Clinical Microbiology and Infection xxx (xxxx) xxx

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Original article

Risk of lower respiratory tract infections: a genome-wide association study with Mendelian randomization analysis in three independent European populations

Helene M. Flatby [1, 2, *], Humaira Rasheed [3, 4], Anuradha Ravi [1, 2], Laurent F. Thomas [3, 5, 6, 7], Kristin V. Liyanarachi [1, 8], Jan E. Afset [1, 5, 7, 9], Andrew T. DeWan [10, 1], Ben M. Brumpton [3, 4, 11], Kristian Hveem [3, 12], Bjørn O. Åsvold [13, 11, 3], Gunnar S. Simonsen [14, 15], Anne-Sofie Furberg [14, 16], Jan K. Damås [1, 17, 8], Erik Solligård [1, 2], Tormod Rogne [1, 10, 2]

[1] *Gemini Centre for Sepsis Research, Department of Circulation and Medical Imaging, NTNU, Norwegian University of Science and Technology, Trondheim, Norway*
[2] *Clinic of Anaesthesia and Intensive Care, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway*
[3] *K.G. Jebsen Centre for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway*
[4] *Clinic of Medicine, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway*
[5] *Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway*
[6] *BioCore—Bioinformatics Core Facility, Norwegian University of Science and Technology, Trondheim, Norway*
[7] *Clinic of Laboratory Medicine, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway*
[8] *Department of Infectious Diseases, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway*
[9] *Department of Medical Microbiology, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway*
[10] *Department of Chronic Disease Epidemiology and Center for Perinatal, Pediatric and Environmental Epidemiology, Yale School of Public Health, New Haven, CT, USA*
[11] *HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Levanger, Norway*
[12] *Department of Research, Innovation, and Education, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway*
[13] *Department of Endocrinology, Clinic of Medicine, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway*
[14] *Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway*
[15] *Research Group for Host–Microbe Interaction, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway*
[16] *Faculty of Health and Social Sciences, Molde University College, Molde, Norway*
[17] *Centre of Molecular Inflammation Research, Department of Clinical and Molecular Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway*

ARTICLE INFO

Article history:
Received 14 April 2021
Received in revised form
20 October 2021
Accepted 1 November 2021
Available online xxx

Editor: M. Leeflang

Keywords:
Body mass index
Genome-wide association study
Lower respiratory tract infection
Mendelian randomization

ABSTRACT

Objective: Lower respiratory tract infections (LRTIs) are a leading cause of morbidity and mortality worldwide. Few studies have previously investigated genetic susceptibility and potential risk factors for LRTI.

Methods: We used data from the UK Biobank, Trøndelag Health Study (HUNT), and FinnGen to conduct a genome-wide association study (GWAS). Cases were subjects hospitalized with LRTI, and controls were subjects with no such hospitalization. We conducted stratification and interaction analyses to evaluate whether the genetic effect of LRTI differed by sex or smoking. Mendelian randomization (MR) analyses were conducted to identify the unconfounded relationship between cardiometabolic risk factors and LRTI.

Results: A total of 25 320 cases and 575 294 controls were included. The 15q25.1 locus reached genome-wide significance in the meta-analysis (rs10519203: OR 0.94, p 3.87e-11). The protective effect of effect allele of rs10519203 was present among smokers (OR 0.90, 95%CI 0.87–0.92, p 1.38e-15) but not among never-smokers (OR 1.01, 95%CI 0.97–1.06, p 5.20e-01). In MR analyses, we found that increasing body mass

Abbreviations: BMI, body mass index; EAF, effect allele frequency; GWAS, genome-wide association study; HUNT, Trøndelag Health Study; ICD, International Classification of Diseases; IVW, inverse-variance weighted; LD, linkage disequilibrium; LDL-C, low-density lipoprotein cholesterol; LRTI, lower respiratory tract infection; OR, odds ratio; SBP, systolic blood pressure; SNP, single-nucleotide polymorphism; T2DM, type 2 diabetes mellitus.

* Corresponding author: Helene M. Flatby, Department of Circulation and Medical Imaging, NTNU, Prinsesse Kristinas gate 3, Akutten og Hjerte-lunge-senteret, 3. etg, Trondheim, 7491, Norway.
E-mail address: helene.flatby@ntnu.no (H.M. Flatby).

https://doi.org/10.1016/j.cmi.2021.11.004

Pneumonia
Respiratory infection
Smoking

index (OR 1.31, 95%CI 1.24−1.40, p 3.78e-18), lifetime smoking (OR 2.83, 95%CI 2.34−3.42, p 6.56e-27), and systolic blood pressure robustly increased the risk of LRTIs (OR 1.11, 95%CI 1.02−1.22, p 1.48e-02).
Conclusion: A region in 15q25.1 was strongly associated with LRTI susceptibility. Reduction in the prevalence of smoking, overweight, obesity, and hypertension may reduce the disease burden of LRTIs.
Helene M. Flatby, Clin Microbiol Infect 2021;▪:1

Introduction

Lower respiratory tract infections (LRTIs) are the sixth leading cause of mortality for all ages, with 2.38 million deaths globally each year [1]. There is significant variability in the host response to LRTI, even among individuals considered at low risk, which indicates a genetic component to the pathogenesis and prognosis [2]. Previous genetic studies have focused primarily on sepsis caused by pneumonia [3], childhood pneumonia [4], and viral pneumonia [5−7]. Genome-wide association studies (GWASs) on host susceptibility to bacterial LRTI in adults are scarce. Although a variety of pathogens can cause LRTIs, the majority are of bacterial origin [8].

