Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Human genetic factors associated with pneumonia risk, a cue for COVID-19 susceptibility

Debleena Guin, Saroj Yadav, Priyanka Singh, Pooja Singh, Sarita Thakran, Samiksha Kukal, Neha Kanojia, Priyanka Rani Paul, Bijay Pattnaik, Viren Sardana, Sandeep Grover, Yasha Hasija, Luciano Saso, Anurag Agrawal, Ritushree Kukreti

A R T I C L E   I N F O

Keywords: Pneumonia, Community acquired pneumonia, Nosocomial pneumonia, CYP1A1, Genetic variants, Meta-analysis, COVID-19

A B S T R A C T

Pneumonia, an acute respiratory tract infection, is one of the major causes of mortality worldwide. Depending on the site of acquisition, pneumonia can be community acquired pneumonia (CAP) or nosocomial pneumonia (NP). The risk of pneumonia is partially driven by host genetics. CYP1A1 is a widely studied pulmonary CYP family gene primarily expressed in peripheral airway epithelium. The CYP1A1 genetic variants, included in this study, alter the gene activity and are known to contribute in lung inflammation, which may cause pneumonia pathogenesis. In this study, we performed a meta-analysis to establish the possible contribution of CYP1A1 gene, and its three variants (rs2606345, rs1048943 and rs4646903) towards the genetic etiology of pneumonia risk. Using PRISMA guidelines, we systematically reviewed and meta-analysed case-control studies, evaluating risk of pneumonia in patients carrying the risk alleles of CYP1A1 variants. Heterogeneity across the studies was evaluated using I² statistics. Based on heterogeneity, a random-effect (using maximum likelihood) or fixed-effect (using inverse variance) model was applied to estimate the effect size. Pooled odds ratio (OR) was calculated to estimate the overall effect of the risk allele association with pneumonia susceptibility. Egger’s regression test and funnel plot were used to assess publication bias. Subgroup analysis was performed based on pneumonia type (CAP and NP), population, and age group. A total of ten articles were identified as eligible studies, which included 3049 cases and 2249 healthy controls. The meta-analysis findings revealed CYP1A1 variants, rs2606345 (T vs G; OR = 1.12 (0.75–1.50); p = 0.02; I² = 84.89%), and rs1048943 (G vs T; OR = 1.19 (0.76–1.61); p = 0.02; I² = 0.00%) as risk markers whereas rs4646903 showed no statistical significance for susceptibility to pneumonia. On subgroup analysis, both the genetic variants showed significant association with CAP but not with NP. We additionally performed a spatial analysis to identify the key factors possibly explaining the variability across countries in the prevalence of the coronavirus disease 2019 (COVID-19), a viral pneumonia. We observed a significant association between the risk allele of rs2606345 and rs1048943, with a higher COVID-19 prevalence worldwide, providing us important links in understanding the variability in COVID-19 prevalence.

https://doi.org/10.1016/j.meegid.2022.105299

Received 10 January 2022; Received in revised form 30 March 2022; Accepted 2 May 2022

Available online 8 May 2022

1567-1348/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

Pneumonia is an acute inflammatory condition of the lungs usually caused by bacterial, viral or fungal infection (Mackenzie, 2016). According to the site of acquisition, pneumonia is classified as community acquired (CAP) or nosocomial (NP) (Mackenzie, 2016). The global burden of disease study (2015) stated that lower respiratory infections like pneumonia are the third most common cause of death globally (WHO, 2022). Though the biology of the infecting microbe is important, host genetic plays a crucial role in the pathogenesis of pneumonia (Dela Cruz et al., 2018). Pneumonia occurs when pathogens enter the alveoli, infect, multiply and encourage a host immune response. These responses cause inflammation of the lung tissues, marking the pathogenesis of pneumonia (Jain et al., 2021). Host genetic factors that participate in these processes starting from pathogen entry, infection, inflammation, and resolution can all be considered as good candidates in genetic association studies of pneumonia and its complications (Cooke and Hill, 2001; Kumar et al., 2014). The difference in epidemiology, pathogenesis, host, microbiology, common causal organisms, and pathophysiology between CAP and NP depends on the mode of acquisition of pneumonia infection, on host risk factors and other environmental changes (Herold and Sailer, 2004; Torres et al., 2021). Thus, understanding the functional impact of genetic determinants of susceptibility to pneumonia, both CAP and NP, independently is crucial for determining the mechanisms behind pneumonia pathogenesis.

