The Journal of Infectious Diseases

PROMISE SUPPLEMENT

A Genome-Wide Association Study of Respiratory Syncytial Virus Infection Severity in Infants

Mari Johnson[1,2†], Irina Chelysheva[1,2†], Deniz Öner[3], Joseph McGinley[1,2], Gu-Lung Lin[1,2], Daniel O'Connor[1,2], Hannah Robinson[1,2], Simon B Drysdale[1,2], Emma Gammin[1,2], Sophie Vernon[1,2], Jill Muller[1,2], Helen Wolfenden[4]; Sharon Westcar[4], Lazarus Anguvaa[5], Ryan S Thwaites[6], Louis Bont[7], Joanne Wildenbeest[7], Federico Martinón-Torres[8,9,10], Jeroen Aerssens[3], Peter JM Openshaw[6], Andrew J Pollard[1,2]

[1]Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom; [2]NIHR Oxford Biomedical Research Centre and Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom; [3]Biomarkers Infectious Diseases, Janssen Pharmaceutica NV, Beerse, Belgium; [4]Royal Berkshire NHS Foundation Trust, Reading, United Kingdom; [5]Milton Keynes General Hospital NHS Trust, Milton Keynes, United Kingdom; [6]National Heart and Lung Institute, Imperial College London, London, United Kingdom; [7]Department of Paediatric Infectious diseases and Immunology, Wilhelmina Children's Hospital, University Medical Centre Utrecht, Utrecht, Netherlands; [8]Translational Pediatrics and Infectious Diseases, Department of Pediatrics, Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Galicia, Spain; [9]Genetics, Vaccines and Infections Research Group (GenViP), Instituto de Investigación Sanitaria de Santiago, Universidade de Santiago de Compostela, Santiago de Compostela, Galicia, Spain; [10]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain

Background: Respiratory syncytial virus (RSV) is a significant cause of infant morbidity and mortality worldwide; however, understanding the genetic risk factors of severe RSV is incomplete. Neutrophils and monocytes have been previously identified as major cell subsets involved in airway inflammation, however the pathophysiology of these events is also not fully characterised. Given that the majority of children experience at least one RSV infection by the age of two years,

[†]Contributed equally
mjohnso5@fredhutch.org

© The Author(s) 2024. Published by Oxford University Press on behalf of Infectious Diseases Society of America. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com This article is published and distributed under the terms of the Oxford University Press, Standard Journals Publication Model (https://academic.oup.com/pages/standard-publication-reuse-rights)

DOI: 10.1093/infdis/jiae029

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

but not all develop severe disease, we used genomic and transcriptomic data to explore potential mechanistic biomarkers of disease severity.

Methods: We conducted a genome-wide association study (GWAS) to investigate the genetic factors underlying RSV severity, assessed by the ReSVinet scale, in a cohort of 251 infants aged from 1 week old to 1 year of age. Genotyping data was collected from multiple European study sites as part of the RESCEU Consortium. Data and were analysed following quality control and genotype imputation using the TOPMed server. Generalised linear regression models were employed to assess the impact of genotype on RSV severity. Matrix eQTL in R was used to model the impact of candidate SNPs genotype on gene expression as measured by microarray.

Results: While no SNPs reached the genome-wide statistical significance threshold ($p < 5\times10^{-8}$), we identified 816 candidate SNPs with a p-value of $< 1\times10^{-4}$ and 75 SNPs with a p-value of $< 1\times10^{-5}$. Subsequent functional annotation of candidate SNPs highlighted variants previously identified in GWAS studies of inflammatory diseases, and genes relevant to neutrophil trafficking and cytoskeletal functions, including *LSP1 and RAB27A*. Furthermore, the eQTL analysis revealed that SNPs within the *RAB27A* locus significantly altered gene expression (FDR $p < 0.05$).

Conclusions: These findings may provide insights into genetic mechanisms driving severe RSV infection, offering biologically relevant information for future investigations.

Key Words: RSV, GWAS, eQTL, infection, inflammation

INTRODUCTION

Respiratory syncytial virus (RSV) causes seasonal epidemics of respiratory infections and is the most common cause of respiratory infection in neonates and young children worldwide. While most experience mild symptoms, for approximately 1 in 56 healthy term infants, infection can lead to hospitalisation, placing a significant burden on healthcare systems, particularly in low- and middle-income settings[1]. In severe cases, the infection can progress to the lower respiratory tract, causing bronchiolitis, pneumonia, and potentially fatal respiratory failure[2]. Identifying genetic risk factors for severe respiratory distress is therefore crucial for managing disease.

RSV is an enveloped, single-stranded, negative-sense RNA virus belonging to the sub-family Pneumoviridae. Its genome encodes 11 structural and non-structural proteins, essential for viral replication and immune evasion[3]. Infants and children with underlying medical conditions, such as preterm birth, chronic lung disease, or congenital heart disease, are particularly susceptible to severe RSV infections[4,5]. However, even without underlying risk factors, RSV infection can cause serious disease, and most hospital admissions are among previously healthy infants[6]. Currently, prevention strategies focus on passive immunisation with monoclonal antibodies to high-risk infants during the RSV season or maternal vaccination [7,8].

