Respiratory Research

RESEARCH

Open Access

A genome-wide association study of adults with community-acquired pneumonia

Eva Suarez-Pajes[1†], Itahisa Marcelino-Rodriguez[1,2†], Elisa Hernández Brito[3†], Silvia Gonzalez-Barbuzano[1], Melody Ramirez-Falcon[1], Eva Tosco-Herrera[1], Luis A. Rubio-Rodríguez[4], María Luisa Briones[5], Olga Rajas[6], Luis Borderías[7], Jose Ferreres[8,9], Antoni Payeras[10], Leonardo Lorente[11], Javier Aspa[6], Jose M. Lorenzo Salazar[4], José Manuel Valencia-Gallardo[12], Nieves Carbonell[8,9], Jorge L. Freixinet[13], Felipe Rodríguez de Castro[12,14], Jordi Solé Violán[15,16,17], Carlos Flores[1,4,15,17*†] and Carlos Rodríguez-Gallego[3,14,15,17*†]

Abstract

Background Community-acquired pneumonia (CAP) is associated with high morbidity and hospitalization rate. In infectious diseases, host genetics plays a critical role in susceptibility and immune response, and the immune pathways involved are highly dependent on the microorganism and its route of infection. Here we aimed to identify genetic risk loci for CAP using a case-control genome-wide association study (GWAS).

Methods We performed a GWAS on 3,765 Spanish individuals, including 257 adult patients hospitalized with CAP and 3,508 population controls. Pneumococcal CAP was documented in 30% of patients; the remaining 70% were selected among patients with unidentified microbiological etiology. We tested 7,6 million imputed genotypes using logistic regressions. UK Biobank GWAS of bacterial pneumonia were used for results validation. Subsequently, we prioritized genes and likely causal variants based on Bayesian fine mapping and functional evidence. Imputation and association of classical HLA alleles and amino acids were also conducted.

Results Six independent sentinel variants reached the genome-wide significance ($p < 5 \times 10^{-8}$), three on chromosome 6p21.32, and one for each of the chromosomes 4q28.2, 11p12, and 20q11.22. Only one variant at 6p21.32 was validated in independent GWAS of bacterial and pneumococcal pneumonia. Our analyses prioritized *C4orf33* on 4q28.2, *TAPBP* on 6p21.32, and *ZNF341* on 20q11.22. Interestingly, genetic defects of *TAPBP* and *ZNF341* are previously known inborn errors of immunity predisposing to bacterial pneumonia, including pneumococcus and *Haemophilus influenzae*. Associations were all non-significant for the classical HLA alleles.

Conclusions We completed a GWAS of CAP and identified four novel risk loci involved in CAP susceptibility.

†Eva Suarez-Pajes, Itahisa Marcelino-Rodriguez and Elisa Hernández Brito contributed equally to this work and share first authorship.

Carlos Flores and Carlos Rodríguez-Gallego are senior authors.

*Correspondence:
Carlos Flores
cflores@ull.edu.es
Carlos Rodríguez-Gallego
jrodgal@gobiernodecanarias.org

Full list of author information is available at the end of the article

Keywords GWAS, Pneumonia, Host genetics, Inborn errors of immunity

Introduction

Infections are one of the main causes of death globally and nearly one in eight deaths continue to be due to bacterial infections [1]. Community acquired pneumonia (CAP) is considered a major public health problem due to its high morbidity and mortality [2, 3]. Yearly CAP incidence varies widely worldwide, with estimates between 11 and 169 cases per 10,000 persons. The available data for adults in Spain is 46.3 cases per 10,000 inhabitants [4]. The most common CAP complications are sepsis or severe respiratory failure, and mortality by CAP is mostly associated to hospitalized patients, reaching 23% in intensive care units (ICUs) [4, 5].

Host genetics plays a central role in the response to pathogens and contribute to explain the differences in susceptibility and severity among patients [6–8]. Specifically, single nucleotide polymorphism (SNP)-based heritability assessments in pneumonia support that genetic host factors explain a greater proportion of severity than of susceptibility [7]. However, there is a paucity of genetic studies aimed at identifying genetic factors involved in CAP susceptibility or prognosis. In addition, most of them have focused on candidate genes, especially on genes involved in the immune response such as those encoding the *MBL*, *SFTPA1*, *SFTPA2*, *SFTPD*, *IL6*, and *IL10*, among others [9, 10]. A few other studies have relied on genome-wide association studies (GWAS) to reveal pneumonia susceptibility loci in the human leukocyte antigen (HLA), *MUC5AC*, *IL6R*, and *TNFRSF1A* [11, 12], and pneumonia severity loci in *CFTR*, *R3HCC1L*, and *HBB* [7]. However, it must be noted that these studies have not distinguished the source of infection, implying that the patients are a heterogeneous mixture of patients with CAP and with nosocomial infections, i.e., with hospital-acquired pneumonia (HAP).

CAP is typically caused by several bacteria, including *Streptococcus pneumoniae* and *Haemophilus influenzae*, or viruses [13, 14]. *S. pneumoniae* is one of the leading causes of CAP and has been identified in about one-third of hospitalized cases in Europe, although these frequencies may be underestimated [13–16]. The immune response varies widely depending on the causative pathogen. Inborn errors of immunity (IEI, usually referred to as primary immunodeficiencies) strongly support that predisposition to infection by different microorganisms usually relies on different components of the immune system. For example, IEI impairing type I interferon-mediated immunity predispose to susceptibility to severe pneumonia by SARS-CoV-2 or influenza viruses, whereas IEI predisposing to pneumococcal infection are particularly involved in opsonization or phagocytosis of opsonized bacteria by splenic macrophages [6, 17, 18]. Therefore, studies aimed to identify the genetic basis of susceptibility or severity of infections may benefit from a precise homogenization of the source of infection and the causative microorganism.

To identify genetic variants associated with CAP, here we have conducted a GWAS of hospitalized patients with the only diagnosis of CAP, focusing on patients with pneumococcal infection or without identified causal microorganism.