CYP1A1 is a critical enzyme mediating the metabolism of a broad spectrum of xenobiotics and endobiotics (9). There are several reports highlighting the functionally relevant genetic variants of CYP1A1 to play a clinically important role in several disease phenotypes. A number of studies have investigated the genetic association as well as gene interaction of pneumonia risk with monooxygenase enzyme group, cytochrome P450 (CYP). In a study, a group of researchers identified CYP1A1 gene as a critical regulator of inflammatory responses and phagocytosis in sepsis through signalling pathways that may be promising targets for treating inflammatory diseases (Tian et al., 2020). A study by Fang et al. (2016) demonstrated lower CYP1A1 expression in pigs infected with Mycoplasma hyopneumoniae (M. hyopneumoniae). They further extended their efforts by studying this gene in pulmonary alveolar macrophages (PAM) cell lines suggest CYP1A1 suppresses inflammatory response caused by pneumonia infection (Fang et al., 2016). Interestingly, few other studies indicated the role of CYP1A1 genetic polymorphisms in infectious diseases and consequently establishing its role in inflammatory responses. Previously, it was identified that genetic variants of some host genes (CYP1A1, ACE and IL-6) are associated with the diversity in response to CAP (Salnikova et al., 2014; Zhao et al., 2017). The selection of this gene was established based on its role in physiological and pathological processes during pneumonia infection, particularly in the immune and inflammatory responses (Zhao et al., 2017). From the systematic literature search performed for association of CYP1A1 genetic variants and pneumonia, the most widely reported CYP1A1 single nucleotide polymorphisms (SNPs) obtained were rs2606345, rs1048943 and rs4646903. These SNPs had functional consequences which may ultimately be involved with a disease phenotype like pneumonia. The SNP rs2606345 (C > A), is located in the first intron of the gene, has a functional role of lower gene expression in the presence of allele A (Rotunno et al., 2009; Talwar et al., 2017). Another SNP, rs1048943 (T > A, C, G), resulted in a missense amino acid substitution characterized by the substrate-specific increased activity for minor allele G (Salnikova et al., 2013c). The presence of minor allele ‘G’ of the 3’ untranslated region (UTR) SNP, rs4646903, shows an increased inducibility of CYP1A1 gene expression (Meletiadis et al., 2006; Salnikova et al., 2013c). Thus we can suggest that genetically determined alteration of CYP1A1 expression could contribute to lung inflammation pathogenesis. While there are several evidence of association between CYP1A1 polymorphisms with risk of pneumonia, there are studies which show conflicting results as well (Muñoz et al., 2012; Smith et al., 2001).

In this study we used a meta-analysis approach: 1) to investigate the impact of CYP1A1 risk allele and the risk of pneumonia (including both CAP and NP). This may increase the odds of the incident pneumonia; 2) to determine whether any association between CYP1A1 and pneumonia is generalizable to the coronavirus disease 2019 (COVID-19) as the ongoing COVID-19 pandemic and its consequent prevalence has rarely been examined through the lens of pneumonia. This genetic predisposition with pneumonia may help us in understanding the genetic etiology of COVID-19 infection and its prevalence.

2. Materials and methods

This meta-analysis was conducted as per the recommendations of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines (Liberati et al., 2009) following the PICOS (Population, intervention, comparison, outcome and study design) strategy. The PICOS outline (Population: pneumonia patients; Intervention/Exposure: individuals carrying the risk allele of CYP1A1 variants, rs2606345, rs1048943 and rs4646903; Comparisons: risk allele carriers in cases and controls vs wild type allele carriers in cases and controls; Outcomes: pneumonia susceptibility in patients carrying the risk allele of CYP1A1 variants; and Study design: case-control studies). The selected studies were those in which the relationship between CYP1A1 gene polymorphisms and risk of pneumonia disease has been evaluated. Bibliographic databases like MEDLINE (PubMed), Web of Science, and Science Direct and worldwidescience.org were searched for all articles published till January 13, 2021. The keywords used to identify relevant studies were “CYP1A1”, “Pneumonia”, “genetic variant”, and “single nucleotide polymorphisms” using AND/OR Boolean operators. Cross references of each study retrieved were also examined for inclusion in case they discuss the effect of CYP1A1 genetic variant and its risk in pneumonia.

Two investigators (DG and SY) independently reviewed each study for its inclusion in the meta-analysis. The inclusion criteria were: (1) studies conducted on human population only, (2) included the effect of CYP1A1 genetic variant with available numeric data, (3) adopted a case-control study design (4) provided a detailed assay method. Studies (1) any other type of lung inflammation apart from pneumonia or pneumonia as a consequence of any exposure or pneumonia existing with comorbid conditions, (2) no defined diagnostic criteria for pneumonia, and (3) genotypic data not in accordance with Hardy Weinberg equilibrium were excluded.

Allele frequency data for each case and control were extracted into contingency tables to calculate the odds of pneumonia in patients carrying the risk allele of the associated variants. In case of missing allele frequency data, the corresponding odds ratio (OR) and p value were calculated from genotypic data given. The references of the retrieved articles were manually screened to identify additional studies. In case of studies where genotypic data is given allelic data is calculated to maintain a consensus across studies. The included studies and their characteristics like first author, year of publication, population, disease, genetic variant, odds ratio, genotyping method, risk allele, sample size (cases and control) and male female distribution and their quality assessment score were tabulated. All the included articles described some variant of pneumonia infection, one study discussing Mycoplasma pneumoniae infection (Zhao et al., 2017), CAP (Moroz et al., 2011; Salnikova et al., 2013a; Salnikova et al., 2013c; Salnikova et al., 2010; Smelaya et al., 2011), NP (Salnikova et al., 2014), both CAP and NP (Salnikova et al., 2013b; Salnikova et al., 2008) and relapsing pneumonia (Kortyina et al., 2005). The included cases were diagnosed by experienced professionals based on symptoms, medical histories and the clinical, radiology or laboratory results (chest X-ray, spirometry measures, etc.), and physical examination of new lung infiltrates or lower respiratory tract infection. Controls were age and gender matched healthy volunteers with no previous history of relevant infectious diseases.
The methodological assessment of all the selected articles was performed by two reviewers independently using the modified Newcastle-Ottawa Scale (NOS) for non-randomised studies (Wells et al., 2001). The quality score was assigned on the basis of eight categories primarily based on three broad criteria: selection of study groups; comparability of the groups; and ascertainment of either the exposure or outcome of interest for case-control, respectively. A maximum of one star was awarded for each detail present in the study for each parameter except for comparability, where a maximum of two stars can be given. A cumulative score of the number of stars obtained for each study reflected its quality. In case of conflicting scores, a consensus was reached upon discussing with another author (RK). A study was regarded as a high-quality study when it rated six or more stars. Stata 16.0 (Stata Corporation, College Station, TX) (Sterne and Egger, 2001; Support, S. T. StataCorp, 2019) was used to generate pooled ORs between pneumonia patients and healthy controls (DerSimonian and Laird, 1986; Cochran, 1954). Heterogeneity of data was evaluated using the I2 statistics, with I2 greater than 50% considered significant heterogeneity (Higgins et al., 2003). Based on I2 value, a random effect (maximum likelihood) or fixed effect (inverse variance) model was adopted to perform the meta-analysis (Cochran, 1954). Summary ORs were represented as a point estimate and 95% confidence intervals (CIs) on a forest plot (Light et al., 1994), and publication bias was evaluated using regression based Egger’s test and Begg’s funnel plot (Begg and Mazumdar, 1994; Egger et al., 1997; Sterne et al., 2001). Subgroup analysis was performed based on pneumonia type (CAP and NP), populations (China and Russia) and age group (<12 years and ≥ 12 years). Since M. pneumoniae infection is the most common form of CAP, the data from this article is included with CAP cohort. Patients having frequently recurring (relapsing) pneumonia (J18, according to the ICD-10) with unspecified organism of infection, this has also been included with CAP for easy interpretation. Sensitivity analyses were also performed to assess heterogeneity, to estimate the influence of any individual datasets, by omitting one study at a time and examining their influence on the combined effect.