Given that smoking is causally related to LRTI [9], it is relevant to estimate whether the effect of genetic variants associated with LRTI risk is modulated by smoking behaviour. Interaction analysis between smoking behaviour and genetic variants can be performed to assess how different risk estimates apply to different genetic predispositions.

A range of other risk factors for LRTI has also been identified in traditional epidemiological studies, such as obesity [10], diabetes, air pollution, and individuals of the paediatric and geriatric populations [2,8]. However, it can be challenging to quantify the causal effects of such studies. Mendelian randomization (MR) analyses use genetic variants—typically identified from GWASs—as instrumental variables to assess the causal association between an exposure and an outcome. Due to the fact that genetic variants are allocated to the offspring at random, MR analyses are at a reduced risk of residual confounding and reverse causation compared with traditional multivariable analyses [11]. MR studies have previously found cardiometabolic risk factors—such as body mass index (BMI) [12,13], smoking [14], and systolic blood pressure (SBP)—to affect the risk of LRTI [15].

Our aims were to identify genetic loci linked to LRTI through genome-wide association analyses in three large, independent cohorts of European ancestry—UK Biobank (UK), the Trøndelag Health Study (HUNT, Norway), and FinnGen (Finland)—and to evaluate whether these genetic effects were modulated by smoking behaviour. Finally, we conducted MR analyses by combining data from previously published GWASs with our genome-wide association analyses to identify the unconfounded associations between cardiometabolic risk factors and LRTI.

Methods

Setting

We used individual-level data from the UK Biobank and HUNT, and summary-level data from FinnGen. The cohorts are described in the Supplementary Material (Methods).

Phenotype

Cases and controls were determined based on ICD codes for bacterial LRTI retrieved from hospital records. The same ICD-9 and ICD-10 codes were used to identify cases in both the UK Biobank and HUNT (Supplementary Material Table S1). In our main analysis, cases were participants with at least one ICD-9/10 code as a primary or secondary diagnosis with LRTI. Controls were drawn from the same population as cases and were in all analyses defined as subjects with no registered hospitalization for LRTI (primary or secondary diagnosis).

For FinnGen, we used publicly available summary statistics from the genome-wide association analysis of the risk of bacterial pneumonia (Supplementary Material Table S1). Cases were defined as participants with at least one inpatient ICD-8/9/10 code as a primary or secondary diagnosis for bacterial pneumonia. Subjects without hospitalization with bacterial pneumonia served as controls [16].

A sensitivity analysis was performed for the UK Biobank and HUNT, where we included only the primary diagnosis based on ICD-9/10 codes.

Statistical analysis

Genome-wide association analyses

Genome-wide association analyses were conducted using scalable and accurate implementation of generalized mixed model (SAIGE; Supplementary Material: Methods) [17]. Summary statistics from the three cohorts were meta-analysed using fixed effect and weighted by effect size estimates and standard errors in METAL (version 2011-03-25) [18]. The meta-analysis included genomic control to account for residual population stratification. I^2 and Cochran's Q-test for heterogeneity were calculated to assess heterogeneity across the cohorts [18]. We considered association with $p < 5e\text{-}08$ to be genome-wide significant [19].

Stratified and interaction analyses

To evaluate whether the genetic risk of LRTI was modulated by smoking or sex, we conducted stratification and interaction analyses. We performed genome-wide association analyses stratified by smoking status (ever-smokers and never-smokers) and by sex.

In addition, we carried out an interaction analysis of smoking habits using the UK Biobank. Logistic regression models were used to calculate the association between smoking and LRTI by genotype group, presence or absence of the effect allele of the lead SNP, using never-smoker without the effect allele as the reference group. An interaction term was calculated between smoking status (never-smoker versus ever-smoker) and SNP genotype using the glm() function in R (version 3.6.1). Birth year, sex, genotype chip, and ancestry-informative principal components 1 to 6 were included as covariates.

Mendelian randomization

We performed two-sample MR analyses evaluating the unconfounded effect of the following five cardiometabolic traits on the risk of LRTI: BMI, type 2 diabetes mellitus (T2DM), low-density lipoprotein cholesterol (LDL-C), SBP, and lifetime smoking index. Genetic variants to be used as instruments for the exposures were obtained from publicly available GWAS summary data sources

Please cite this article as: Flatby HM et al., Risk of lower respiratory tract infections: a genome-wide association study with Mendelian randomization analysis in three independent European populations, Clinical Microbiology and Infection, https://doi.org/10.1016/j.cmi.2021.11.004