Materials and methods

Study design

We conducted a one stage case-control GWAS of adult subjects of European ancestry from Spain (Fig. 1). No statistical calculation for adequate sample size was performed before the study was conducted. A total of 259 adult hospitalized patients with CAP diagnosis were recruited between March 2001 and 2016 from six Spanish hospitals and constituted the cases. These patients were included in previous candidate gene association studies of CAP [9, 19]. It is assumed that pneumococci cause most CAP cases in which negative test results were found using conventional microbiological methods [14, 16, 20]. Therefore, to keep homogeneous the causative microorganism of CAP, we only included patients with confirmed pneumococcal infection or those in whom no identified causative microorganism was identified. The study inclusion criteria and phenotype descriptions are available in the Supplementary material.

As controls, we used the genetic data available from 3,526 donors from the Spanish DNA Biobank (https://www.bancoadn.org), collected from the National Blood Service and have been used in a previous GWAS of severe COVID-19 [21]. All control participants were unrelated and clinically uncharacterized adults. They self-reported being of Spanish origin and having no personal or familial history of diseases, such as infectious, cancerous, circulatory, endocrine, mental, or behavioral, nervous, visual, auditory, respiratory, and immunological, among others.

Genotyping, quality control, and variant imputation

Both cases and controls were genotyped with the Axiom Spain Biobank Array (Thermo Fisher Scientific) following the manufacturer's instructions in the Santiago de Compostela Node of the National Genotyping Center (CeGen-ISCIII; http://www.usc.es/cegen). Genotyping quality control and variant imputation procedures are detailed in the Supplementary material. We conducted a principal component (PC) analysis (PCA) to derive the

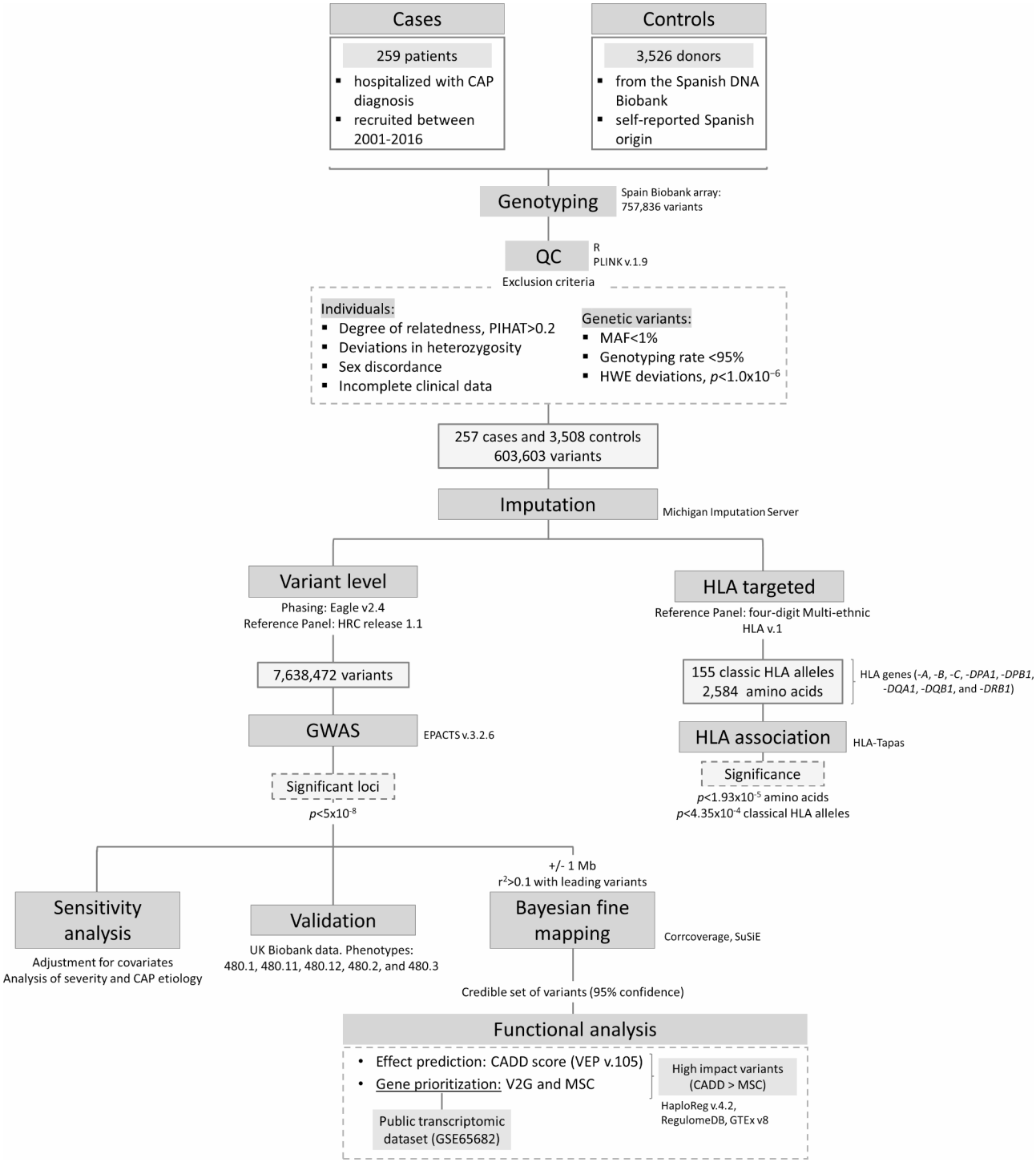

Fig. 1 Flowchart of the study design. CADD: Combined Annotation Dependent Depletion, CAP: Community-acquired Pneumonia, GWAS: Genome-wide association study, HLA: Human Leukocyte Antigen, HRC: Haplotype Reference Consortium, HWE: Hardy-Weinberg equilibrium, MAF: Minor Allele Frequency, MSC: Mutation Significance Cutoff, QC: Quality Controls, V2G: Variant to Gene, VEP: Variant Effect Predictor. UK Biobank phenotype codes: bacterial pneumonia (480.1), pneumococcal pneumonia (480.11), *Pseudomonas* pneumonia (480.12), Viral pneumonia (480.2), and (480.3) pneumonia due to fungus (mycoses)

main PCs for model adjustments and to identify genetic outliers (Figure S1). These procedures left us with a total of 603,603 genetic variants for 257 hospitalized CAP patients and 3,508 controls.