We additionally conducted a spatial analysis (among countries) of the factors that might account for the variability among COVID-19 prevalence (total cases per million). Here we used country-specific demographic and socio-economic variables (such as population density, GDP, median age, and many others), to discover any association pattern. The COVID-19 dataset used in the study was downloaded from ourworldindata.org on May 24, 2021. The population specific allele frequency data for rs2606345 and rs1048943 were obtained from 1000 genome browser (Consortium, T. G. P., 2015) on March 08, 2021. For this purpose, we initially ran a linear regression fit between COVID-19 prevalence, and the country-wise distribution of risk allele of CYP1A1 SNPs (rs2606345, and rs1048943). To further strengthen the robustness and precision of above association, and to determine the influence of other confounding variables along with the allelic distribution of CYP1A1 variants with COVID-19 prevalence, we considered 20 other predictor variables (socio-economic and demographic factors). To reduce the skewness of the data, all the variables were log-transformed before entering into the regression models. We additionally removed predictor variables with >30% missing data. We next examined the univariate relationships between predictor variables and COVID-19 prevalence to find candidate variables for our final multivariable model. The variables with \( p < 0.05 \) were considered for the multivariable models. All the analyses were performed in R 3.6.3.

3. Results

Through the initial search, a total of 1406 articles were identified (7 from PubMed, 9 from Web of science, 245 from Science Direct and 1145 from worldwidescience.org). Based on initial screening of titles and abstract 1292 publications were excluded after removing duplicates (n = 30), leaving 84 articles for full text review. Among them 17 articles were removed as they discussed some other gene but not CYP1A1 gene or its genetic variants and 59 articles were removed as they did not discuss any genetic association. Finally, 8 case control studies and 2 studies from their cross-references that met the pre-defined criteria, were included for the quantitative analysis. The flow chart for the study selection process is represented in Fig. 1. The current systematic search included ten studies totalling 5298 subjects (3049 cases and 2249 healthy controls). The study population primarily comprised of well-characterized cohorts in Russia (Korytina et al., 2005; Moroz et al., 2011; Salnikova et al., 2013a; Salnikova et al., 2013b, 2013c; Salnikova et al., 2014; Salnikova et al., 2006; Salnikova et al., 2010; Smelaya et al., 2011) and one included Chinese cohort (Zhao et al., 2017). The mean age of all the pooled participants was 29.33 ± 5.49 years (31.16 years for cases and 27.49 years for controls). All the studies discuss the association of CYP1A1 genetic variants (rs2606345, rs1048943, rs4646903) with the risk of pneumonia. The demographic characteristics and clinical details of all the included studies are provided in Table 1. For cumulative quality assessment, three of ten articles were deemed as good quality (cut off score of >7), six articles (≥5–6 score) were categorised under moderate, and finally any scores below 5 were judged as poor quality which included one article (Supplementary table 1).

3.1. Meta-Analysis Results

This meta-analysis compares pneumonia patients as cases, comprising CAP and NP subjects both, with healthy controls for association of CYP1A1 genetic variants with pneumonia susceptibility. To maintain the precision in assessing the effect size in each meta-analyses performed, we removed all the studies (n = 3) with overlapping samples before meta-analysis (Moroz et al., 2011; Salnikova et al., 2013b, 2013c). Finally, a total of seven studies were included for the meta-analyses. Among the seven included studies, four provided the data on rs2606345 (Salnikova et al., 2013a; Salnikova et al., 2014; Smelaya et al., 2011; Zhao et al., 2017), five provided data on rs1048943 (Korytina et al., 2005; Salnikova et al., 2013a; Salnikova et al., 2014; Salnikova et al., 2008; Salnikova et al., 2010) and three provided the data on rs4646903 (Salnikova et al., 2013a; Salnikova et al., 2014; Salnikova LE, 2010). The details of included studies, their sample size, population, risk allele, and allele distribution in cases and controls for each SNP are represented in Table 2. Of all the CYP1A1 variants studied, we observed the most significant association of rs2606345, and rs1048943 but no significant association was established for rs4646903. Our meta-analysis demonstrated that CYP1A1 genetic polymorphisms significantly correlated with the increased risk of pneumonia under the allelic model for rs2606345 (T vs G; \( OR = 1.12 \) (0.75–1.50); \( p < 0.02; I^2 = 84.89\% \)), and rs1048943 (G vs T; \( OR = 1.19 \) (0.76–1.61); \( p < 0.02; I^2 = 0.00\% \)) as risk markers (Fig. 2 and 3) however no statistical significance was achieved for rs4646903 (C vs T; \( OR = 0.86 \) (0.65–1.07); \( p = 0.4; I^2 = 0.00\% \)) (Fig. 4), when compared with healthy controls.