Table 1
Background characteristics at entry in the UK Biobank and HUNT study

	UK Biobank			HUNT		
	Cases (n = 12 807)	Controls (n = 395 091)	All (n = 407 898)	Cases (n = 7766)	Controls (n = 61 656)	All (n = 69 422)
Female sex[a]	5583 (43.6)	215 031 (54.4)	220 614 (54.1)	3638 (46.8)	33 146 (53.7)	36 784 (52.9)
Age at recruitment (years)[b]	62 (57–66)	58 (51–63)	58 (51–63)	64 (51–72)	43 (30–56)	45 (31–59)
Ever-smoker[a]	9086 (71.0)	237 031 (60.0)	246 117 (60.3)	5274 (82.0)	34 597 (64.5)	39 871 (66.4)
Diabetes (self-reported)[a]	184 (1.4)	2843 (0.7)	3027 (0.7)	718 (9.2)	2693 (4.4)	3411 (4.9)
Body mass index, kg/m^2 [c]	28.3 (5.6)	27.4 (4.7)	27.4 (4.8)	26.7 (5.8)	26.8 (4.9)	26.8 (5.0)
LDL cholesterol, mmol/L[c]	3.4 (0.9)	3.6 (0.9)	3.6 (0.9)	3.7 (1.2)	3.5 (1.1)	3.5 (1.1)
Systolic blood pressure, mmHg[c]	140.3 (19.3)	138.2 (18.6)	138.3 (18.6)	141.2 (27.5)	132.8 (22.3)	133.7 (23.1)

LDL, low-density lipoprotein.
[a] Data are presented as number (%).
[b] Data are presented as median (25th and 75th centile).
[c] Data are presented as mean (SD).

(Supplementary Material Table S2). The instrumental variables used as exposures had a p < 5e-08 for their respective trait and were independent (r^2 < 0.001 and kb 10 000) to ensure that the genetic variants were not prone to weak instrument bias. Our results from the genome-wide association analyses in each of the three cohorts and the meta-analysis were used to evaluate the effect of the genetic instruments on the outcome. The analyses were conducted using the TwoSampleMR (version 0.5.4) [20] and MendelianRandomization packages (version 0.5.0) [21] in R (version 4.0.0). The main analysis was performed using the inverse-variance weighted (IVW) method [20]. Three key assumptions should be met for an instrumental variable to be valid: the instrument should be robustly associated with the exposure, share no common cause with the outcome, and only affect the outcome through the risk factor [11]. Sensitivity analyses were performed using MR Egger, weighted median, weighted mode [20], and the contamination mixture method [21] to check whether these assumptions were met.

Phenome-wide association

Phenome-wide association analyses were performed through the Open Targets platform (genetics.opentargets.org, accessed 14th September 2020) for the lead variant that reached genome-wide significance in the meta-analysis [22].

Ethical approval

The Regional Committee for Medical Research, Health Region IV, in Norway (REK) has approved the HUNT study, and this project is regulated in conjunction with the Norwegian Social Science Data Services (NSD). The UK Biobank study has ethical approval from the North West Multi-Centre Research Ethics Committee (MREC). Approval for individual projects is covered by the Research Tissue Bank (RTB). The FinnGen Biobank was evaluated and approved by the Coordinating Ethics Committee of the Helsinki and Uusimaa Hospital District.

Results

Cohort demographics for UK Biobank, HUNT, and FinnGen

The baseline characteristics for cases and controls in the UK Biobank and HUNT are shown in Table 1. The case prevalence of LRTI was 12 807 (3.7 %), 7766 (11.2%), and 4747 (3.8%), in UK Biobank, HUNT, and FinnGen, respectively, with a male predominance among cases in all three cohorts (56.4%, 53.2%, and 53.9% respectively). Mean age at the first event was lower for cases in FinnGen (57 years) compared with UK Biobank and HUNT (62 and 64 years, respectively). In the UK Biobank and HUNT, cases at baseline were older than controls, had higher SBP, and were more likely to be ever-smokers and self-report having diabetes (Table 1). This information was unavailable for cases and controls in FinnGen.

Genome-wide association analyses

Evaluating each cohort separately, no region reached genome-wide significance (Manhattan plots presented in Supplementary Material Figs. 7–9).

The meta-analysis of the three cohorts totalled 25 320 cases and 575 294 controls. One locus reached genome-wide significance: lead SNP rs10519203 in hydroxylysine kinase (*HYKK*) on chromosome 15 located in the region 15q25.1 (Fig. 1). The genome-wide-significant SNPs in the 15q25.1 region were mapped to *IREB2*, *HYKK*, *CHRNA5*, *CHRNA3*, and *CHRNB4* genes, where many SNPs were in high linkage disequilibrium (LD) with the lead SNP (Fig. 2). The effect allele (A) of rs10519203 (effect allele frequency (EAF) 0.66 in each of the three cohorts) was associated with LRTI with an OR of 0.94 (95%CI 0.92–0.95, p 3.86e-11) in the meta-analysis, with no heterogeneity between the individual cohorts (I^2 0%) (Table 2; Manhattan plots presented in Supplementary Material Figs. 7–9).

The genomic inflation factor in the meta-analysis (λ 1.02), UK Biobank (λ 1.01), HUNT (λ 0.99) and FinnGen (λ 1.01) did not indicate inflation of the results (Supplementary Material Figs. 10–13).

The sensitivity analysis using only primary diagnosis yielded results similar to those of the main analysis (Supplementary Material Table S3).