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

Early twin studies suggested a genetic component to developing severe RSV disease, as monozygotic twins reported a higher concordance between hospitalisation rates compared with dizygotic twins[9]. A candidate gene association study (CGAS) conducted in 2007 also identified several innate immune (*IFNA13, IL15, TLR8 and STAT1*) and chemotaxis genes (*CCL8, ITGB2,* and *VCAM1)* associated with RSV and influenza[10]. Recently, a genome-wide association study (GWAS) of RSV bronchiolitis and infection susceptibility have highlighted novel candidate single nucleotide polymorphisms (SNPs), including those within the *CXCR3* gene locus, and genes within olfactory and taste receptors[11,12]. Despite these findings, there is limited functional characterisation of genetic risk loci.

The majority of GWAS variants reside in non-coding regions of DNA and are thought to exert their effects through gene-transcription regulation. Expression quantitative trait loci (eQTL) analysis can model the effect of genotype on gene expression, highlighting potential causal variants[13]. No eQTL analysis of RSV infection severity has been conducted to date. Consequently, this study has two aims: first, to identify genes and pathways associated with the severity of RSV infection through a GWAS, and second, to perform eQTL analysis of SNPs associated with RSV severity.

METHODS

Study design

For our GWAS of RSV severity, we included participants who were recruited as part of the REspiratory Syncytial virus Consortium in EUrope (RESCEU) project, which has been described in detail previously (ClinicalTrials.gov identifier: NCT03756766; Jefferies et al.,2020)[14]. The study prospectively enrolled RSV-infected infants under 1 year of age who were otherwise healthy (N=500) from hospitals and community clinics across the Netherlands, Spain, and the United Kingdom during the 2017–2020 RSV seasons (from October to April).

Infants with pre-existing medical conditions including preterm birth (<37 weeks' gestation) were included as they were suspected to develop severe RSV (N=50). Exclusion criteria included infants who had previously received previous antivirals, steroids or montelukast within 7 days pre-enrolment, and/or were exposed to an investigational RSV vaccine or medication. From the recruited cohort, 261 cases were included for further quality control, and downstream analysis of RSV severity. Samples were selected based on clinical data completion and sample availability (Table. 1).

Infants were tested for RSV either through point-of-care testing using the Alere i RSV assay (Abbott) in both community and hospital settings or through routine tests such as rapid antigen detection or polymerase chain reaction (PCR) at a central laboratory, where available[15]. RSV disease severity was quantified at the time of recruitment through an extensive clinical

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

questionnaire meant to calculate a ReSVinet score. The score quantifies RSV disease severity on a range of 0-20 accounting for respiratory rate, blood oxygen, apnoea, dyspnoea, feeding intolerance and other markers[16]. Additional data included hospitalisation status and respiratory support requirements (Table. 1).

After presentation with RSV disease either in the hospital or at health centres, consent for genetic screening was acquired from parents according to the principles of ICH GCP and national clinical trial regulations. United Kingdom study authorisation was granted from the Health Research Authority (Integrated Research Application System project I.D 231136), and Research Ethics Committee approval has been granted by South Central and Hampshire A (National Research Ethics Service reference number 17/SC/0522). In Spain, this was granted by Comité de Ética de la Investigación de Santiago-Lugo (2017/395), and in the Netherlands by the Medical Ethical Committee, University Medical Center Utrecht (reference number 17/563).

Genome wide association study

DNA extraction

DNA was extracted from a buccal swab taken at the time of recruitment and stored in a DNA preservation buffer. Buccal swabs were stored at –20 degrees Celsius until required for extraction. Buccal swabs were compressed into the medium to ensure suspension of cheek cells in the medium. DNA were extracted from buccal swab medium using the QIAGEN DNeasy blood and tissue kit (Cat. No. / ID: 69504) using a QIASymphony machine to perform the extraction steps according to the standard DNeasy tissue extraction protocol.

SNP genotyping assay

All samples were genotyped on the Axiom PMRA platform from Affymetrix with a protocol recommended by the manufacturer.

SNP Quality control

Prior to association analysis, SNPs with high levels of missingness (>10%) were removed. Likewise, samples with a high proportion of missing data across all genotyped SNPs were removed (>10%), followed by a second round of SNP filtering (>5% missingness). Four samples at this stage failed QC. Additionally, we removed SNPs that significantly deviated from the expected allele frequency assuming the Hardy-Weinberg Equilibrium (HWE) ($P < 1\times10^{-5}$). Lastly, SNPs were aligned with the TOPMed Hg38 Freeze 5 reference genome[17]. SNPs were checked for strand alignment, and variants with mismatching frequencies ($R < 0.8$) were removed in addition to AT-CG and (insertion-deletion) INDEL SNPs.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

Sample relatedness and population structure

By calculating identity by descent after pruning variants for linkage disequilibrium ($R < 0.2$) in PLINK, we estimated the genetic relatedness of our participants. One sample from each pair that shared alleles at a higher than the expected frequency was removed ($\pi > 0.185$). In total, four sibling pairs were detected. To account for population structure when performing association analyses, the first 6 principal components were included, as determined by the BE algorithm[18] (Suppl. Fig.1). Outlying samples were removed if they deviated +- 2SD from the mean principal component value, and two outlier samples were removed at this stage.

Imputation

Haplotypes were phased using the Eagle v2 algorithm with the TOPMed Hg38 Freeze 5 reference genome, and Minimac4 imputed SNP data using the TOPMed server[19]. Variants were filtered based on allele frequency concordance between the study data and the reference genome ($R > 0.7$) and re-filtered for missingness and HWE. SNPs with a minor allele frequency of <1% were excluded from the association analysis based on statistical power.