3.2. Test for publication bias

As shown in Supplementary fig. 1, visualization of the Begg’s funnel plot suggested that Egger’s linear regression test yielded evidence of publication bias among the included studies, therefore, we further performed subgroup analysis and sensitivity analysis to assess the robustness and consistency of our meta-analysis findings.

3.3. Subgroup and sensitivity analysis

Due to heterogeneity in the meta-analysis, we attempted to subgroup the studies. Subgroup analysis was performed based on pneumonia type (CAP and NP), different population (China and Russia), and age group (<12 years and ≥ 12 years) (Table 3, Supplementary fig. 5). Significant association was observed for CAP subgroup for rs2606345 (\( OR = 1.43 \) (1.19–1.66); \( p < 0.0001 \)) and rs1048943 (\( OR = 1.29 \) (0.76–1.18); \( p = 0.02 \)). Sensitivity analysis results are as represented in Table 4, showing
3.4. Genetic variability of CYP1A1 genetic variants in global populations and a possible link to COVID-19

Since rs2606345 (A) is the major allele in Europeans (66.6%) but not in other populations (African 5%, Asian 5–30%) (Consortium, T. G. P., 2015), we explored its possible association, and the risk allele (C) of rs1048943, with high variability in regional prevalence of the ongoing COVID-19 pandemic, through a spatial analysis. The COVID-19 prevalence varied widely across countries (Fig. 5). To assess the pattern of association between country-specific prevalence with all the candidate predictor variables (as shown in Supplementary table 2) from a univariate regression analysis for each individual predictor. We found that all the variables except population, population density, diabetes prevalence, and male smokers were significantly associated with prevalence in these univariate models. Excluding these non-significant variables and variables with >30% missing data (ICU patients per million, hospital patients per million, total tests per thousand, poverty, and female smokers), all the remaining variables were included the multivariable linear regression model where we sequentially added them in one-by-one and evaluated the model fit. The variable producing the best fit was retained in the model. Our final model included total death per million (p < 2 \times 10^{-16}), percentage of population with age > 65 older (p = 1.78 \times 10^{-5}), human development index (p = 1.73 \times 10^{-10}), stringency index (p = 0.0286), SNP frequency for rs2606345A (p = 0.0005) and SNP frequency for rs1048943G (p = 0.01). This model had good agreement between these variables with total cases per million across countries (r^2 = 0.8983, p < 2.2 \times 10^{-16}). The linear model regression exploring the variation in COVID-19 prevalence with the allele frequency data (for rs2606345 and rs1048943) for each country are depicted in Fig. 6 and 7. This model showed a significant association...
Table 1
Main characteristic of studies included in meta-analysis for CYP1A1 genetic variants associated with risk of pneumonia.

| No | Study details | Population (Ethnicity) | Genotyping method | Disease Studied CYP1A1 variants (alt. allele) | Genotypic variant details | Allelic variant details | Score Quality |
|----|---------------|------------------------|-------------------|---------------------------------------------|---------------------------|------------------------|--------------|
| 1  | Zhao J., et al. (2017) (Zhao et al., 2017) | China (East Asian) | PCR Sequencing | MPP | rs2606345 (T) | TT (<0.0001) | T(0.764) 1.07 (0.68–1.66) 7 |
| 2  | Salnikova, L. E., et al. (2013) (Salnikova et al., 2011a, 2013b, 2013c) | Russia (European) | Allele specific tetra-primer PCR | CAP | rs2606345 (T), rs4646903 (C), rs1048943 (G) | T/T rec (3.9 × 10⁻⁵), TT (0.093), AA (0.188) | T(0.00001), T (0.117), A (0.19) 5 |
| 3  | Salnikova, L. E., et al. (2014) (Salnikova et al., 2011a) | Russia (European) | PCR-CTPP | NP | rs2606345 (T), rs4646903 (C), rs1048943 (G) | TT (0.324), TT(0.377), AA (0.88) | T(0.00045), T (0.22), A (0.56) 7 |
| 4  | Salnikova, L. E., et al. (2008) (Salnikova et al., 2008) | Russia (European) | PCR - Genotyping | CAP | rs1048943 (G) | AA (0.035) | A (0.039) 4 |
| 5  | Korytina, G. F., et al. (2005) (Korytina et al., 2005) | Russia (European) | PCR-RFLP | RP | rs1048943 (G) | AA (0.031) | A(0.009) 6 |
| 6  | Salnikova, L. E., et al. (2010) (Salnikova et al., 2010) | Russia (European) | Allele specific PCR genotyping | CAP | rs2606345 (T), rs4646903 (C), rs1048943 (G) | TT (0.01) | T(0.01) 5 |
| 7  | Smelaya T.V., et al. (2011) (Smelaya et al., 2011) | Russia (European) | Comprehensive PCR based | CAP | NA | TT (1 × 10⁻⁶) | T(0.0001) | T (0.48) (0.35–0.65) (continued on next page) |
Table 1 (continued)