Stratified and interaction analyses

We observed the same genetic associations with LRTI when we carried out genome-wide analyses stratified by sex (Supplementary Material Table S4). Stratified by smoking status, we observed that the effect of rs10519203 was present among ever-smokers (OR 0.90, 95%CI 0.87–0.92, p 1.38e-15) but not among never-smokers (OR 1.01, 95%CI 0.97–1.06, p 5.20e-01). Furthermore, in the analysis limited to never-smokers, we identified a locus in *NR4A2* (OR 1.68, 95%CI 1.40–2.02, p 4.22e-08) that reached genome-wide significance that was not identified in the main analysis (Supplementary Material Table S4). In the interaction analysis we observed a non-additive interaction between smoking and the protective allele A of rs10519203 (OR 0.79, 95%CI 0.69–0.89, p 1.00e-04). Through the association between smoking and LRTI by genotype group, we found that the protective effect of the A-allele was present only for ever-smokers (Table 3).

Please cite this article as: Flatby HM et al., Risk of lower respiratory tract infections: a genome-wide association study with Mendelian randomization analysis in three independent European populations, Clinical Microbiology and Infection, https://doi.org/10.1016/j.cmi.2021.11.004

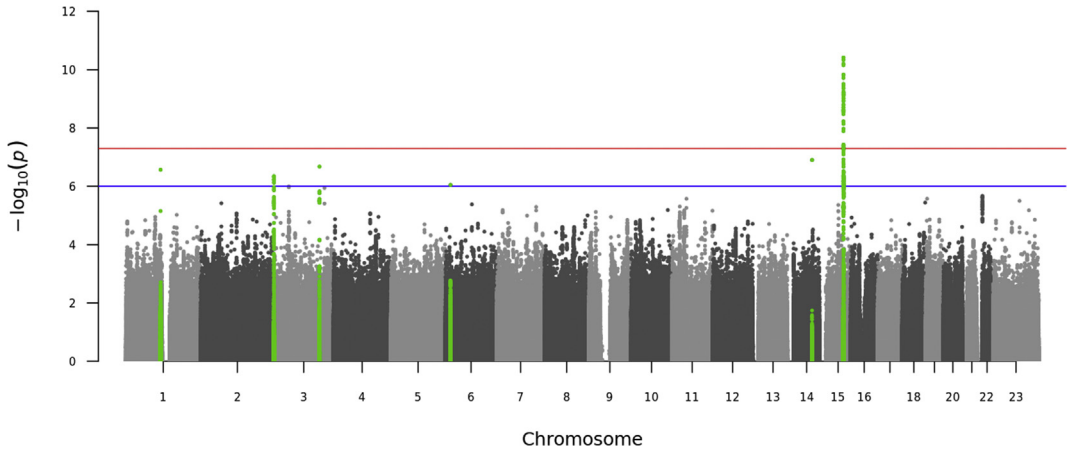

Fig. 1. Manhattan plot of results for the meta-analysis on primary or secondary diagnosis of lower respiratory tract infection (LRTI). Manhattan plot showing results for the fixed-effect meta-analysis on primary or secondary diagnosis of LRTI. The *x*-axis shows the genomic position (chromosomes 1–23), and the *y* axis represents the negative logarithm (base 10) of the variant p value. The blue line indicates genome-wide suggestive associations (p value <1e-06), and the red line indicates genome-wide significant associations (p value <5e-08). Genome-wide suggestive loci (±500 kb of lead variant) are highlighted in green.