GWAS modelling

In order to estimate if genotype was associated with RSV severity in our cohort, we fitted a generalised linear regression model of genotype versus RSV severity as determined by ReSVinet scores (0-20) using PLINK 2.0. Covariates included participant age, sex, pre-existing conditions including preterm birth, plus the first 6 principal components. The significance of the effect size of genotype upon ReSVinet score was estimated by comparing the model slope with a null model using a likelihood ratio test under a Chi-squared distribution.

Functional annotation of candidate snps

SNPs identified in the GWAS of RSV severity that were significant above $P < 1\times10^{-4}$ were input into the FUMA v1.4.1 and SNPNexus variant annotation pipelines[20,21]. Features included predicted variant deleteriousness, and those previously reported in the GWAS catalogue[22]. Chromatin sites and regulatory features of loci were examined using RegulomeDB and the Roadmap database[23,24], and the Reactome database was used to examine pathway enrichment of SNPs, using the Fisher's Exact Test to determine significance[21]. Variants with prior GWAS associations were queried in the GTEx V8 database for tissue-specific eQTLs in whole blood and lung tissue[25].

Expression Quantitative Trait Loci (eQTL) Analysis

RNA extraction and transcriptomic analysis

Blood samples were taken within 48 hours of admission, and RNA was extracted from whole blood using QIAsymphony PAXgene Blood-RNA Kit (QIAGEN) according to the manufacturer's instructions and processed for ClariomTM™ GOScreen microarray (Thermo Fisher) after quality

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

control. First-strand cDNA was synthesised using Poly-dT and random primers with a 5′-adaptor sequence and a 3’-adaptor for single-stranded cDNA, followed by low-cycle PCR amplification for labelling with the GeneChip Pico Reagent kit. In vitro transcription (IVT), then amplified antisense mRNA (cRNA) for a second cDNA synthesis round, producing double-stranded cDNA. The targets were hybridised onto a single GO Screen plate after fragmentation, denaturation, and end-labeling according to manufacturer instructions. A GeneChip Fluidics Station 450 and GeneChip scanner 3000 7G stained and scanned single sample cartridge arrays, while GeneTitan Multi-Channel Instruments stained and imaged array plates.

Expression quantitative trait loci analysis

Genotypes from candidate SNPs ($P < 1\times10^{-4}$), were modelled with gene expression using a linear regression model using MatrixEQTL in R[26]. Covariates included age, sex, pre-existing conditions (including premature birth), genetic principal components 1–6, plus transcriptome principal components 1-9, as determined by PCA[18]. SNPs located within 1 Mb of the gene transcription start site were classified as cis-eQTL. The false discovery rate (FDR) was used to adjust for multiple testing, and eQTLs of relevant variants were explored using R v. 4.2, and statistical tests between groups were performed using the Wilcoxon test.

Code availability

Full details of these scripts and methods can be found at: https://github.com/mari-lynne/RSV_GWAS

RESULTS

GWAS of RSV severity in a paediatric cohort

Cohort characteristics

Between the years 2017–2020, 261 RSV infected infants were recruited across the United Kingdom, Netherlands, and Spain, and enrolled for genomic and transcriptomic analyses following infection (Methods; Study Design). The median age of our study cohort was ~3 months old (Table. 1). Following diagnosis, participants were observed for clinical symptoms including temperature and wheeze and need for mechanical ventilation which was combined to give a composite ReSVinet score. Forty infants were recruited with pre-existing conditions, including 14 preterm infants (<37 weeks) (Table. 1).

When evaluating the genetic ancestry of our study cohort, the majority of participants clustered within the European super-population (Suppl. Fig. 1). To account for population structure, principal components (PC) 1-6 were included in the final regression model, and two participants with outlying PCA distribution +/- 2 SD from the mean PC were removed (Suppl. Fig. 1). Four

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

siblings were also removed to account for sample relatedness. Genotyping data were evaluated for SNP missingness and sample quality, following which an additional 4 samples were excluded. In total 251 participants were included from the initial 261 RSV cases used for GWAS analysis, with a median ReSVinet score of 8 (Table. 1).

Host genetic factors involved in RSV infection severity

To investigate genetic factors that may contribute to RSV severity, we conducted a GWAS using a linear model of ReSVinet , a clinical composite score which evaluates respiratory disease severity from a scale of 0-20. Age and previous comorbidities including preterm birth were included as covariates in the GWAS model, along with principal components 1-6. Following imputation, variants were filtered to include those with a minor allele frequency > 0.01, which left 8,917,131 SNPs for downstream analyses. No variants associated with RSV infection severity at the genome-wide significance level ($P > 5\times10^{-8}$). However, given the relatively small sample size, our study was underpowered to detect significant associations of moderate effect sizes (β 0.2) using the standard genome-wide threshold[27].

When subsequently examining the Manhattan plot, genetic loci with greater significance were discernible (Fig. 1). Further examination of the distribution of P-values under the observed and null hypothesis, showed deviations from expected P-values at those below $P < 1x10^{-4}$ (Suppl. Fig 2). Based on these data we used a suggestive significance threshold of $P < 1 \times 10^{-4}$ for downstream analysis. In total, 816 variants associated with RSV severity were identified below the suggestive significance threshold of $P < 1x10^{-4}$, and 75 variants had a P value $< 1x10^{-5}$ (Fig. 1).