| No | Study details | Case | Control | Variant details | Genotypic | Allelic | Score | Quality |
|----|---------------|------|---------|----------------|-----------|---------|-------|---------|
|    | No Study (year) [Ref] Population (Ethnicity) Genotyping method Disease | M | F | Total | Age (in years) | M | F | Total | Age (in years) | Studied CYP1A1 variants (alt. allele) | p value | OR (95% CI) | Score |
| 8  | Moroz, V. V., et al. (2011) (Moroz et al., 2011) Russia (European) Allele specific tetra-primer PCR | 307 | 27 | 334 | 26.9 ± 0.8 | 130 | 11 | 141 | 29.1 ± 0.6 | rs2606345 (T), rs4646903 (C), rs1048943 (G) | 64.04, 11.76, 0.04 | T 1.90 (1.41–2.55), T 1.43 (0.91–2.25), A 0.59 (0.27–1.31) | 6 |
|    |                                           | 176 | 40 | 216 | 43.0 ± 2.0 | 83 | 22 | 105 | 41.0 ± 1.6 | rs2606345 (T), rs4646903 (C), rs1048943 (G) | T 1.31 (0.93–1.85), T 0.91 (0.52–1.59), A 1.32 (0.58–3.00) |
| 9  | Salnikova, L. E., et al. (2013) (Salnikova et al., 2013a, 2013b, 2013c) Russia (European) Allele specific tetra-primer PCR | 321 | 29 | 350 | 27.2 ± 0.8 | 343 | 89 | 432 | 30.0 ± 0.7 | rs2606345 (T), rs4646903 (C), rs1048943 (G) | 69.15, 11.77, 0.04 | T 1.58 (1.27–1.96), T 1.28 (0.91–1.80), A 0.86 (0.53–1.39) | 6 |
|    |                                   | 224 | 44 | 266# | 43.1 ± 1.2 | 116 | 35 | 150 | 42.5 ± 1.5 | rs2606345 (T), rs4646903 (C), | 62.38, 11.13, 0.04 | T 0.86 (0.64–1.15), T 1.25 (0.77–2.04), A 0.94 (0.45–1.96) |

Bold characters highlight important phenotypic groupings and their total counts in the respective study.

M, male; F, female; PCR, polymerase chain reaction; Age of the participants shown in Mean ± Standard deviation. PCR-CTPP, polymerase chain reaction- confronting two-pair primer; RFLP, restriction fragment length polymorphism; MPP, mycoplasma pneumoniae pneumonia; CAP, community acquired pneumonia; NP, nosocomial pneumonia; RP, relapsing pneumonia; alt. Allele, alternate allele for respective SNP; allele frequency (denoted in per cent) of alternate allele calculated from respective study in control population, OR, odds ratio; CI, confidence interval; dom, dominant model; rec, recessive model.

All p values represented are uncorrected. #male/female count not given for 3 samples.

Quality assessment was performed using modified NOS scale (Wells G A, 2001) the detailed scoring can be found in Suppl. Table 1.

All the citations are as in the main manuscript file.
Table 2

| Gene       | Total No. of Population Total samples | Total No. of Risk allele samples | Risk allele present | Risk allele absent | Risk allele present | Risk allele absent | p value OR (95% CI) | F-model test | Egger’s test |
|------------|--------------------------------------|---------------------------------|--------------------|------------------|--------------------|------------------|---------------------|--------------|--------------|
| CYP1A1     | 2663                                 | 2115                            | 735                | 2218             | 1918               | 1533             | 0.02               | 1.12         | 0.002        |
| rs2606345  | 609                                  | 451                             | 158                | 1518             | 1501               | 161              | 0.40               | 0.86         | 0.1828       |
| rs1048943  | 1609                                 | 165                             | 1533               | 1698             | 161                | 161              | 0.40               | 0.86         | 0.1828       |

All patients include patients of CAP and NP both; OR, odds ratio; CI, confidence interval; F, Fixed effect model.

All p values calculated using chi-square test.

4. Discussion

This study is an attempt to investigate the role of host genetic factors in pneumonia susceptibility. Most infectious diseases, like pneumonia, are a consequence of a complex network of host genetics and pathogen genetic factors that may be inducible by several non-genetic factors. The host immune-mediated response determines the course of the disease, its susceptibility, progression, and severity. Such immune response related genes may serve as good candidates in establishing a genetic association in infectious diseases. A few genetic studies investigated the role of CYP1A1 gene and its variants with development and outcome of CAP and NP (Salnikova et al., 2008; Smelaya et al., 2011). CYP1A1 are a family of monooxygenase enzymes involved in biotransformation reactions (Danielson, 2002). Presence of CYP1A1 polymorphic variants changes the gene expression and/ activity resulting in the altered redox balance (Stading et al., 2020). This imbalance can cause chronic inflammation in the lungs aggravating the disease pathogenesis (Hussain et al., 2014; Stading et al., 2020). This study provides the first meta-analysis results reiterating the prior experimental observations that suggest a genetic contribution of CYP1A1 variants (rs2606345, rs1048943, rs4646903) establishing a more precise risk estimate of pneumonia. We additionally noted, the genetic variant (rs2606345 and rs1048943) that are crucial for CYP1A1 activity, is statistically associated with COVID-19 prevalence. This study may provide us insights towards understanding the role of CYP1A1 genetics, with inflammation, pneumonia as well as COVID-19 susceptibility.