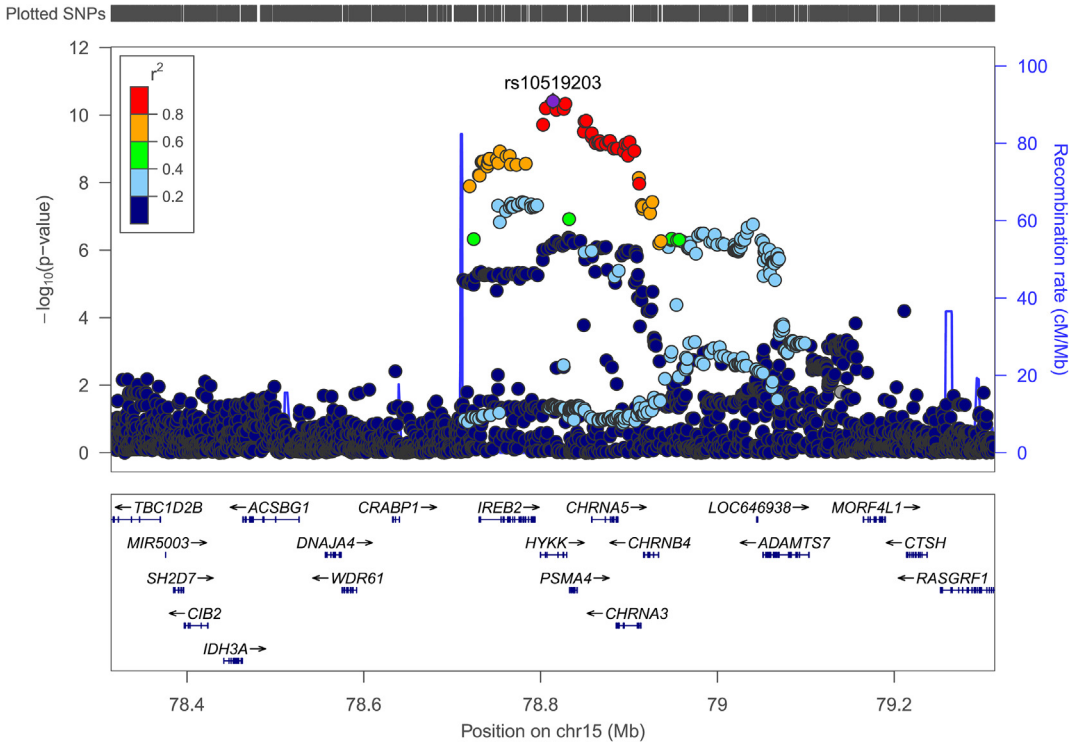

Fig. 2. Regional association plot for chromosome 15 locus (rs10519203) from the meta-analysis. Locus zoom plot presenting the results of the fixed-effects meta-analysis in chr15q25.1 region on primary or secondary diagnosis of lower respiratory tract infection (LRTI). The *y* axis represents the negative logarithm (base 10) of the variant p value, and the *x*-axis represents the position on the chromosome 15, with the names and location of genes and nearest genes shown at the bottom. The variant rs10519203 with the lowest p value in the region is marked by a purple diamond. The colours of the other variants linkage disequilibrium with the lead SNP estimated by r^2 (*red* $r^2 \geq 0.8$, *gold* $0.6 \leq r^2 < 0.8$, *green* $0.4 \leq r^2 < 0.6$, *cyan* $0.2 \leq r^2 < 0.4$, *blue* $r^2 < 0.2$, *grey* r^2 unknown). The European population from the 1000 Genomes Project [30], November 2014 release, was used as a reference, on genome build hg19.

Mendelian randomization analyses

Using a two-sample MR framework, we found that increasing BMI, lifetime smoking, and SBP were associated with an increased risk of developing LRTI. The OR for developing LRTIs in the meta-analysis was 1.31 (95%CI 1.24–1.40, p 3.78e-18) and 2.83 (95%CI 2.34–3.42, p 6.56e-27) for BMI and lifetime smoking respectively, with comparable results in the three individual cohorts (Fig. 3). High SBP and genetic predisposition to T2DM were found to increase the risk of LRTIs, where the ORs for developing LRTI were 1.11 (95%CI 1.02–1.22, p 1.48e-02) and 1.04 (95%CI 1.02–1.07, p 1.79e-03), respectively. For SBP, comparable effect estimates were present in the UK Biobank and HUNT, but no effect was present in FinnGen. For T2DM, similar effects were observed in the three cohorts, with less precision for HUNT and FinnGen. LDL-C was not found to increase the risk of LRTI.

There were clear signs of pleiotropic effects in the association between T2DM and LRTI, which suggests that, overall, there was no

Please cite this article as: Flatby HM et al., Risk of lower respiratory tract infections: a genome-wide association study with Mendelian randomization analysis in three independent European populations, Clinical Microbiology and Infection, https://doi.org/10.1016/j.cmi.2021.11.004

Table 2
Genetic variants with p value <1e−06 for the fixed-effect meta-analysis on primary or secondary diagnosis of lower respiratory tract infection

rsID[a]	Chr	Pos[b]	Category[c]	Closest gene	EA/ OA	Meta-analysis				UK Biobank				HUNT				FinnGen			
						EAF	OR (95%CI)	p value	I^2	I^2 p value	EAF	OR (95%CI)	p value	*INFO*	EAF	OR (95%CI)	p value	R^2	EAF	OR (95%CI)	p value
rs12569146	1	115794671	intron	NGF	C/T	0.05	0.88 (0.84 −0.93)	2.72e-07	0	5.21e-01	0.05	0.91 (0.84 −0.99)	2.35e-02	0.97	0.03	0.85 (0.77 −0.94)	2.03e-03	0.97	0.09	0.87 (0.81 −0.94)	2.66e-04
rs112752491	3	280850	intron	CHL1	G/A	0.08	1.09 (1.06 −1.13)	4.49e-07	85.8	6.12e-02	0.08	1.09 (1.04 −1.14)	3.18e-04	0.97	0.08	1.16 (1.08 −1.24)	2.09e-05	0.97	0.06	1.02 (0.93 −1.11)	6.95e-01
rs9823831	3	151584657	intergenic	SUCNR1	C/T	0.35	0.95 (0.93 −0.97)	2.11e-07	39.4	1.92e-01	0.35	0.94 (0.91 −0.96)	2.08e-06	0.98	0.34	0.98 (0.94 −1.02)	2.70e-01	0.97	0.36	0.94 (0.90 −0.99)	8.77e-03
rs114244623	6	17356806	intergenic	CAP2	A/G	0.02	1.21 (1.12 −1.31)	8.99e-07	13.3	3.16e-01	0.02	1.15 (1.04 −1.28)	7.34e-03	0.97	0.02	1.29 (1.13 −1.48)	1.83e-04	0.96	0.01	1.32 (1.05 −1.68)	2.01e-02
rs150325821	14	81842837	intron	STON2	T/C	0.01	1.25 (1.15 −1.36)	1.25e-07	41.9	1.79e-01	0.01	1.19 (1.07 −1.32)	1.29e-03	0.90	0.01	1.43 (1.21 −1.68)	1.80e-05	0.98	0.01	1.22 (0.97 −1.54)	9.48e-02
rs10519203	**15**	**78814046**	**intron**	**HYKK**	**A/G**	**0.66**	**0.94 (0.92 −0.95)**	**3.86e-11**	**0**	**6.17e-01**	**0.66**	**0.93 (0.91 −0.96)**	**4.16e-07**	**1.00**	**0.66**	**0.93 (0.89 −0.96)**	**1.20e-04**	**0.98**	**0.66**	**0.95 (0.91 −1.00)**	**3.16e-02**
rs34138960	15	78831668	Upstream-gene	PSMA4	G/A	0.23	0.94 (0.92 −0.97)	4.23e-07	0	4.24e-01	0.23	0.93 (0.89 −0.94)	8.17e-06	0.99	0.22	0.95 (0.91 −0.99)	1.74e-02	0.99	0.27	0.97 (0.92 −1.01)	1.56e-01