Functional annotation and pathway analysis of candidate snps associated with RSV infection severity

To identify variants that might be relevant in determining the severity of RSV infection, we explored genomic features and public GWAS data of candidate SNPs using SNPNexus and FUMA[20,21]. Interestingly, the most significant SNPs surrounding the *LSP1* gene locus on chromosome 11 region p15.5 ($P < 1\times10^{-6}$) were located upstream of the gene's promoter site (Suppl. Fig. 4). Furthermore, *LSP1* variants observed in our study were previously identified in GWAS studies of inflammatory bowel disease, ulcerative colitis, and white blood cell counts, suggesting a role for these SNPs in attenuating inflammation (Suppl. Table 1). Additionally, variant rs3843306 in chromosome 1 region p22.2 was associated with pulmonary function decline in asthmatics[28], while a chromosome 10 variant (10p12.33), rs2477664, in the *MRC1* (macrophage mannose receptor) gene locus was reported previously in a GWAS of haematological traits (Suppl. Table 1). Taken together, these results highlight potential candidate genes which may be relevant in the immune response during RSV infection.

In order to evaluate whether GWAS-associated genes were related to biologically relevant functions, we performed a pathway enrichment analysis of genes mapped to their closest variant within 10 KB of a transcription start or stop site[21]. One hundred thirteen GWAS-related genes

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

were evaluated against genes annotated in the Reactome pathway database for over-enrichment[29] (Suppl. Table 1). Immune system-related pathways including antigen cross-presentation and butyrophilin (BTN) family interactions (*MRC1, PSMB1, BTNL8*), were significantly enriched for GWAS-associated genes, while enriched disease pathways included genes involved in transmembrane support (*ABCC6, PSMB1*). The inactivation of CDC42 and RAC1 was also highlighted by the associations of SNPs in the *SRGAP3* gene locus, a GTPase-activator protein for RAC1 (Suppl. Table 1).

Expression quantitative trait loci analysis of candidate variants associated with RSV infection severity

We identified that a significant proportion of our candidate variants were located within histone regions, as well as identifying several SNPs that were within transcription factors (Suppl. Fig. 3). In order to investigate the influence of these variants upon gene transcription, we performed an eQTL analysis of candidate SNPs with gene expression measured in whole blood taken at clinical admission. In total, 816 SNPs were evaluated against approximately 1600 protein-coding genes measured on Affymetrix microarray chips. Cis-eQTLs included SNPs that were located within 1 Mb of a gene locus. Matrix eQTL was employed to test associations for gene transcription, with age, sex, and genetic principal components as model covariates.

After correction for multiple testing, we identified two variants in linkage disequilibrium (rs11071126, rs1501398), located upstream of the *RAB27A* gene that were significant eQTL. These variants were associated with reduced RSV severity (GWAS $P < 1.39\times10^{-5,}$ $\beta = -1.83$), and a decrease in RAB27A expression in infected infants (*adj. P* = 0.045, $\beta = -0.29$) (Fig. 2). As *LSP1* gene expression was not significant for eQTL in the blood, we examined *LSP1* variants (11p15.5) for associations with tissue-specific gene expression using the GTEx database[25]. Subsequently, the *LSP1* variants were significantly associated with differences in *LSP1* expression both in lung tissue and whole blood (rs907612, $P = 5.2 \times10^{-17}$, $P = 2.36 \times10^{-9}$). SNP mapping revealed that these variants were located within regulatory regions of the gene locus, including transcription start sites (Suppl. Fig. 4).

DISCUSSION

Here we present the first eQTL analysis of candidate SNPs associated with RSV severity, highlighting variants in the *RAB27A* gene locus that were associated with reduced RSV severity and lower gene expression. RAB27A is a small GTPase and a key regulator of neutrophil activation and exocytosis, and overexpressed during severe RSV [30,31]. Moreover, in *RAB7A/B* knock out mouse models, neutrophil recruitment to the lung was significantly reduced[32]. Neutrophil infiltration in the lung is a hallmark of severe RSV infection, therefore variants influencing these pathways are of particular interest. Additionally, genome annotation highlighted variants within the *SRGAP3* locus associated with reduced RSV severity. *SRGAP3* encodes another GTPase-

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

activating protein linked to the inactivation of Rac1 and CDC42. Recently, Malhi et al., 2021, observed that disrupting Rho-GTPases and Rac1 activity through statins reduced RSV infection *in vitro*, highlighting potential therapeutic pathways to explore[33].

While no genome-wide significant hits were identified, we did observe several candidate variants associated with RSV severity which have previously been reported in GWAS of inflammatory traits, including variants in the *MRC1*, *BTNL8* and *LSP1* gene loci. Lymphocyte-specific protein 1, or *LSP1*, is primarily expressed in neutrophils, monocytes and T-cells, and its expression significantly increases during severe RSV paediatric infection, where it plays a crucial role in neutrophil activation and migration in the lung[31,34]. During severe infections, increased recruitment of neutrophils in the respiratory tract can cause the release of proinflammatory cytokines, chemokines, and reactive oxygen species[35].

Subsequently, *LSP1* gene knockouts in asthmatic mice, reduced lung neutrophil infiltration, demonstrating how high levels of *LSP1* induced neutrophil migration may be detrimental to disease progression[36]. In our cohort, we observed that *LSP1* GWAS variants were associated with differential gene expression in the lung and whole blood using data from the GTEx consortium[25], however it was not a significant eQTL in our dataset. Larger sample sizes and potentially different tissue expression will therefore be needed to confirm these associations.