The CYP P450 enzymes are primarily known to metabolise xenobiotics (Danielson, 2002), and also endobiotics that are derived from different physiological or pathological processes such as inflammation (Bui et al., 2011). There have been studies that tried to deduce the interplay between CYP1A1 gene and lung infections. Some studies attempted to establish an association of the CYP1A1 genetic variants with pneumonia (Korytina et al., 2005; Salnikova et al., 2013c; Salnikova et al., 2014; Zhao et al., 2017), others related the changes in its expression and activity with the infection (Fang et al., 2016; Fang et al., 2015). On the basis of such evidence, CYP1A1 seems to be an important risk factor for pneumonia. Several investigations reported the change in CYP1A1 activity by infection or inflammatory stimuli; where the activity is induced in some cases or downregulated in other (Morgan, 2001; Santes-Palacios et al., 2016). Tian et al. (2020) in their study found upregulation of CYP1A1 in peritoneal macrophages activated by bacterial lipopolysaccharides (LPS) (Tian et al., 2020). Additionally, an increased production of pro-inflammatory cytokines like TNF-α and IL-6 was observed in CYP1A1 overexpressed macrophage cells after LPS stimulation. Moreover, CYP1A1 upregulation reduced the bacterial phagocytosis by decreasing the expression of macrophage channel, SR-A. The authors concluded CYP1A1 as an important driver of infection and sepsis (Tian et al., 2020). On the contrary, in LPS-stimulated bovine mammary epithelial cells, CYP1A1 expression was drastically suppressed as compared to controls along with the elevated expression of TNF-α and IL-6. This effect was attenuated upon CYP1A1 overexpression (Hussain et al., 2014; Zhang et al., 2018). Further, Fang et al. (2015) also reported CYP1A1 downregulation in pig infected with M. hyopneumoniae (46). This study was extended in pulmonary alveolar macrophage (PAM) cells, where inflammatory response, caused by M. hyopneumoniae infection, was suppressed upon CYP1A1 overexpression, via PPAR-γ signalling pathway (8). Comparably, the genetic
polymorphism included in our study, rs2606345, determining lower CYP1A1 expression, exacerbate similar role in lung inflammation. Such evidence demonstrated that inflammatory stimuli or infection regulate CYP1A1 expression, in turn affecting cytokine production. These responses are species and tissue specific (Morgan et al., 1994; Ryan and Levin, 1993). Thus, the above findings depict immense potential of CYP1A1 in inflammation and may serve as a useful target in mitigating infections like pneumonia.

Our study revealed the association of rs2606345 and rs1048943 with risk of pneumonia, particularly CAP. The alternate allele (A in plus-strand or T in minus-strand) of rs2606345 increased pneumonia susceptibility. Despite following a stringent criterion for inclusion of studies in the present study, the possibility of observer bias in diagnosing pneumonia cannot be ruled out when dealing with retrospective records. We, therefore advice readers caution, when interpreting the results. There is a need to conduct large scale cohort studies in future to validate our findings. We also observed, this allele to be the major allele in European (66.6%) and Russian (~80%) population unlike in the other populations (African 5%, Asian 5–30%, American 39%) (Consortium, T. G. P, 2015). Interestingly, on spatial analysis, we noted significant association of this allele frequency distribution with total cases per million due to the recent outbreak of COVID-19 (Fig. 6). Likewise, for rs1048943, higher the risk allele ‘C’ frequency also showed association with COVID-19 prevalence, across countries worldwide (Fig. 7). We would also like to warn that since this is an ongoing pandemic, the numbers are changing with time and this is a circumstantial evidence.

The COVID-19, a viral pneumonia (Berlin et al., 2020; Gandhi et al., 2020), is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Kim et al., 2020; Wu et al., 2020; Yao et al., 2020; Zhang and Holmes, 2020). This activates the host immune system to release several inflammatory cytokines (IL-1, IL-6, TNFα, and IFNγ) (Li and Fan, 2020). There is a very fine balance between the immune response produced by the host cells on encountering SARS-CoV-2 and lung damage. The immune cells accelerate the production of cytokines to kill the pathogen. In case of cytokine rush, such accelerated immune response may in turn, damage the host cells. Several studies correlated the downregulation of CYP1A1 with increasing pro-inflammatory cytokines levels (IL-6, TNF-α, IL-1, IL-1β) (Wang et al., 2022). In this study we
speculate that CYP1A1 could be crucial in COVID-19 prevalence. Since this study establishes the association of CYP1A1 genetic variants with pneumonia susceptibility and one of the major symptoms in patients dying with COVID-19 infection is pneumonia (Guan et al., 2020; Surendra et al., 2021). Pneumonia may be considered a proxy phenotype for studying association with CYP1A1 with COVID-19. We also observed in this study, there is a positive correlation of ‘A’ allele frequency of CYP1A1 SNP (rs2606345) with COVID-19 prevalence among populations worldwide. Since CYP1A1 plays a vital role in innate immune response (inflammatory responses) against any kind of infection

### Table 3
Sub-group analysis for included studies comparing pneumonia patients with healthy controls for association of CYP1A1 genetic variants with pneumonia risk based on different subgroups like pneumonia type, population and age.