Statistical analysis and the functional assessment of associated loci

Variant association testing

To test the association of genetic variants with CAP, we used additive logistic regression models across all the imputed variants satisfying a good imputation quality (Rsq≥0.3) and a MAF≥1% using EPACTS v.3.2.6 [22]. The association model was adjusted for sex and the first 3 PCs and the results were assessed using the genomic inflation factor (λ) calculated with the *gap* package for R. Variant associations were considered statistically significant at a threshold $p<5.0\times10^{-8}$. Independent sentinel variants were identified as those surpassing $p<5.0\times10^{-8}$ and showing weak linkage disequilibrium (LD; $r^2<0.1$) with others in each locus after clumping in PLINK. The pseudo-R^2 of Nagelkerke was calculated based on a polygenic risk scores (PRS) model to determine the amount of variance being explained by the sentinel variants. Study power for detecting the statistically associated variants was assessed ad hoc. Detailed information is available in the Supplementary material.

A sensitivity analysis of the sentinel variants was conducted by including other covariates in the logistic regression model (e.g., age), the etiology of CAP, excluding the genetic outliers based on the PCA, and also testing their association with other two recorded severe pneumonia outcomes in the case series by separate and combined: (i) severe sepsis or septic shock according to the criteria available at the moment of the start of the recruitment [23]; and (ii) severe respiratory failure, defined as oxygen saturation<90% on room air, or a partial pressure of oxygen [PaO_2]<60 mmHg.

Since there is a lack of independent GWAS of CAP which can serve as a formal replication of the findings, we accessed publicly available summary GWAS of pneumonia of infectious, bacterial, fungal, and viral origin from UK Biobank (UKBB) data [24] to validate the association of the independent sentinel variants. We considered that the results were validated if the variant in the UKBB showed significance after Bonferroni correction for the number of variants assessed (p-value$<8.33\times10^{-3}$) and the same direction of effect as in our study.

Bayesian fine mapping and functional analysis

We performed a fine mapping on the association results in a 2 Megabase pairs (Mb) region around the independent sentinel variants to identify the credible variant set that most likely harbors the causal variant with 95% confidence. For this, we used the *corrcoverage* package [25] for R assuming that is only one causal variant in each locus. To validate the results, we also inferred the credible set with 95% confidence using the R package SuSiE [26] and assumed 3, 5, or 10 causal variants. Further details of the fine mapping are available in the Supplementary material.

We assessed the biological consequences of the variants included in the credible sets of the associated loci. In addition, we used the Variant-to-Gene (V2G) score to prioritize the genes that were most likely affected by the functional evidence based on data from the Open Targets Genetics portal [27]. To interpret the CADD score and predict the biological impact of the variants on the prioritized genes, we used the Mutation Significance Cutoff (MSC) v.1.6 of genes [28] with a confidence interval of 99%. For those variants deemed to predict a high biological impact (with a CADD>MSC), we performed an in silico analysis to determine their potential regulatory roles. Further details of the functional analysis are available in the Supplementary material.

To assess tissue expression of the genes prioritized in the associated loci and the existence of eQTL in the sentinel variants on artery, esophagus, lung, and whole blood, we used The Genotype-Tissue Expression (GTEx) Release v.8 data [29]. Protein quantitative loci (pQTLs) were evaluated using data available in Open Target Genetics. In parallel, we accessed a public transcriptomic dataset (GSE65682) available to assess expression differences in the genes prioritized in the associated loci among 108 ICU patients with CAP diagnosis and 42 healthy controls. Detailed information of this analysis is available in the Supplementary material.

Association of classic HLA alleles and amino acids

Due to the sequence complexity of the human leukocyte antigen (HLA) complex, its key role in immunity, and the previously reported association with pneumonia susceptibility [12], we also performed a targeted association testing of the variation in genes of the HLA complex with CAP. To this end, we imputed the genetic variation at eight classical HLA genes (*-A*, *-B*, *-C*, *-DPA1*, *-DPB1*, *-DQA1*, *-DQB1*, and *-DRB1*) and tested the association of the amino acids and the classical four-digit HLA alleles. The significance was established through the Bonferroni correction at $p<1.93\times10^{-5}$ for the amino acids and $p<4.35\times10^{-4}$ for the classical HLA alleles. Further information of this analysis is available in the Supplementary material.

Results

The study tested the association with CAP in a total of 7,638,472 variants from 3,508 controls and 257 patients (Fig. 2), in which *S. pneumoniae* was identified in 30% of cases, and the remaining 70% were patients where the