Chr, chromosome; CI, confidence interval; EA, effect allele; EAF, effect allele frequency; I^2, heterogeneity; I^2 p value, heterogeneity *P* value.
OA, other allele; OR, odds ratio; Pos, chromosome position; R^2, imputation score from Minimac3; SNP, single-nucleotide polymorphism.
INFO, imputation score from IMPUTE2.
[a] Boldface indicates SNPs with genome-wide significant evidence of association in the meta-analysis (*P* value <5e-08).
[b] Position of each SNP is given along the chromosome Build 37.
[c] Predicted function.

Table 3
Interaction analysis in the UK Biobank for the effect between rs10519203 and smoking status

Alleles	Smoking	OR (95%CI)	p value
GG	No	Reference	Reference
GG	Yes	1.93 (1.72−2.17)	1.39e-28
AG or AA	No	1.05 (0.94−1.17)	3.90e-01
AG or AA	Yes	1.59 (1.44−1.77)	1.25e-18

CI, confidence interval; OR, odds ratio.

robust association between the two in our data (Supplementary Material Table S5). The MR sensitivity analyses otherwise supported the findings from the IVW analyses.

Phenome-wide association analysis

Phenome-wide association analysis of rs10519203 identified 360 diseases and traits in the UK Biobank and GWAS Catalogue. After Bonferroni correction ($p < 1.39e-04$), 64 traits were significantly correlated with rs10519203 (Supplementary Material Fig. S14), and the full results are available as Supplementary Data 1. We identified that the presence of the effect allele (A) of rs10519203 had a protective effect on smoking behaviour, such as cigarettes smoked per day (lowest *p* 3.00e-286, OR 0.91) and different respiratory diseases such as lung cancer (lowest *p* 2.39e-16, OR 0.93) and emphysema (lowest *p* 3.21e-13, OR 0.77).

Discussion

By use of three large, independent, European cohorts, we identified the 15q25.1 region as being strongly associated with LRTI susceptibility. The effect of the lead variant, rs10519203, was present among smokers but not among never-smokers. In the Mendelian randomization analyses, we found that increasing BMI, lifetime smoking, and SBP were associated with an increased risk of LRTI.

The genetic variant with the strongest association with LRTI was the intron variant rs10519203 in *HYKK*, a protein-coding gene involved in lysine catabolism. However, due to extensive LD in this 15q25.1 region, it is possible that the causative allele(s) reside within any of the other genes in the region. A recent pre-print also identified the 15q25.1 region as being associated with viral and bacterial pneumonia [23], but our analyses yielded a stronger signal due to the larger sample size. Previously published studies have linked variants in the 15q25.1 region to chronic obstructive pulmonary disease [24], emphysema [25], lung cancer [26], smoking behaviour [27], and nicotine, alcohol, and cocaine dependence [28].

Using MR analyses, we found an unconfounded association between BMI, lifetime smoking, and SBP with the risk of developing LRTI. These results align with previously published MR studies [12−15].

We observed no sex-specific effect in the genome-wide analyses stratified by sex. In the analysis stratified by smoking status, however, we observed that the protective effect of the effect allele for rs10519203 was present only among smokers. This aligns well with our phenome-wide association analysis, where we found that the effect allele of rs10519203 had a protective effect for traits related to smoking behaviour and several other respiratory diseases. We also identified a locus associated with the risk of LRTI that was present among never-smokers but not among ever-smokers. Variants in this locus on chromosome 2 have been linked to reduced forced vital capacity [29].