| S.No. | SNP     | Subgroup characteristics | Number of studies | Heterogeneity | OR (95% CI) | p value |
|-------|---------|--------------------------|-------------------|---------------|-------------|---------|
| 1     | rs2606345 | CAP                      | 3                 | 0.00          | 1.43 (1.19-1.66) | <0.0001 |
|       |         | NP                       | 2                 | 83.87         | 0.78 (0.51-1.04) | 0.06    |
|       |         | Russia                   | 4                 | 88.63         | 1.14 (0.70-1.59) | 0.03    |
|       |         | Population               |                   |               |             |         |
|       |         | China                    | 1                 | –             | 1.07 (0.58-1.56) | 0.76    |
|       |         | Russia                   | 4                 | –             | 1.14 (0.70-1.59) | 0.76    |
|       |         | Age                      |                   |               |             |         |
|       |         | <12 years                | 1                 | –             | 1.07 (0.58-1.56) | 0.76    |
|       |         | >12 years                | 4                 | 88.63         | 1.14 (0.70-1.59) | 0.03    |
| 2     | rs1048943 | CAP                      | 3                 | 0.00          | 1.29 (0.76-1.18) | 0.02    |
|       |         | NP                       | 2                 | 0.00          | 1.02 (0.30-1.73) | 0.63    |
|       |         | Age                      |                   |               |             |         |
|       |         | <12 years                | 1                 | –             | 3.86 (0.42-8.14) | 0.007*  |
|       |         | >12 years                | 4                 | 0.00          | 1.16 (0.74-1.59) | 0.18    |
| 3     | rs4646903 | CAP                      | 2                 | 0.00          | 0.81 (0.59-1.04) | 0.15    |
|       |         | NP                       | 1                 | –             | 1.25 (0.61-1.89) | 0.36    |

Bold characters highlight significantly associated alleles with respective p values.
CAP, Community acquired pneumonia; NP, nosocomial pneumonia;
All p values calculated using chi square test, except * where p value calculated using Fischer exact test.

### Table 4
Sensitivity analysis after each study was excluded by turns.

| S.No. | SNP     | No. Studies | Study Omitted | Pooled OR (95% CI) for remainders | p value | Heterogeneity |
|-------|---------|-------------|---------------|-----------------------------------|---------|---------------|
| 1     | rs2606345 | 4           | Zhao et al. (2017) | 1.14 (0.70-1.59) | 0.025 | 88.63 | 0.00 |
|       |         |             | Salnikova et al. (2014) | 1.12 (0.66-1.58) | 0.01 | 87.63 | 0.00 |
|       |         |             | Smelaya TV et al. (2011)a | 1.05 (0.63-1.47) | 0.18 | 86.36 | 0.00 |
|       |         |             | Salmikova et al. (2013a, 2013b, 2013c) | 1.34 (1.13-1.55) | 0.0002 | 12.48 | 0.22 |
|       |         |             | Salnikova et al. (2014) | 1.00 (0.62-1.39) | 0.09 | 81.53 | 0.00 |
|       |         |             | Salnikova et al. (2008) | 1.30 (0.70-1.82) | 0.008 | 88.00 | 0.00 |
|       |         |             | Salmikova et al. (2008) | 1.17 (0.74-1.61) | 0.05 | 0.00 | 0.48 |
|       |         |             | Korytina et al. (2005) | 1.16 (0.74-1.59) | 0.18 | 0.00 | 0.78 |
|       |         |             | Salmikova et al. (2010) | 1.13 (0.68-1.58) | 0.09 | 0.00 | 0.58 |
|       |         |             | Salmikova et al. (2013a, 2013b, 2013c) | 1.22 (0.60-1.85) | 0.02 | 0.00 | 0.46 |
| 2     | rs1048943 | 5           | Salnikova et al. (2014) | 0.81 (0.59-1.04) | 0.15 | 0.00 | 0.64 |
|       |         |             | Salmikova et al. (2010) | 0.85 (0.60-1.09) | 0.48 | 44.64 | 0.18 |
|       |         |             | Salmikova et al. (2013a, 2013b, 2013c) | 1.00 (0.65-1.35) | 0.79 | 0.00 | 0.36 |

Smelaya TV et al. (2011)a includes CAP patients; Smelaya TV et al. (2011)b includes NP patients.
Significant p values are represented with bold characters.
Fig. 5. A geospatial distribution map of prevalence (the number of cases per million) as on 24 May 2021 worldwide due to COVID-19. COVID-19 prevalence is extracted from ourworldindata.org till 24 May 2021. The white colored areas in the map show the absence of data. A half open interval includes only one of its end-points and is denoted by mixing notations for open and closed intervals. For e.g., $(0–1]$ means greater than 0 and less than or equal to 1 and $[0,1)$ means greater than or equal to 0 and less than 1.

Fig. 6. A geospatial frequency maps depicting the distribution of risk allele ‘A’ frequency of rs2606345 (CYP1A1).

Fig. 7. A geospatial frequency maps depicting the distribution of risk allele ‘C’ frequency of rs1048943 (CYP1A1). The data is obtained from population frequency data of the 1000genome browser on 8 March 2021. The white coloured areas in the map show the absence of data. A half open interval includes only one of its end-points and is denoted by mixing notations for open and closed intervals. For e.g., $(0–1]$ means greater than 0 and less than or equal to 1 and $[0,1)$ means greater than or equal to 0 and less than 1.
other respiratory diseases, etc.) could not be included. Further, the unrequired for true estimation of the summary effect. Another major lim
the limited population data and sample size in this meta-analysis are
the CYP1A1 gene region and its association with pneumonia, both CAP and
CYP1A1 polymorphisms with CAP, where rs2606345 and rs1048943 are
The variation in prevalence of COVID-19 across countries (Miller et al.,
In a recent in vitro study demonstrated, CYP1A1 gene expression to be
inevitable to rule out the presence of other functional variants in the
expression of CYP1A1 in other tissues like adipose, breast, liver, and skin
not yet to be explored. We also observed that genetic variant, rs2606345 and rs1048943, governing
other ethnic groups prevented us from drawing conclusion of causality of the variant as presence of association in different ethnicities would have provided evidence in this direction. Though this is a statistical analysis, our findings along with the previous reports revealing the role of CYP1A1 in inflammatory response to infection, may help us deduce the importance of this gene in infectious disease like pneumonia. The direct evidences of the molecular mechanism by which CYP1A1 contributes to pneumonia pathogenesis is still yet to be explored. We also observed that genetic variant, rs2606345 and rs1048943, governing CYP1A1 expression may give us clues towards the genetic basis of COVID-19 prevalence across populations worldwide.