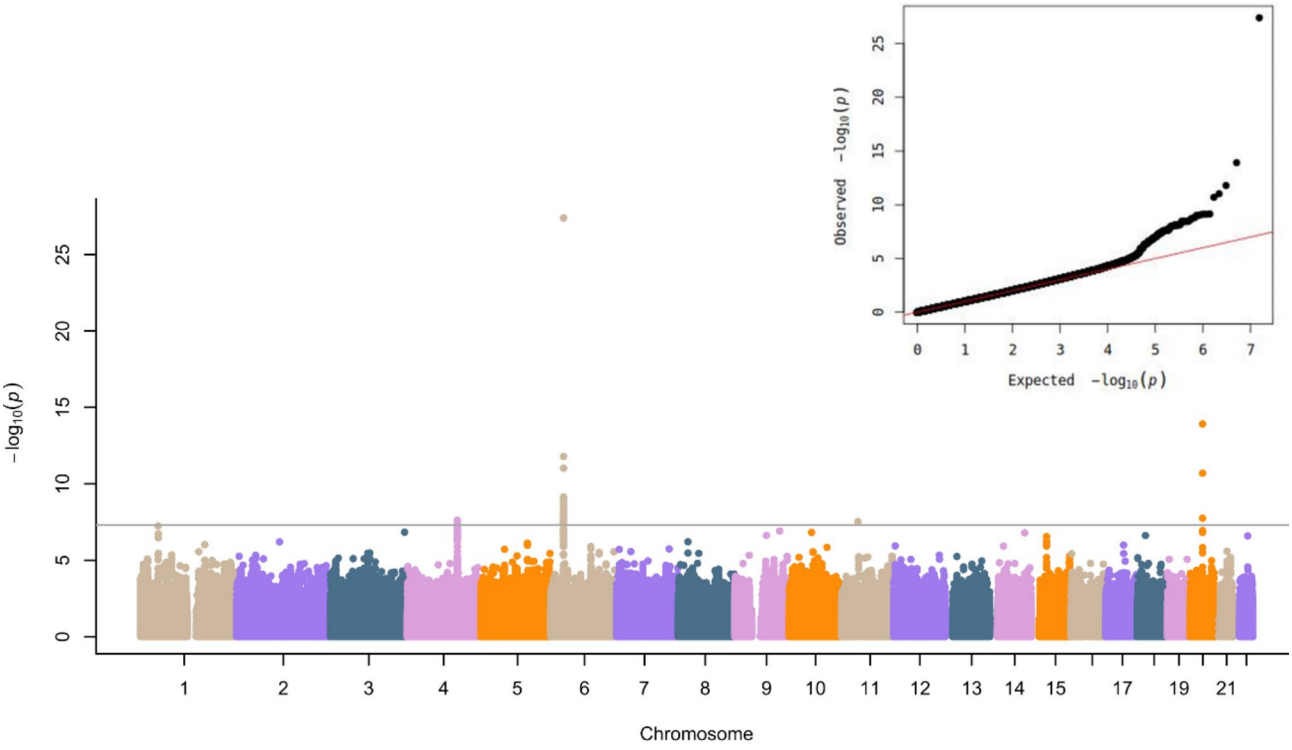

Fig. 2 Manhattan plot of the genome-wide association study results for adult CAP. The x-axis represents the chromosome positions while the y-axis represents the transformed *p*-values (–log10 [*p*-value]) obtained by logistic regression models using EPACTS v.3.2.6. Each point represents the result of a genetic variant analyzed. The variants that were deemed to be significantly associated with CAP exceed the significance threshold set at *p*-value = 5.0×10^{-8}, represented by the horizontal line. The inset on the top right represents the Quantile-Quantile (QQ) plot comparing the observed -log10 (*p*-values) from the GWAS analysis on the y-axis with the expected *p*-values under the null hypothesis on the x-axis. The genomic inflation factor (λ = 1.04) did not suggest inflation of the results obtained

Table 1 Results of the sentinel variants independently associated with CAP

Variant	Position	*p*-value	MAF	OR [95% CI]	A1/A2	Rsq	Func.Gene	Nearest gene(s)
rs34955650	4:130254264	2.41×10^{-8}	0.05	2.46 [1.80–3.38]	C/T	0.99	intergenic	*C4orf33*, LINC02466
rs213226	6:33209310	8.64×10^{-9}	0.17	1.84 [1.49–2.26]	A/G	0.36	intergenic	*RING1*, *HCG25*
rs456261	6:33258443	4.00×10^{-28}	0.02	8.23 [5.65–11.99]	G/A	0.33	intronic	*PFDN6*
rs2076775	6:33394253	7.26×10^{-10}	0.14	0.50 [0.41–0.63]	C/G	0.36	intronic	*SYNGAP1*
rs117203606	11:42330132	2.90×10^{-8}	0.11	0.21 [0.12–0.36]	G/A	Genotyped	intergenic	*LINC02740*, *HNRNPKP3*
rs45577437	20:32341041	1.21×10^{-14}	0.44	0.46 [0.37–0.56]	C/T	Genotyped	exonic	*ZNF341*

Position: chromosome and base pair according to GRCh37/hg19; MAF: minor allele frequency in the study population; OR: odds ratio; CI: confidence interval; A1: Non-effect allele; A2: Effect allele; Rsq: Imputation R squared; Func.Gene: gene location

causative microorganism was not identified. The clinical and demographic characteristics of the study subjects are shown in Supplementary Table 1. Overall, we did not detect inflation of the association results as the lambda of the study barely deviated from the expected under the null (λ = 1.04). Association testing revealed a total of 67 genome-wide significant variants (Supplementary Table S2) which were located on chromosomes 4q28.2, 6p21.32, 11p12, and 20q11.22. Regional association plots for these results are provided in the Supplementary Figure S2.

There were six independently associated sentinel variants (Table 1): rs34955650 at 4q28.2 ($p = 2.41 \times 10^{-8}$) intergenic to *C4orf33* and LINC02466; three at 6p21.32, rs456261 and rs2076775 that are both intronic to *PFDN6* ($p = 4.00 \times 10^{-28}$) and *SYNGAP1* ($p = 7.26 \times 10^{-10}$), respectively, and rs213226 that is intergenic to *RING1* and *HCG25* ($p = 8.64 \times 10^{-9}$); rs117203606 at 11p12 ($p = 2.90 \times 10^{-8}$) which is intergenic to LINC02740 and *HNRNPKP3*; and rs45577437 at 20q11.22 ($p = 1.21 \times 10^{-14}$) that is exonic to *ZNF341*. However, the results for rs456261 should be interpreted with caution given its unexpected low MAF (Table 1) compared with the gnomAD data from non-Finnish Europeans and because no other variant in the region was in strong LD with it. A PRS model using the six independent sentinel SNPs explained 11% variance in our study (p-value = 2.09×10^{-3}, OR = 1.67[95%CI: 1.53–1.82]).

However, we warn that this must be considered an overestimate and that independent studies should be ascertained to validate the results. An ad hoc statistical power calculation indicated that the study had > 80% power to detect variant associations with an OR > 1.80.

