycoprotein of respiratory syncytial virus and neutralizes virion infectivity. *J. Infect. Dis.* **180**, 2009–2013 (1999).
- Tulic, M. K. *et al.* TLR4 polymorphisms mediate impaired responses to respiratory syncytial virus and lipopolysaccharide. *J. Immunol.* **179**, 132–140 (2007).
- Inoue, Y. *et al.* CD14 – 550 C/T, which is related to the serum level of soluble CD14, is associated with the development of respiratory syncytial virus bronchiolitis in the Japanese population. *J. Infect. Dis.* **195**, 1618–1624 (2007).
- Choi, E. H., Lee, H. J., Yoo, T. & Chanock, S. J. A common haplotype of interleukin-4 gene *IL4* is associated with severe respiratory syncytial virus disease in Korean children. *J. Infect. Dis.* **186**, 1207–1211 (2002).
- Janssen, R. *et al.* Genetic susceptibility to respiratory syncytial virus bronchiolitis is predominantly associated with innate immune genes. *J. Infect. Dis.* **196**, 826–834 (2007).
- Pasanen, A. *et al.* Genome-Wide Association Study of Polymorphisms Predisposing to Bronchiolitis. *Sci. Rep.* **7**, 41653 (2017).
- MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
- Cirulli, E. T. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* **11**, 415–425 (2010).
- Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).

22. Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
23. Syrmis, M. W. *et al.* Comparison of a multiplexed MassARRAY system with real-time allele-specific PCR technology for genotyping of methicillin-resistant *Staphylococcus aureus*. *Clin. Microbiol. Infect.* **17**, 1804–1810 (2011).
24. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
25. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
26. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
27. Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
28. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
29. Quang, D., Chen, Y. & Xie, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761–763 (2015).
30. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
31. Amigo, J., Salas, A., Phillips, C. & Carracedo, Á. SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access. *BMC Bioinformatics* **9**, 428 (2008).
32. Amigo, J., Salas, A. & Phillips, C. ENGINES: exploring single nucleotide variation in entire human genomes. *BMC Bioinformatics* **12**, 105 (2011).
33. Pardo-Seco, J., Gómez-Carballa, A., Amigo, J., Martínón-Torres, F. & Salas, A. A genome-wide study of modern-day Tuscans: revisiting Herodotus's theory on the origin of the Etruscans. *PLoS One* **9**, e105920 (2014).
34. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
35. Gómez-Carballa, A. *et al.* Indian signatures in the westernmost edge of the European romani diaspora: new insight from mitogenomes. *PLoS One* **8**, e75397 (2013).
36. Pardo-Seco, J. *et al.* Genomic continuity of Argentinean Mennonites. *Sci. Rep.* **6**, 36392 (2016).
37. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
38. Dering, C., Hemmelmann, C., Pugh, E. & Ziegler, A. Statistical analysis of rare sequence variants: an overview of collapsing methods. *Genet. Epidemiol.* **35**(Suppl 1), S12–17 (2011).
39. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
40. Mi, H. *et al.* PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45**, D183–D189 (2017).
41. Justicia-Grande, A. J. *et al.* Development and Validation of a New Clinical Scale for Infants with Acute Respiratory Infection: The ReSVinet Scale. *PLoS One* **11**, e0157665 (2016).
42. Dopazo, J. *et al.* 267 Spanish Exomes Reveal Population-Specific Differences in Disease-Related Genetic Variation. *Mol Biol Evol* **33**, 1205–1218 (2016).
43. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
44. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
45. Thierry-Mieg, D. & Thierry-Mieg, J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.* **7**(Suppl 1), S1211–14 (2006).
46. Aronsson, F., Robertson, B., Ljunggren, H. G. & Kristensson, K. Invasion and persistence of the neuroadapted influenza virus A/WSN/33 in the mouse olfactory system. *Viral Immunol.* **16**, 415–423 (2003).
47. van Riel, D., Verdijk, R. & Kuiken, T. The olfactory nerve: a shortcut for influenza and other viral diseases into the central nervous system. *J Pathol* **235**, 277–287 (2015).
48. Shinya, K. *et al.* Subclinical brain injury caused by H5N1 influenza virus infection. *J. Virol.* **85**, 5202–5207 (2011).
49. van Riel, D. *et al.* Evidence for influenza virus CNS invasion along the olfactory route in an immunocompromised infant. *J. Infect. Dis.* **210**, 419–423 (2014).
50. Ng, Y. T., Cox, C., Atkins, J. & Butler, I. J. Encephalopathy associated with respiratory syncytial virus bronchiolitis. *J. Child Neurol.* **16**, 105–108 (2001).
51. van Riel, D., Verdijk, R. & Kuiken, T. The olfactory nerve: a shortcut for influenza and other viral diseases into the central nervous system. *J. Pathol.* **235**, 277–287 (2015).
52. Boivin, G. A. *et al.* Mapping of clinical and expression quantitative trait loci in a sex-dependent effect of host susceptibility to mouse-adapted influenza H3N2/HK/1/68. *J. Immunol.* **188**, 3949–3960 (2012).
53. Shivkumar, M. *et al.* Herpes simplex virus 1 targets the murine olfactory neuroepithelium for host entry. *J. Virol.* **87**, 10477–10488 (2013).
54. Kalbe, B. *et al.* Olfactory Receptors Modulate Physiological Processes in Human Airway Smooth Muscle Cells. *Front. Physiol.* **7**, 339 (2016).
55. Aisenberg, W. H. *et al.* Defining an olfactory receptor function in airway smooth muscle cells. *Sci. Rep.* **6**, 38231 (2016).
56. Li, J. J. *et al.* Activation of olfactory receptors on mouse pulmonary macrophages promotes monocyte chemotactic protein-1 production. *PLoS One* **8**, e80148 (2013).
57. Blackwell, J. M., Jamieson, S. E. & Burgner, D. HLA and infectious diseases. *Clin. Microbiol. Rev.* **22**, 370–385 (2009).
58. Ghangueux, J. P. & Palese, P. *Modulation of host gene expression and innate immunity by viruses.* (Springer, 2005).
59. Anderson, L. J. & Graham, B. H. *Challenges and opportunities for respiratory syncytial virus vaccines.* (Springer-Verlag, 2013).
60. Fjaerli, H. O. *et al.* Whole blood gene expression in infants with respiratory syncytial virus bronchiolitis. *BMC Infect. Dis.* **6**, 175 (2006).
61. Guo, X. *et al.* Respiratory syncytial virus infection upregulates *NLRC5* and major histocompatibility complex class I expression through RIG-I induction in airway epithelial cells. *J. Virol.* **89**, 7636–7645 (2015).
62. McDonald, J. U. *et al.* A simple screening approach to prioritize genes for functional analysis identifies a role for interferon regulatory factor 7 in the control of respiratory syncytial virus disease. *mSystems* **1** (2016).
63. Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V. & McGuckin, M. A. Mucins in the mucosal barrier to infection. *Mucosal Immunol.* **1**, 183–197 (2008).
64. Vinall, L. E. *et al.* Altered expression and allelic association of the hypervariable membrane mucin *MUC1* in *Helicobacter pylori* gastritis. *Gastroenterology* **123**, 41–49 (2002).
65. Baños-Lara Mdel, R., Piao, B. & Guerrero-Plata, A. Differential mucin expression by respiratory syncytial virus and human metapneumovirus infection in human epithelial cells. *Mediators Inflamm.* **2015**, 347292 (2015).
66. Currier, M. G. *et al.* *EGFR* interacts with the fusion protein of respiratory syncytial virus strain 2-20 and mediates infection and mucin expression. *PLoS Pathog.* **12**, e1005622 (2016).
67. Stokes, K. L. *et al.* The respiratory syncytial virus fusion protein and neutrophils mediate the airway mucin response to pathogenic respiratory syncytial virus infection. *J. Virol.* **87**, 10070–10082 (2013).