Pathway enrichment analysis of GWAS loci also highlighted several immune-related genes including *MRC1*, *PSMB1*, and *BTNL8*. The *MRC1* gene encodes the mannose receptor C-type 1, which is involved in macrophage activation and Class 1 antigen presentation[37]. In our study, the *MRC1* variant rs2477664 was associated with reduced severity during RSV infection. Furthermore, examining single cell sequencing data from lung tissue revealed that *MRC1* and *PSMB1* are highly expressed in alveolar macrophages, which are important in controlling RSV replication and progression through the respiratory tract [38,39]. Therefore, it is possible that early antigen presentation might be an important factor in controlling RSV disease. Interestingly, Pasanen et al., 2017, also identified genes in overlapping pathways including antigen presentation, and highlighted genes also present in our study, including *DSCAM* and *TRABD2B* which have yet to be explored in the context of RSV infection[11].

The main limitation of this study was the small sample size relative for a GWAS study and the population's homogeneity. Given the high number of comparisons made across hundreds of thousands of SNPs, large sample sizes are typically needed to detect robust associations with disease. Consequently, examination of these findings in new cohorts will be critical to confirming these results. Only two other GWAS studies of RSV severity have been conducted at the time of writing, using bronchiolitis and prolonged RSV infection as the diagnostic criterion for severe RSV, respectively[11,40]. While neither of these studies detected any genome-wide significant results, they may be useful for further meta-analyses. Future research may continue to investigate the role of GWAS associated genes in the host-immune response to RSV, particularly those linked to neutrophil pathways and antigen presentation, shedding light on potential therapeutic avenues.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

Footnotes

Conflicts of Interest: FMT and PJMO have received honoraria from GSK group of companies, Pfizer, Inc., Sanofi Pasteur, MSD, Seqirus, AstraZeneca, Moderna and Janssen for taking part in advisory boards and expert meetings and for acting as a speaker in congresses outside the scope of the submitted work. FMT has also acted as principal investigator in randomized controlled trials of the above-mentioned companies as well as Ablynx, Gilead, Regeneron, Roche, Abbott, Novavax, and MedImmune, with honoraria paid to his institution.

SBD has received honoraria from MSD and Sanofi for taking part in RSV advisory boards and has provided consultancy and/or investigator roles in relation to product development for Janssen, AstraZeneca, Pfizer, Moderna, Valneva, MSD, iLiAD and Sanofi with fees paid to St George's, University of London. SBD is a member of the UK Department of Health and Social Care's (DHSC) Joint Committee on Vaccination and Immunisation (JCVI) RSV subcommittee and Medicines and Healthcare products Regulatory Agency's (MHRA) Paediatric Medicine Expert Advisory Group (PMEAG), but the reviews expressed herein do not necessarily represent those of DHSC, JCVI, MHRA or PMEAG. AJP is chair of Department of Health and Social Care's (DHSC) Joint Committee on Vaccination and Immunisation. JGW participated in the advisory board of Janssen and Sanofi with fees paid to UMCU and has been an investigator for clinical trials sponsored by pharmaceutical companies including AstraZeneca, Merck, Pfizer, Sanofi, and Janssen. All funds have been paid to UMCU.

Acknowledgements: This work was supported by the Respiratory Syncytial Virus Consortium in Europe (RESCEU) project and Preparing for RSV Immunisation and Surveillance in Europe (PROMISE). RESCEU and PROMISE have received funding from the Innovative Medicines Initiative 2 Joint Undertaking (Grant Numbers 116019 and 101034339, respectively). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

References

1. Wildenbeest JG, Billard MN, Zuurbier RP, et al. The burden of respiratory syncytial virus in healthy term-born infants in Europe: a prospective birth cohort study. The Lancet Respiratory Medicine. 2023;11(4):341-353.
2. Florin TA, Plint AC, Zorc JJ. Viral Bronchiolitis. The Lancet. 2017;389(10065):211-224.
3. Collins PL, Fearns R, Graham BS. Respiratory Syncytial Virus: Virology, Reverse Genetics, and Pathogenesis of Disease. In: Knipe DM, Howley PM, eds. Fields Virology. 6th ed. Vol 1. Philadelphia, PA: Wolters Kluwer Health; 2018:1303-1340.
4. Welliver RC, Checchia PA, Bauman JH, Fernandes AW, Mahadevia PJ, Hall CB. Fatality Rates in Published Reports of RSV Hospitalizations among High-Risk and Otherwise Healthy Children. Current Medical Research and Opinion. 2008;24(1):235-241.
5. Hall CB, Weinberg GA, Iwane MK, et al. The Burden of Respiratory Syncytial Virus Infection in Young Children. New England Journal of Medicine. 2009;360(6):588-598.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