Models adjusting for other covariates did not substantially modify these results (Supplementary Table S3). The significance and effect size of the sentinel variants remained unaltered by the exclusion of the genetic outliers (Supplementary Table S4). Sensitivity analyses stratifying by the origin of CAP show that the associations of the six sentinel variants are maintained at least when patients have a non-definite CAP etiology, probably because there was still sufficient sample size to detect the effect of variants (Supplementary Table S5). The results for the association of the two severe pneumonia outcomes considered by separate or combined (i.e., severe sepsis or septic shock and severe respiratory insufficiency) for the six independently sentinel variants are shown in Supplementary Table S6. For these sub-analyses, despite the low sample size reduction, all variants reach significance at the nominal level and the direction of effect is still maintained.

To validate the results, we then evaluated the association of the independent sentinel variants in UKBB GWAS of pneumonia by bacteria, *Pseudomonas*, pneumococcus, fungi, or virus. Results for three out of the six independent sentinel variants had nominal significance ($p < 0.05$) and the same direction of effect. However, only rs2076775 was significantly associated with bacterial and pneumococcal pneumonia after Bonferroni correction (Supplementary Table S7).

Bayesian fine mapping by *corrcoverage* around each of the four chromosome loci to identify the most likely causal variants driving the association was able to delineate a credible set of 52 variants for 4q28.2 and 25 variants for 6p21.32 (Supplementary Table S8 and Figure S3). The variant from each credible set with the highest V2G score was used to assign the most likely gene involved in the association. At 6p21.32, the ranking prioritized the *TAPBP* gene (V2G max. score = 0.41 and MSC = 4.26), encoding the TAP binding protein. At 4q28.2, the *C4orf33* gene was prioritized (V2G max. score = 0.15 and MSC = 4.87). For downstream functional analysis, we selected those genetic variants with the highest probability of biological effect based on the CADD score from the delineated credible sets in the associated loci. For the 6p21.32 locus, we assessed 10 variants (CADD > MSC), and the predictions suggest a relevant biological impact since some of these SNPs predict the affectation of transcription factor binding and regulatory motif and DNase I hypersensitive sites. Furthermore, rs381847, rs2247385, and rs456261 may affect enhancer (H3K4me1 and H3K27ac) and promoter (H3K4me3 and H3K9ac) histone marks in several cell lines, including lung tissue and immune cells. Moreover, these 10 variants are eQTLs of immunity genes, including HLA genes and *TAPBP* (Supplementary Table S9). For the credible set of 4q28.2, we selected 13 variants with CADD > MSC. Two of these are rs17014611 and rs35004602, both linked to predictions of enhancer histone marks affectation in lung and immune cell lines, and DNase I hypersensitive sites in fetal lung (Supplementary Table S10). Fine mapping with SuSiE assuming different numbers of causal variants increased the number of SNPs within the credible sets and the uncertainty. The credible sets in the chromosomes 4q28.2 and 6p21.32 were also fragmented in two sets (Supplementary Figure S3) although it did not change the prioritized gene on each case. Neither *corrcoverage* nor SuSiE were able to provide credible sets for 11p12 and 20q11.22.

According to GTEx, the *TAPBP* gene is highly expressed (i.e., has higher average levels of Transcripts Per Millions) on spleen, lungs, lymphocytes, and whole blood. In fact, the three independent sentinel variants of 6p21.32 were pQTLs on blood plasma for the T*APBP* gene and eQTLs on artery, esophagus, lung, and whole blood (Supplementary Table S11). No significant eQTLs were found for any tissue at GTEx for the other three independent sentinel variants at chromosomes 4q28.2, 11p12, and 20q11.22 (rs34955650, rs117203606, and rs45577437). Furthermore, we found a significant upregulation of *TAPBP* (lowest $p = 1.44 \times 10^{-6}$ among the four probe sets available) and for *C4orf33* ($p = 9.67 \times 10^{-5}$) among ICU patients with CAP compared to controls (GEO: GSE65682). *ZNF341* did not show a significant gene expression difference (Supplementary Figure S4).

Finally, we assessed the association of 155 classical HLA alleles and the 2,584 amino acids for eight classical HLA genes with CAP. Despite the key implication of the HLA genes in infectious diseases, all associations tested were found non-significant (Supplementary Figure S5, and Table S12).

Discussion

Here, we describe the results of a GWAS of CAP conducted in Spanish population. We identified six independently associated variants from four chromosome loci (4q28.2, 6p21.32, 11p12, and 20q11.22) reaching genome-wide significance. Results for one of them rs2076775 (6p21.32) were validated with susceptibility to bacterial and pneumococcal pneumonia in independent studies with cases that were likely a mixture of CAP and HAP patients.

One of the independent variants detected is located on 4q28.2, intergenic to *C4orf33* and a long non-coding RNA (LINC02466) which has been involved in cancer [30]. Besides, the sentinel variants on chromosome 6

have been previously associated with platelet and blood cell count, type-I diabetes, or celiac disease, among other traits (GWAS data available at Open Target Genetics). The variant rs2076775, with evidence of validation in UKBB, is intronic to *SYNGAP1* which encodes for the synaptic Ras GTPase activating protein 1 which has been associated with neurodevelopmental disorders [31]. This variant has been previously associated ($p<5\times10^{-8}$) with hematopoietic cell count, rheumatoid arthritis, type 1 diabetes based on Open Targets Genetics data. Finally, another variant was prioritized in 11p12 intergenic to a long non-coding RNA and a pseudogene (*HNRNPKP3*). To our knowledge, these genes have not been associated with prior susceptibility or severity of infections. However, formal replication of results and further studies of functional characterization are needed to assess the biological effects of the identified genes and variants in pneumonia.