A key strength of our study is the large number of cases and controls included, which increases the statistical power and

Please cite this article as: Flatby HM et al., Risk of lower respiratory tract infections: a genome-wide association study with Mendelian randomization analysis in three independent European populations, Clinical Microbiology and Infection, https://doi.org/10.1016/j.cmi.2021.11.004

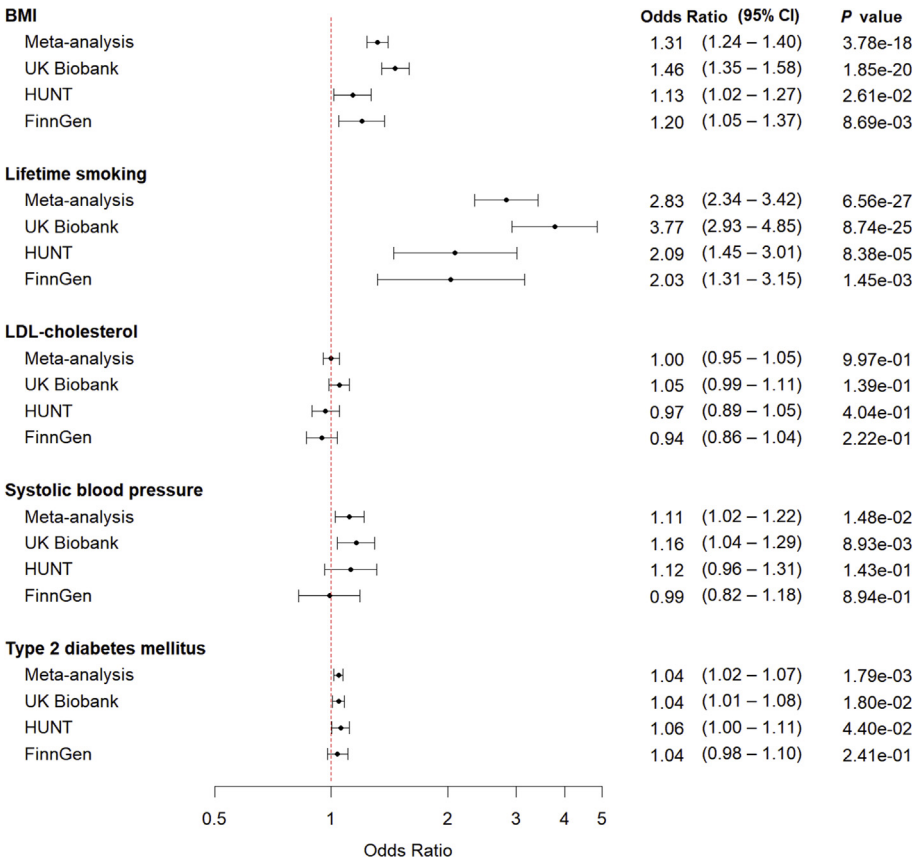

Fig. 3. Mendelian randomization analyses of cardiometabolic risk factors on risk of lower respiratory tract infections (LRTIs). Forest plot of the two-sample inverse-variance weighted Mendelian randomization analyses of cardiometabolic risk factors identified as genetically correlated with LRTIs. Each risk factor was evaluated separately for the meta-analysis, UK Biobank, HUNT, FinnGen, and the corresponding risk factors. The x-axis represents the results expressed as per standard deviation increase in genetically proxied levels of the exposure for continuous traits (body mass index (BMI), lifetime smoking, low-density lipoprotein (LDL) cholesterol, systolic blood pressure) and as per unit increase in the log odds ratio for genetically proxied type 2 diabetes liability.

allowed us to carry out analyses stratified by sex and smoking status. Another strength is the use of a narrow phenotype definition for bacterial LRTIs, which may have different underlying pathophysiology compared with viral LRTIs. The phenotype was based on hospital record linkage, and there is a risk of non-differential misclassification. We sought to limit the random error introduced by this misclassification by using a large sample. Ideally, the GWAS results used in the meta-analyses should all be conducted with the same ascertainment criteria on comparable populations and following the same study design. In the present study, different genotyping platforms and imputation programmes were used for the three cohorts, and the phenotype definition was different in FinnGen compared with UK Biobank and HUNT, which may increase the heterogeneity between the cohorts. However, for the lead genetic loci, we found very little variety between the included cohorts. To avoid bias due to population stratification, we evaluated only subjects of European ancestry, and it is unclear whether our findings apply to other ancestry groups, highlighting the need for research on other ancestry groups. Finally, the genetic instruments used in the MR analyses may have had pleiotropic effects, which could produce spurious findings. To limit this risk, we conducted several sensitivity analyses which revealed that the observed positive association between T2DM and LRTI in the main MR analysis might be due to bias, while the other findings were robust.

In conclusion, we identified a region in 15q25.1 to be strongly associated with LRTI susceptibility. Using an MR framework, we found that increased BMI, lifetime smoking, and SBP were associated with a greater risk of LRTI. Reduction in the prevalence of smoking, overweight, obesity, and hypertension may reduce the disease burden of LRTIs.

Author contributions

All authors made substantial contributions to the interpretation of the data and critically revised the manuscript. All authors have approved the submitted version and are personally accountable for their own contributions.

Transparency declaration

This work was supported by Samarbeidsorganet Helse Midt-Norge and Norwegian University of Science and Technology, NTNU. The funders had no role in study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit the article for publication.

Web resources

We used the following web-based resources: LD Hub (http://ldsc.broadinstitute.org), Locus Zoom (http://locuszoom.org), and Open Targets platform (https://genetics.opentargets.org).