6. Hall CB, Weinberg GA, Blumkin AK, et al. Respiratory syncytial virus-associated hospitalizations among children less than 24 months of age. Pediatrics. 2013;132(2):e341–e348.
7. Wu H, Pfarr DS, Tang Y, et al. Therapeutic Antibodies as a Treatment Option for Respiratory Syncytial Virus: A Review of the Current Status of Palivizumab and New Agents under Development. Drugs. 2020;80(1):9-28.
8. Mazur NI, Terstappen J, Baral R, et al. Respiratory syncytial virus prevention within reach: the vaccine and monoclonal antibody landscape. The Lancet Infectious Diseases. 2023;23(1):e2–e21.
9. Thomsen SF, Stensballe LG, Skytthe A, Kyvik KO, Backer V, Bisgaard H. Increased Concordance of Severe Respiratory Syncytial Virus Infection in Identical Twins. Pediatrics. 2008;121(3):493-496.
10. Janssen R, et al. Genetic Susceptibility to Respiratory Syncytial Virus Bronchiolitis Is Predominantly Associated with Innate Immune Genes. The Journal of Infectious Diseases. 2007;196(6):826-834.
11. Pasanen A, Karjalainen M, Bont L, et al. Genome-Wide Association Study of Polymorphisms Predisposing to Bronchiolitis. Scientific Reports. 2017;7:41653.
12. Salas A, Pardo-Seco J, Cebey-López M, et al. Whole exome sequencing reveals new candidate genes in host genomic susceptibility to respiratory syncytial virus disease. Scientific Reports. 2017;7(1):1-13.
13. Veyrieras JB, Kudaravalli S, Kim SY, et al. High-Resolution Mapping of Expression-QTLs Yields Insight into Human Gene Regulation. PLOS Genetics. 2008;4(10):e1000214.
14. Jefferies K, Drysdale SB, Robinson H, et al. Presumed Risk Factors and Biomarkers for Severe Respiratory Syncytial Virus Disease and Related Sequelae: Protocol for an Observational Multicenter, Case-Control Study From the Respiratory Syncytial Virus Consortium in Europe (RESCEU). The Journal of infectious diseases. 2020;222(Suppl 7):S658–S665.
15. Peters RM, Schnee SV, Tabatabai J, Schnitzler P, Pfeil J. Evaluation of Alere i RSV for Rapid Detection of Respiratory Syncytial Virus in Children Hospitalized with Acute Respiratory Tract Infection. Journal of clinical microbiology. 2017;55(4):1032–1036.
16. Justicia-Grande AJ, Pardo-Seco J, Cebey-López M, Vilanova-Trillo L, Gómez-Carballa A, Rivero-Calle I, Puente-Puig M, Curros-Novo C, Gómez-Rial J, Salas A, Martinón-Sánchez JM, Redondo-Collazo L, Rodríguez-Tenreiro C, Martinón-Torres F; Respiratory Syncytial Virus network (ReSVinet). Development and Validation of a New Clinical Scale for Infants with Acute Respiratory Infection: The ReSVinet Scale. PLoS One. 2016 Jun 21;11(6):e0157665.
17. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nature Genetics. 2021;53(7):1097–1103.
18. Zhou HJ, Li L, Li Y, et al. PCA outperforms popular hidden variable inference methods for molecular QTL mapping. Genome Biology. 2022;23:210.
19. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: Faster genotype imputation. Bioinformatics. 2015;31(5):782–784.
20. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nature Communications. 2017;8:1, 8(1):1–11.
21. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic Acids Research. 2020;48(W1):W185-W192.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

22. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nature Protocols. 2015;10(10):1556–1566.
23. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. Genome Research. 2012;22(9):1790-1797.
24. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317-330.
25. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580-585.
26. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28(10):1353-1358.
27. Hoggart, C. J., Clark, T. G., De Iorio, M., Whittaker, J. C., & Balding, D. J. (2008). Genome-wide significance for dense SNP and resequencing data. Genetic epidemiology, 32(2), 179–185.
28. Imboden M, Bouzigon E, Curjuric I, et al. Genome-wide association study of lung function decline in adults with and without asthma. J Allergy Clin Immunol. 2012;129(5):1218-1228.
29. Gillespie M, Jassal B, Stephan R., Milacic M, Rothfels K, Senff-Ribeiro A, Griss J, Sevilla C, Matthews L, Gong C, Deng C, Varusai T, Ragueneau E, Haider Y, May B, Shamovsky V, Weiser J, Brunson T, Sanati N, Beckman L, D'Eustachio P. (2022). The reactome pathway knowledgebase 2022. Nucleic acids research, 50(D1), D687–D692.
30. Catz SD. The role of Rab27a in the regulation of neutrophil function. Cellular microbiology. 2014;16(9):1301-1310.
31. Dapat C, Kumaki S, Sakurai H, Nishimura H, Labayo HKM, Okamoto M, Saito M, Oshitani H. Gene signature of children with severe respiratory syncytial virus infection. Pediatric Research. 2021 May;89(7):1664-1672.
32. Singh RK, Furze RC, Birrell MA, Rankin SM, Hume AN, Seabra MC. A role for Rab27 in neutrophil chemotaxis and lung recruitment. BMC Cell Biol. 2014;15:39
33. Malhi M, Norris MJ, Duan W, Moraes TJ, Maynes JT. Statin-mediated disruption of Rho GTPase prenylation and activity inhibits respiratory syncytial virus infection. Communications Biology. 2021;4(1):1239.
34. Le NPK, Channabasappa S, Hossain M, Liu L, Singh B. Leukocyte-specific protein 1 regulates neutrophil recruitment in acute lung inflammation. American Journal of Physiology. Lung Cellular and Molecular Physiology. 2015;309
35. Carvajal JJ, Avellaneda AM, Salazar-Ardiles C, Maya JE, Kalergis AM, Lay MK. Host Components Contributing to Respiratory Syncytial Virus Pathogenesis. Frontiers in Immunology. 2019 Sep 12;10:2152.
36. Le NPK, do Nascimento AF, Schneberger D, et al. Deficiency of leukocyte-specific protein 1 (LSP1) alleviates asthmatic inflammation in a mouse model. Respiratory Research. 2022;23:165.
37. Taylor PR, Martinez-Pomares L, Stacey M, Lin HH, Brown GD, Gordon S. Macrophage receptors and immune recognition. Annu Rev Immunol. 2005;23:901-44
38. Vieira Braga FA, Kar G, Berg M, et al. A cellular census of human lungs identifies novel cell states in health and in asthma. Nature Medicine. 2019 Jul;25(7):1153-1163
39. Makris S, Bajorek M, Culley FJ, Goritzka M, Johansson C. Alveolar Macrophages Can Control Respiratory Syncytial Virus Infection in the Absence of Type I Interferons. J Innate Immun. 2016;8(5):452-63.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