The sentinel variant of the 20q11.22 locus observed in this study is located in exon 5 of *ZNF341* and predicts a missense change (p.Pro185Ser). The association at this locus was also supported by two other genome-wide significant intronic variants residing in this gene. ZNF341 acts as a DNA-binding transcription factor, primarily as an activator of Signal Transducer and Activator of Transcription *STAT3* gene, and, to a lesser extent, a number of other genes such as *STAT1* [32–34]. Mutations at *ZNF341*, have been reported to cause of hyper-immunoglobulin E syndrome (HIES). HIES is characterized by elevated serum IgE levels, recurrent bacterial and *Candida* infections, eczema with cold staphylococcal skin abscesses, and other non-immunologic features that affect the skeleton, dentition, and connective tissue. HIES patients have recurrent skin and lung infections, including pneumonia, caused by *S. aureus*, *S. pneumoniae*, or *H. influenzae*, and the pulmonary recovery can involve abnormalities characterized by bronchiectasis and pneumatoceles [33–35]. Based on this evidence, variants affecting the function of *ZNF341* could also play an important role in CAP and its severity, although functional studies and replication cohorts are necessary to validate and describe the role of the repotted variants.

The study also prioritized possibly damaging variants in the *TAPBP* gene that were eQTLs for that gene in different tissues and pQTLs on plasma. It also revealed that *TAPBP* gene expression was upregulated among ICU patients with CAP compared to controls, although the role of profound inflammatory dysregulation in critical patients cannot be discarded. The *TAPBP* gene encodes the transporter associated with antigen processing (TAP) binding protein, also called tapasin. Tapasin is part of the peptide loading complex (PLC), which coordinates loading of high affinity peptides onto nascent HLA class I (HLA-I) molecules [36, 37]. Tapasin has a central role in the classical pathway of HLA-I presentation of endogenous peptides and in cross-presentation of exogenous antigens by professional antigen presenting cells [37]. Multiple HLA-I molecules have been reported to be tapasin-dependent, among them the *HLA-B*08:01* allele [38]. In spite that a previous GWAS found this class I HLA allele associated with pneumonia susceptibility [12], *HLA-B*08:01* was not associated with CAP in our study ($p=0.45$). Gene defects in *TAPBP* or in other PLC components cause an extremely rare IEI known as bare lymphocyte syndrome type I (BLS-I) [39–41]. Despite the clinical and biological heterogeneity, BLS-I patients usually develop symptoms well into late childhood, although some patients remain asymptomatic even in adulthood. They are characterized by recurrent respiratory tract infections, by *S. pneumoniae* and *H. influenzae* among others, associated with the development of bronchiectasis and respiratory insufficiency, as well as to cutaneous manifestations [39, 40, 42].

We acknowledge some limitations of the study. First and foremost, the study was based on a small sample of cases. Studies with larger sample sizes would permit providing a more robust estimation of the effects of the reported associations, especially for the low MAF variants, and possibly to reveal additional susceptibility loci beyond those identified here. Secondly, half of the patients had a low Pneumonia Severity Index. However, when the analyses only included the patients with severe sepsis or septic shock, or those with severe respiratory insufficiency, association results of the sentinel variants were maintained. Thirdly, our study lacks a formal assessment of the full spectrum of genetic variants (including other types of variation beyond SNPs and small INDELs, and variants with MAF lower than 1% which accumulate most of the disease-causing variation) for which complementary approaches based on Next-Generation Sequencing, such as whole exome or genome sequencing, would be ideal. Fourth, the precise causative pathogen of CAP was undetermined in 70% of the patients studied. Therefore, besides highlighting the interest in confirming the findings in patients with CAP by other pathogens, we could not draw conclusions regarding the associations with the response to a specific pathogen. Despite this, only one of the variants (rs34955650) did not reach the nominal statistical significance in the analysis of the reduced sample of CAP patients by *S. pneumoniae* infection. Furthermore, previous studies suggest that pneumococci cause most CAP cases in which negative test results were found using conventional microbiological methods, particularly in Spain [14, 16, 20]. However, infection by other bacteria and viruses is still expected. In addition, a substantial proportion of patients with CAP present coinfections, particularly pneumococcus [14, 20, 43]. Finally, we used controls that could have introduced

some confounding in the results since, despite hospitalization for CAP was not recorded for these donors during the recruitment, we cannot discard that they could develop it during their lifetime. Controls not ascertained for the disease risk are widely used in genetic studies for multiple infectious diseases, providing equivalent results as if controls comprise mild or asymptomatic patients [7, 8, 21]. In addition, it was not possible to perform additional sensitivity analyses or consider other variables of interest, such as smoking, due to the limited characterization of the personal history of health of these individuals.

Despite the limitations, this study has been completed for a specific sub-phenotype of pneumonia and is one of the first steps in understanding the genetics of this condition. Besides, mutations at *ZNF341* and *TAPBP*, cause previously known IEIs involved in the susceptibility to bacterial infections, particularly pneumonia. Thus, sequencing studies, together with experimental analyses of function, will be key to evaluate whether deleterious mutations in the novel CAP loci could define new IEIs predisposing to bacterial pneumonia [44]. Although the clinical application of the presented results is limited for the moment, they are valuable for downstream targeted analyses to potentially assist in CAP risk stratification and to inform of potential drug targets.

Conclusion

In summary, we report four novel loci associated with CAP, including two genes that were previously known to cause IEIs predisposing to bacterial pneumonia. Complementary studies are required to replicate the findings in larger studies and to better define the mechanistic links of these variants and genes on predisposition to pneumonia.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12931-024-03009-4.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We would like to thank CeGen for granting the access to the array data from the control samples provided by the "Banco Nacional de ADN-Instituto de Salud Carlos III (BNADN-ISCIII; www.bancoadn.org).

Author contributions

ES-P, IM-R, SG-B, MR-F, ET-H, LAR-R, JMLS performed the analyses, and experiments; EHB, MLB, OR, LB, JF, AP, LL, JA, JMV-G, NC, JF, FRC, JSV, CF and CG participated in data collection and clinical data; ESP, IM-R, CG, and CF wrote the first draft of the manuscript. ES-P, IM-R, EHB, SG-B, MR-F, JSV, CG, and CF refined the main and supplementary texts, figures, and tables of the manuscript. IM-R, CG and CF designed and supervised the study. CG and CF obtained funding. All authors read and approved the final manuscript.