Please cite this article as: Flatby HM et al., Risk of lower respiratory tract infections: a genome-wide association study with Mendelian randomization analysis in three independent European populations, Clinical Microbiology and Infection, https://doi.org/10.1016/j.cmi.2021.11.004

Acknowledgements

This research was conducted using data from the UK Biobank resource under the application number 40135. The HUNT Study is a collaboration between the HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. We want to acknowledge the participants and investigators of the FinnGen study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cmi.2021.11.004.

References

[1] Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age–sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017;390:1151–210.
[2] Dela Cruz CS, Wunderink RG, Christiani DC, Cormier SA, Crothers K, Doerschuk CM, et al. Future research directions in pneumonia. NHLBI working group report. Am J Respir Crit Care Med 2018;198:256–63.
[3] Rautanen A, Mills TC, Gordon AC, Hutton P, Steffens M, Nuamah R, et al. Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study. Lancet Respir Med 2015;3:53–60.
[4] Hayden LP, Cho MH, McDonald MN, Crapo JD, Beaty TH, Silverman EK, et al. Susceptibility to childhood pneumonia: a genome-wide analysis. Am J Respir Cell Mol Biol 2017;56:20–8.
[5] Zuniga J, Buendia-Roldan I, Zhao Y, Jimenez L, Torres D, Romo J, et al. Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. Eur Respir J 2012;39:604–10.
[6] 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;8:599.
[7] Chen H-H, Shaw DM, Petty LE, Graff M, Bohlender RJ, Polikowsky HG, et al. Host genetic effects in pneumonia. Am J Hum Genet 2021;108:194–201.
[8] Troeger C, Blacker B, Khalil IA, Rao PC, Cao J, Zimsen SR, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Infect Dis 2018;18:1191–210.
[9] Nuorti JP, Butler JC, Farley MM, Harrison LH, McGeer A, Kolczak MS, et al. Cigarette smoking and invasive pneumococcal disease. N Engl J Med 2000;342:681–9.
[10] Nie W, Zhang Y, Jee SH, Jung KJ, Li B, Xiu Q. Obesity survival paradox in pneumonia: a meta-analysis. BMC Med 2014;12:61.
[11] Davies NM, Holmes MV, Smith GD. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 2018;362:k601.
[12] Winter-Jensen M, Afzal S, Jess T, Nordestgaard BG, Allin KH. Body mass index and risk of infections: a Mendelian randomization study of 101,447 individuals. Eur J Epidemiol 2020;35:347–54.
[13] Butler-Laporte G, Harroud A, Forgetta V, Richards JB. Elevated body mass index is associated with an increased risk of infectious disease admissions and mortality: a mendelian randomization study. Clin Microbiol Infect 2020;27:710–6.
[14] King C, Mulugeta A, Nabi F, Walton R, Zhou A, Hyppönen E. Mendelian randomization case-control PheWAS in UK Biobank shows evidence of causality for smoking intensity in 28 distinct clinical conditions. EClinicalMedicine 2020;26:100488.
[15] Zekavat SM, Honigberg M, Pirruccello JP, Kohli P, Karlson EW, Newton-Cheh C, et al. Elevated blood pressure increases pneumonia risk: epidemiological association and Mendelian randomization in the UK Biobank. Med 2020;2:137–148.e4.
[16] FinnGen. FinnGen documentation of R3 release. 2020. Available from: https://finngen.gitbook.io/documentation/v/r3/. [Accessed 2 June 2020].
[17] Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case–control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 2018;50:1335–41.
[18] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–1.
[19] Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. Eur J Hum Genet 2016;24:1202–5.
[20] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife 2018;7:e34408.
[21] Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol 2017;46:1734–9.
[22] Mountjoy E, Schmidt EM, Carmona M, Peat G, Miranda A, Fumis L, et al. Open Targets Genetics: an open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. bioRxiv 2020. https://doi.org/10.1101/2020.09.16.299271.
[23] Campos A, Kho PF, Vazquez-Prada KX, Garcia-Marin LM, Martin NG, Cuellar-Partida G, et al. Genetic susceptibility to pneumonia: a GWAS meta-analysis between UK Biobank and FinnGen. medRxiv 2020. https://doi.org/10.1101/2020.06.22.20103556.
[24] Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 2009;5:e1000421.
[25] Lambrechts D, Buysschaert I, Zanen P, Coolen J, Lays N, Cuppens H, et al. The 15q24/25 susceptibility variant for lung cancer and chronic obstructive pulmonary disease is associated with emphysema. Am J Respir Crit Care Med 2010;181:486–93.
[26] Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 2008;452:633–7.
[27] Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet 2010;42:436–40.
[28] Wang JC, Grucza R, Cruchaga C, Hinrichs AL, Bertelsen S, Budde JP, et al. Genetic variation in the CHRNA5 gene affects mRNA levels and is associated with risk for alcohol dependence. Mol Psychiatr 2009;14:501–10.
[29] Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat Genet 2019;51:481–93.
[30] Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature 2015;526:68–74.

Please cite this article as: Flatby HM et al., Risk of lower respiratory tract infections: a genome-wide association study with Mendelian randomization analysis in three independent European populations, Clinical Microbiology and Infection, https://doi.org/10.1016/j.cmi.2021.11.004