40. Lawless D, McKennan CG, Das SR, Junier T, Xu ZM, Anderson LJ, Gebretsadik T, Shilts MH, Larkin E, Rosas-Salazar C, Chappell JD, Fellay J, Hartert TV. Viral Genetic Determinants of Prolonged Respiratory Syncytial Virus Infection Among Infants in a Healthy Term Birth Cohort. J Infect Dis. 2023 May 12;227(10):1194-1202e

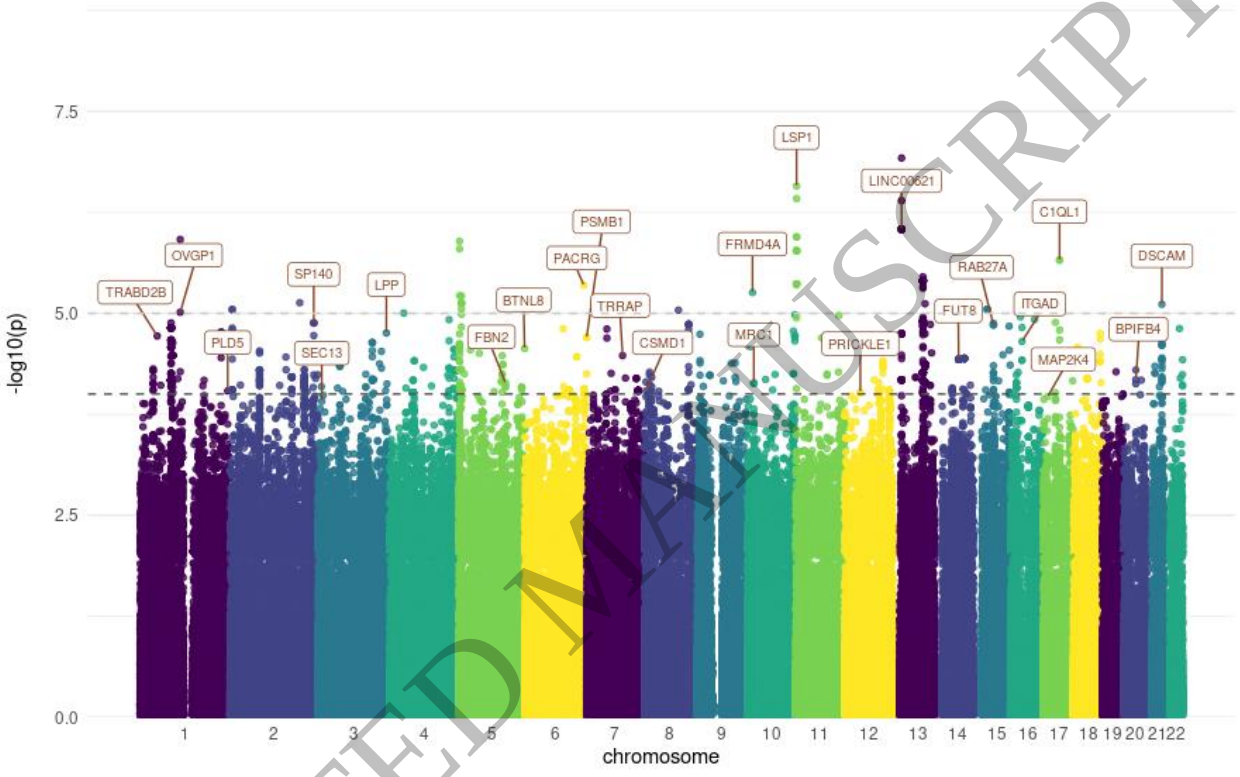
Table 1 A breakdown of the demographics for infants included in the GWAS of RSV severity. Following quality control, 251 samples were included in the GWAS. Patient characteristics are stratified by the age groups of infants aged between 2 weeks and 1 year of age. Infection severity was determined by ReSVinet, a combined clinical score. Fever is recorded as degrees above 37°C. For numeric variables, the median and quartile range, or the mean +/- a standard deviation are recorded.