Funding

Instituto de Salud Carlos III (PI13/01456, PI16/00759, PI17/00610, PI20/00876, PI23/00980) and Ministerio de Ciencia e Innovación (RTC-2017-6471-1; AEI/FEDER), co-financed by the European Regional Development Funds (ERDF), "A way of making Europe" from the EU; ITER agreements (OA17/008 and OA23/043); Grupo DISA (OA18/017), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (PIFIISC19/43); Fundación Mapfre-Guanarteme (OA19/072); Spanish Society of Pulmonology and Thoracic Surgery-Fundación Española del Pulmón; Cabildo Insular de Tenerife (CGIEU0000219140); and Agencia Canaria de Investigación, Innovación y Sociedad de la Información de la Consejería de Economía, Conocimiento y Empleo y por el Fondo Social Europeo (FSE) Programa Operativo Integrado de Canarias 2014–2020, Eje 3 Tema Prioritario 74 (85%) (TESIS2022010042 and TESIS2021010046). EH-B was supported by a grant from Universidad de Las Palmas de Gran Canaria. The Spain Biobank array genotyping service was carried out at CEGEN-PRB3-ISCIII; which is supported by PT17/0019, of the PE I + D + i 2013–2016, funded by Instituto de Salud Carlos III, and co-financed by ERDF.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

The study was conducted according to The Code of Ethics of the World Medical Association (Declaration of Helsinki), and the Research Ethics Committees from the coordinating centers approved this study (Hospital Universitario de Gran Canaria Dr. Negrín Ethics Committee FIS PI 16/00759; Hospital Universitario Nuestra Señora de Candelaria Ethics Committee PI-19/12). Written informed consent was obtained from all participants or their representatives.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Instituto de Investigación Sanitaria de Canarias (IISC), Santa Cruz de Tenerife, Spain
[2]Area of Preventive Medicine and Public Health, Universidad de La Laguna, Santa Cruz de Tenerife, Spain
[3]Department of Immunology, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, Spain
[4]Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain
[5]Department of Respiratory Diseases, Hospital Clínico y Universitario de Valencia, Valencia, Spain
[6]Department of Respiratory Diseases, Hospital Universitario de la Princesa, Madrid, Spain
[7]Department of Respiratory Diseases, Hospital Universitario San Jorge, Huesca, Spain
[8]Intensive Care Unit, Hospital Clínico de Valencia, Valencia, Spain
[9]INCLIVA Biomedical Research Institute, Valencia, Spain
[10]Department of Internal Medicine, Hospital Son Llatzer, Palma de Mallorca, Spain
[11]Intensive Care Unit, Hospital Universitario de Canarias, San Cristóbal de La Laguna, Spain
[12]Department of Respiratory Diseases, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, Spain
[13]Instituto Universitario de Investigaciones Biomédicas y Sanitarias, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[14]Department of Medical and Surgical Sciences, School of Medicine, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[15]Department of Clinical Sciences, University Fernando Pessoa Canarias, Las Palmas de Gran Canaria, Spain

[16]Critical Care Unit, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, Spain
[17]CIBER de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain

Received: 11 June 2024 / Accepted: 8 October 2024
Published online: 16 October 2024