68. Lukacs, N. W., Strieter, R. M., Chensue, S. W., Widmer, M. & Kunkel, S. L. TNF-alpha mediates recruitment of neutrophils and eosinophils during airway inflammation. *J. Immunol.* **154**, 5411–5417 (1995).
69. Tekkanat, K. K. *et al.* IL-13-induced airway hyperreactivity during respiratory syncytial virus infection is STAT6 dependent. *J. Immunol.* **166**, 3542–3548 (2001).
70. Moore, M. L. *et al.* A chimeric A2 strain of respiratory syncytial virus (RSV) with the fusion protein of RSV strain line 19 exhibits enhanced viral load, mucus, and airway dysfunction. *J. Virol.* **83**, 4185–4194 (2009).
71. Lee, R. J. & Cohen, N. A. Taste receptors in innate immunity. *Cell Mol. Life Sci.* **72**, 217–236 (2015).
72. Malki, A. *et al.* Class I odorant receptors, *TAS1R* and *TAS2R* taste receptors, are markers for subpopulations of circulating leukocytes. *J. Leukoc. Biol.* **97**, 533–545 (2015).
73. Lee, R. J. *et al.* *T2R38* taste receptor polymorphisms underlie susceptibility to upper respiratory infection. *J. Clin. Invest.* **122**, 4145–4159 (2012).
74. Zhou, J. *et al.* Genetic association of *TLR4* Asp299Gly, *TLR4* Thr399Ile, and *CD14* C-159T polymorphisms with the risk of severe RSV infection: a meta-analysis. *Influenza Other Respir. Viruses* **10**, 224–233 (2016).
75. Torgerson, D. G. *et al.* Pooled Sequencing of Candidate Genes Implicates Rare Variants in the Development of Asthma Following Severe RSV Bronchiolitis in Infancy. *PLoS One* **10**, e0142649 (2015).
76. Goutaki, M. *et al.* The role of *TLR4* and *CD14* polymorphisms in the pathogenesis of respiratory syncytial virus bronchiolitis in greek infants. *Int. J. Immunopathol. Pharmacol.* **27**, 563–572 (2014).
77. Goutaki, M. *et al.* The role of *TLR4* and *CD14* polymorphisms in the pathogenesis of respiratory syncytial virus bronchiolitis in greek infants. *Int. J. Immunopathol. Pharmacol.* **27**, 563–572 (2014).

## Acknowledgements

This study received support from the Instituto de Salud Carlos III (Proyecto de Investigación en Salud, Acción Estratégica en Salud): project GePEM ISCIII/PI16/01478/Cofinanciado FEDER) (A.S.) and project ReSVinext ISCIII/PI16/01569/Cofinanciado FEDER (F.M.-T.); Consellería de Sanidade, Xunta de Galicia (RHI07/2-intensificación actividad investigadora, PS09749 and 10PXIB918184PR), Instituto de Salud Carlos III (Intensificación de la actividad investigadora 2007–2012, PI16/01569), Fondo de Investigación Sanitaria (FIS; PI070069/PI1000540) del plan nacional de I + D + I and ‘fondos FEDER’ (F.M.-T.), and 2016-PG071 Consolidación e Estructuración REDES 2016GI-1344 G3VIP (Grupo Gallego de Genética Vacunas Infecciones y Pediatría, ED341D R2016/021) (A.S. and F.M.-T). We would like to thank also the collaboration of the Translational Research Network in Pediatric Infectious Diseases (RITIP – [www.ritip.org](http://www.ritip.org)) and their members. We extend our gratitude to the children and parents/guardians who participated in this study and to the nursery and laboratory services from the Hospital Clínico Universitario de Santiago de Compostela for their effort.

## Author Contributions

A.S., F.M.-T. conceived and designed the study. M.C.-L., A.G.-C., P.O.-P., I.R.-C., M.J.-C., J.G.-R. and the GENDRES network were involved in sample recruitment and organization of DNA collections and clinical data. A.S., J.P.-S. and J.A., analyzed the data. A.S. wrote the first draft of the manuscript and it was revised by F.M.-T. and J.P.-S. All the authors contributed to the final version of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-15752-4>.