Variable		< 3 Months (N = 122)	3- 6 Months (N = 57)	≥ 6 Months (N= 72)	Total (N = 251)
Age (days)	Median (IQR)	43.5 (25.0, 59.8)	140.0 (116.0, 156.0)	265.0 (210.0, 302.0)	91.0 (44.0, 187.0)
Sex (n)	Male	87 (71.3%)	35 (61.4%)	32 (44.4%)	154 (61.4%)
	Female	35 (28.7%)	22 (38.6%)	40 (55.6%)	97 (38.7%)
Site (n)	UK	42 (34.4%)	27 (47.4%)	36 (50.0%)	105 (41.8%)
	ND	71 (58.2%)	27 (47.4%)	35 (48.6%)	133 (53.0%)
	Spain	9 (7.4%)	3 (5.3%)	1 (1.4%)	13 (5.2%)
Pre-existing conditions (n)	Yes	16 (13.1%)	11 (19.3%)	15 (20.8%)	42 (16.7%)
Pre-term birth (n)	Yes	7 (5.7%)	2 (3.5%)	5 (6.9%)	14 (5.6%)
Hospitalised (n)	Yes	107 (88.4%)	36 (66.7%)	33 (46.5%)	176 (70.2%)
Respiratory Support	Yes	98 (80.3%)	28 (49.1%)	24 (33.0%)	150 (59.8%)
Fever (°C)	Mean (±SD)	0.27 ± 0.62	0.47 ± 0.80	0.60 ± 0.76	0.41 ± 0.72
ReSVinet Score	Median (IQR)	11 (7, 15)	7 (4, 12)	6 (3, 9)	8 (5, 12)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

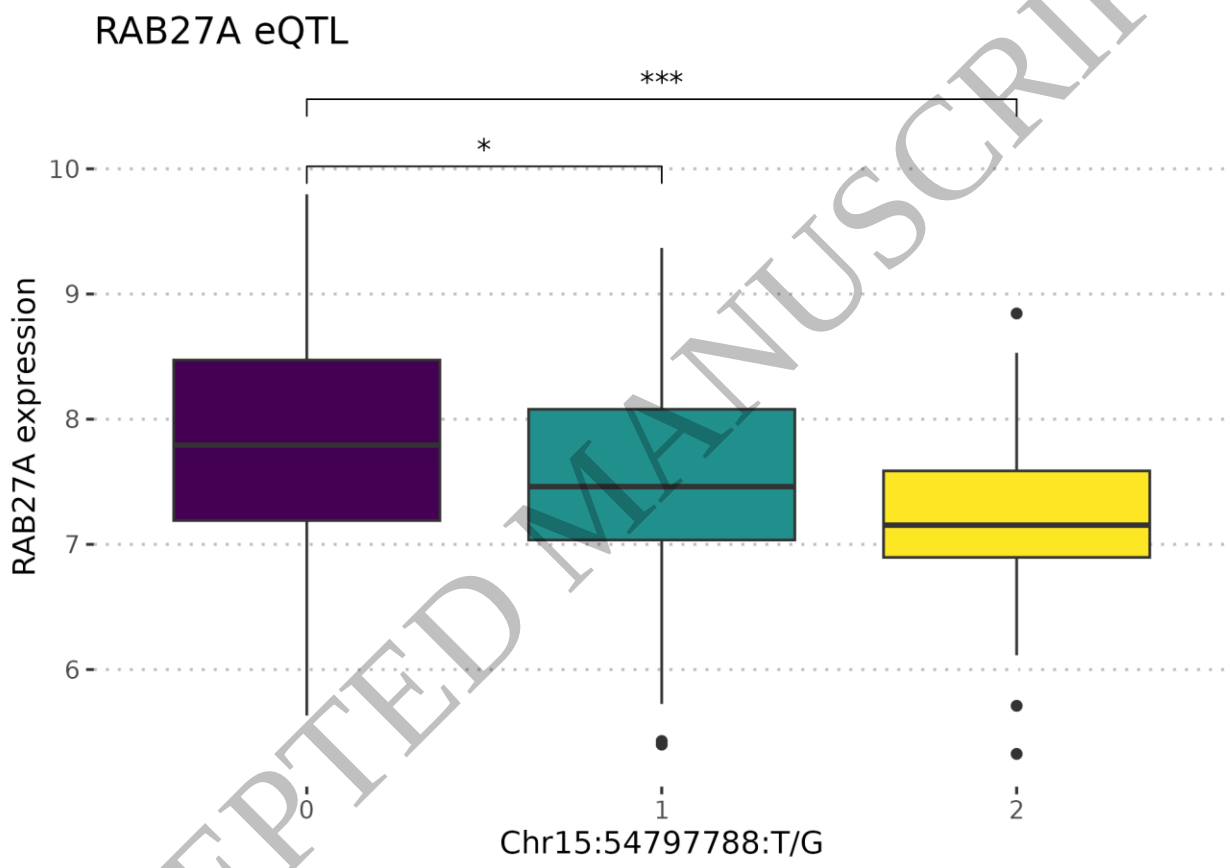
Fig 1) GWAS of RSV severity: Manhattan plot displaying the significance (−log10 unadjusted P value) of each variant in their association with RSV severity as determined by ReSVinet score. SNPs are plotted on the x-axis by chromosome and genomic coordinates. The dark dotted line indicates a suggestive threshold of $P < 1 \times 10^{-4}$. Highlighted are a subset of candidate variants and their overlapping gene loci.



ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024

Fig 2) SNP rs11071126 associates with reduced *RAB27A* gene expression: Presence of the SNP rs11071126 (Chr15:54797788:T:G) was associated with a significant reduction (FDR <0.05) in *RAB27A* gene expression using expression quantitative trait loci (eQTL) analysis. SNPs included in the eQTL analysis were filtered according to a suggestive significance threshold in the GWAS of RSV severity ($P < 1x10^{-4}$). Box plots show the distribution of *RAB27A* expression across individuals and are split according to genotype (0 = TT, 1 = TG, 2 = GG) (*** $P < 0.001$, * $P < 0.05$).



ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiae029/7589256 by Imperial College London Library user on 08 February 2024