References

1. Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399(10325):629–55.
2. Mortensen E, Metersky M. Long-term mortality after Pneumonia. Semin Respir Crit Care Med. 2012;33(03):319–24.
3. Theilacker C, Sprenger R, Leverkus F, Walker J, Häckl D, von Eiff C, et al. Population-based incidence and mortality of community-acquired pneumonia in Germany. PLoS ONE. 2021;16(6):e0253118.
4. Ferreira-Coimbra J, Sarda C, Rello J. Burden of Community-Acquired Pneumonia and Unmet Clinical needs. Adv Ther. 2020;37(4):1302–18.
5. Regunath H, Oba Y. Community-Acquired Pneumonia. 2024.
6. Casanova JL, Abel L. Lethal Infectious diseases as Inborn errors of immunity: toward a synthesis of the germ and genetic theories. Annu Rev Pathol. 2021;16(1):23–50.
7. Chen HH, Shaw DM, Petty LE, Graff M, Bohlender RJ, Polikowsky HG, et al. Host genetic effects in pneumonia. Am J Hum Genet. 2021;108(1):194–201.
8. Degenhardt F, Ellinghaus D, Juzenas S, Lerga-Jaso J, Wendorff M, Maya-Miles D, et al. Detailed stratified GWAS analysis for severe COVID-19 in four European populations. Hum Mol Genet. 2022;31(23):3945–66.
9. García-Laorden MI, Rodríguez de Castro F, Solé-Violán J, Rajas O, Blanquer J, Borderías L, et al. Influence of genetic variability at the surfactant proteins A and D in community-acquired pneumonia: a prospective, observational, genetic study. Crit Care. 2011;15(1):R57.
10. Kloek AT, Brouwer MC, van de Beek D. Host genetic variability and pneumococcal disease: a systematic review and meta-analysis. BMC Med Genomics. 2019;12(1):130.
11. Reay WR, Geaghan MP, Agee M, Alipanahi B, Bell RK, Bryc K, et al. The genetic architecture of pneumonia susceptibility implicates mucin biology and a relationship with psychiatric illness. Nat Commun. 2022;13(1):3756.
12. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat Commun. 2017;8(1):599.
13. Lanks CW, Musani AI, Hsia DW. Community-acquired Pneumonia and Hospital-acquired Pneumonia. Med Clin North Am. 2019;103(3):487–501.
14. Gadsby NJ, Russell CD, McHugh MP, Mark H, Conway Morris A, Laurenson IF, et al. Comprehensive Molecular Testing for Respiratory pathogens in Community-Acquired Pneumonia. Clin Infect Dis. 2016;62(7):817–23.
15. Pick H, Daniel P, Rodrigo C, Bewick T, Ashton D, Lawrence H, et al. Pneumococcal serotype trends, surveillance and risk factors in UK adult pneumonia, 2013–18. Thorax. 2020;75(1):38–49.
16. Torres A, Cillóniz C, Blasi F, Chalmers JD, Gaillat J, Dartois N, et al. Burden of pneumococcal community-acquired pneumonia in adults across Europe: a literature review. Respir Med. 2018;137:6–13.
17. Picard C, Puel A, Bustamante J, Ku CL, Casanova JL. Primary immunodeficiencies associated with pneumococcal disease. Curr Opin Allergy Clin Immunol. 2003;3(6):451–9.
18. Su HC, Jing H, Zhang Y, Casanova JL. Interfering with interferons: a critical mechanism for critical COVID-19 pneumonia. Annu Rev Immunol. 2023;41(1):561–85.
19. Solé-Violán J, Hernández-Brito E, Valencia-Gallardo JM, Herrera-Ramos E, Borderías L, Briones ML, et al. Challenges in understanding host genetics and severity of community-acquired pneumonia. ERJ Open Res. 2021;7(1):00745–2020.
20. Johansson N, Kalin M, Tiveljung-Lindell A, Giske CG, Hedlund J. Etiology of Community-Acquired Pneumonia: increased Microbiological Yield with New Diagnostic methods. Clin Infect Dis. 2010;50(2):202–9.
21. Cruz R, Diz-de Almeida S, López de Heredia M, Quintela I, Ceballos FC, Pita G, et al. Novel genes and sex differences in COVID-19 severity. Hum Mol Genet. 2022;31(22):3789–806.
22. Kang H. February. Efficient and parallelisable association container toolbox (EPACTS). (https://genome.sph.umich.edu/wiki/EPACTS), accessed 2022.
23. Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for Sepsis and Organ failure and guidelines for the use of innovative therapies in Sepsis. Chest. 1992;101(6):1644–55.
24. UKBiobank, ICD PheWeb. (https://pheweb.org/UKB-SAIGE/), accessed December 2023.
25. Hutchinson A, Watson H, Wallace C. Improving the coverage of credible sets in bayesian genetic fine-mapping. PLoS Comput Biol. 2020;16(4):e1007829.
26. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the Sum of single effects model. PLoS Genet. 2022;18(7):e1010299.
27. Open Targets Genetics. (https://genetics.opentargets.org/), accessed October 2022.
28. Itan Y, Shang L, Boisson B, Ciancanelli MJ, Markle JG, Martinez-Barricarte R, et al. The mutation significance cutoff: gene-level thresholds for variant predictions. Nat Methods. 2016;13(2):109–10.
29. The Genotype-Tissue. Expression (GTEx) Portal. (https://gtexportal.org/home/), accessed November 2022.
30. Li Y, Wei W, An S, Jiang J, He J, Zhang H, et al. Identification and analysis of lncRNA, microRNA and mRNA expression profiles and construction of ceRNA network in *Talaromyces marneffei* -infected THP-1 macrophage. PeerJ. 2021;9:e10529.
31. Llamosas N, Arora V, Vij R, Kilinc M, Bijoch L, Rojas C, et al. SYNGAP1 controls the maturation of dendrites, synaptic function, and Network Activity in developing human neurons. J Neurosci. 2020;40(41):7980–94.
32. August A. Who regulates whom: ZNF341 is an additional player in the STAT3/T_H17 song. Sci Immunol. 2018;3(24).
33. Béziat V, Li J, Lin JX, Ma CS, Li P, Bousfiha A et al. A recessive form of hyper-IgE syndrome by disruption of ZNF341-dependent STAT3 transcription and activity. Sci Immunol. 2018;3(24).
34. Frey-Jakobs S, Hartberger JM, Fliegauf M, Bossen C, Wehmeyer ML, Neubauer JC et al. ZNF341 controls STAT3 expression and thereby immunocompetence. Sci Immunol. 2018;3(24).
35. Minegishi Y. Hyper-IgE syndrome, 2021 update. Allergology Int. 2021;70(4):407–14.
36. Blees A, Januliene D, Hofmann T, Koller N, Schmidt C, Trowitzsch S, et al. Structure of the human MHC-I peptide-loading complex. Nature. 2017;551(7681):525–8.
37. Mantel I, Sadiq BA, Blander JM. Spotlight on TAP and its vital role in antigen presentation and cross-presentation. Mol Immunol. 2022;142:105–19.
38. Rizvi SM, Salam N, Geng J, Qi Y, Bream JH, Duggal P, et al. Distinct Assembly profiles of HLA-B molecules. J Immunol. 2014;192(11):4967–76.
39. Shrestha D, Szöllősi J, Jenei A. Bare lymphocyte syndrome: an opportunity to discover our immune system. Immunol Lett. 2012;141(2):147–57.
40. Su HC, Lenardo MJ. Combined Immune deficiencies. Stiehm's Immune deficiencies. Elsevier; 2014. pp. 143–69.
41. Yabe T, Kawamura S, Sato M, Kashiwase K, Tanaka H, Ishikawa Y, et al. A subject with a novel type I bare lymphocyte syndrome has tapasin deficiency due to deletion of 4 exons by Alu-mediated recombination. Blood. 2002;100(4):1496–8.
42. Zimmer J, Andrès E, Donato L, Hanau D, Hentges F, de la Salle H. Clinical and immunological aspects of HLA class I deficiency. QJM: Int J Med. 2005;98(10):719–27.
43. Liu YN, Zhang YF, Xu Q, Qiu Y, Lu QB, Wang T, et al. Infection and co-infection patterns of community-acquired pneumonia in patients of different ages in China from 2009 to 2020: a national surveillance study. Lancet Microbe. 2023;4(5):e330–9.
44. Tangye SG, Al-Herz W, Bousfiha A, Cunningham-Rundles C, Franco JL, Holland SM, et al. Human inborn errors of immunity: 2022 update on the classification from the International Union of Immunological Societies Expert Committee. J Clin Immunol. 2022;42(7):1473–507.

Publisher's note