**Competing Interests:** The authors declare that they have no competing interests.

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



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

© The Author(s) 2017

## Consortia GENDRES network

Antonio Justicia-Grande<sup>3</sup>, Beatriz Morillo<sup>3</sup>, Lorenzo Redondo-Collazo<sup>3</sup>, Carmen Rodríguez-Tenreiro<sup>3</sup>, Ruth Barral-Arca<sup>1,2,3</sup>, Sara Pischedda<sup>1,2,3</sup>, José Peña-Gutián<sup>5</sup>, Carmen Curros Novo<sup>5</sup>, Miriam Puente-Puig<sup>5</sup>, Rosaura Leis-Trabazo<sup>5</sup>, Nazareth Martín-Torres<sup>3,5</sup>, José María Martín-Sánchez<sup>3,5</sup>, Máximo Francisco Fraga-Rodríguez<sup>6</sup>, José Ramón Antúnez<sup>6</sup>, Enrique Bernaola-Iturbe<sup>7</sup>, Laura Moreno-Galarraga<sup>7</sup>, Jorge Álvarez<sup>7</sup>, Teresa González-López<sup>8</sup>, Delfina Suarez-Vázquez<sup>8</sup>, Ángela Vázquez Vázquez<sup>8</sup>, Susana Rey-García<sup>8</sup>, Francisco Giménez-Sánchez<sup>9</sup>, Miguel Sánchez Forte<sup>10</sup>, Cristina Calvo-Rey<sup>11</sup>, María Luz García-García<sup>12</sup>, Ignacio Oulego-Erroz<sup>13</sup>, David Naranjo Vivas<sup>13</sup>, Santiago Lapeña<sup>13</sup>, Paula Alonso-Quintela<sup>13</sup>, Jorge Martínez-Sáenz de Jubera<sup>13</sup>, Estibaliz Garrido-García<sup>13</sup>, Cristina Calvo Monge<sup>14</sup>, Eider Oñate-Vergara<sup>14</sup>, Jesús de la Cruz Moreno<sup>15</sup>, María del Carmen Martínez-Padilla<sup>15</sup>, Manuel Baca-Cots<sup>16</sup>, David Moreno-Pérez<sup>17</sup>, Susana Beatriz-Reyes<sup>18</sup> & María Cruz León-León<sup>18</sup>

<sup>5</sup>Área Asistencial Integrada de Pediatría, Hospital Clínico Universitario, Santiago de Compostela, Galicia, Spain.

<sup>6</sup>Biobank, Servicio Anatomía Patológica, Hospital Clínico Universitario, Santiago de Compostela, Galicia, Spain.

<sup>7</sup>Servicio de Pediatría, Hospital Materno Infantil Virgen del Camino, Pamplona, Navarra, Spain. <sup>8</sup>Departamento de Pediatría, Complejo Hospitalario Universitario de Orense, Galicia, Spain. <sup>9</sup>Servicio de Pediatría, Hospital de la Inmaculada de Granada, Andalucía, Spain. <sup>10</sup>Servicio de Pediatría, Hospital Torrecárdenas, Almería, Andalucía, Spain.

<sup>11</sup>Servicio de Pediatría, Hospital La Paz, Madrid, Spain. <sup>12</sup>Servicio de Pediatría, Hospital Severo Ochoa de Madrid, Madrid, Spain. <sup>13</sup>Servicio de Pediatría, Complejo Asistencial Universitario de León, Castilla-León, Spain. <sup>14</sup>Servicio de Pediatría, Hospital de Donostia, San Sebastián, País Vasco, Spain. <sup>15</sup>Servicio de Pediatría, Complejo Hospitalario de Jaén, Andalucía, Spain. <sup>16</sup>Servicio de Pediatría, Hospital Quirón, Málaga, Andalucía, Spain. <sup>17</sup>Servicio de Pediatría, Hospital Carlos Haya, Málaga, Andalucía, Spain. <sup>18</sup>Servicio de Pediatría, Hospital Virgen de la Arrixaca, Murcia, Spain.