Author contributions

RK devised the concept of the review, reviewed the manuscript and supervised the overall study till final manuscript preparation. DG and SY performed the literature search independently and, reviewed articles and, extracted the data. DG generated the forest and funnel plots. PS (Priyanka Singh) performed the linear regression and calculated Pearson’s correlation and prepared all the spatial analysis images. SG helped us through the meta-analysis and running the statistical analysis through STATA. DG and PS (Pooja Singh) performed the quality assessment of the included articles and prepared the supplementary material. Upon discrepancies, RK resolved the conflict. DG and SY wrote the manuscript. ST cross-checked all the data, re-calculated the numerical values and statistical analysis. NK, SK, PRP helped in manuscript writing, reference management, and preparation of tables. RK and AA reviewed the manuscript, figures and tables. DG, SY and NK edited the manuscript. SY and NK prepared the tables. RK, YH, and BP finally revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding information

Financial support for this research work has been provided by the Council of Scientific and Industrial Research (CSIR), grant number OLP1154.

The data is obtained from population frequency data of the 1000genome browser on 8 March 2021. The white colored areas in the map show the absence of data. A closed open intervals includes only one of its end-points and is denoted by mixing notations for open and closed intervals. For e.g., (0,1] means greater than 0 and less than or equal to 1 and [0,1) means greater than or equal to 0 and less than 1.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

We are highly indebted to our founder director, Prof. S.K. Brahmachari, whose vision in the field of genomics paved us a path. SY, PS (Pooja Singh), ST, NK, and PRP acknowledge CSIR, Government of India for providing their fellowships. DG, PS (Priyanka Singh) and SK acknowledge ICMR, UGC and DBT, Govt. of India, for their financial assistance, respectively. We thank the anonymous reviewers for their helpful suggestions in improving the manuscript. The authors are thankful to Dr. L. E. Salnikova, V. A. Negovsky Research Institute of General Reanimatology, Russian Academy of Medical Sciences, Moscow.
Neurology Division, for sharing genotypic and allelic data from their patient population.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mejde.2022.105299.

References

Begg, C.R., Mazumdar, M., 1994. Operating characteristics of a rank correlation test for publication bias. Biometrics 50, 1088–1101.

Berlin, D.A., Gulick, R.M., Martinson, F.J., 2020. Severe Covid-19. N. Engl. J. Med. 383, 2451–2460. https://doi.org/10.1056/NEJMc2009575.

Bui, P., Imazumii, S., Beedanagari, S.R., Reddy, S.T., Hankinson, O., 2011. Human CYP2S1 metabolizes cytochrome-oxgenase- and lipoxigenase derived eicosanoids. Drug Metab. Dispos. 39 (2), 180–190. https://doi.org/10.1124/dmd.110.035121.

Cochran, W.G., 1954. The combination of estimates from different experiments. Biometrics 10, 101–129. https://doi.org/10.1016/j.mejde.2022.105299.

Kumar, V., Wijmenga, C., Xavier, R.J., 2014. Genetics of immune-mediated disorders: how and how? Drug Metabolism 29, 207–212.

Morgan, E.T., Thomas, K.B., Swanson, R., Vales, T., Hwang, J., Wright, K., 1994. Selective suppression of cytochrome P-450 gene expression by interleukins 1 and 6 in rat liver. Biochem. Biophys. Acta 1219 (475), e483. https://doi.org/10.1016/S0923-3474(10)00198-2.

Mackenzie, A., 2016. The definition and classification of pneumonia. Pneumonia 8, 14. https://doi.org/10.3836/j.issn.0578-1752.2015.14.015.

Meletiadis, J., Chanock, S., Walsh, T.J., 2006. Human pharmacogenomic variations and their implications for antifungal efficacy. Clin. Microbiol. Rev. 19 (4), 763–787.

Miller, L.E., Bhattacharyya, R., Miller, A.L., 2020. Data registries: country-specific variability in Covid-19 prevalence, incidence, and case fatality rate. Data Brief. 32, 106276. https://doi.org/10.1016/j.dab.2020.106276.

Morgan, E.T., 2001. Regulation of cytochrome p450 by inflammatory mediators: why and how? Drug Metabolism 29, 207–212.

Poroshenko, G.G., Rubanovich, A.V., 2010. Genetic predisposition to the development of pneumonia. Oboire. Animatologi. 4, 6. https://doi.org/10.15360/1813-9779-2008-2-6.

Rothman, M., Yu, K., Lubin, J.H., Consonni, D., Pesatori, A.C., Goldstein, A.M., Goldin, L., 2006. The strengths and limitations of meta-analyses based on aggregate data. BMC Med. Res. Methodol. 5, 14. https://doi.org/10.1186/1471-2288-5-14.

Terada, N., Vanhoutte, P.M., 1983. Hypercontractility of rabbit thoracic aorta in situ induced by angiotensin II and norepinephrine. J. Mol. Cell. Cardiol. (Berl) 15 (9), 700–710. https://doi.org/10.1016/0022-2010(83)90066-0.

Vijay-Kumar, A., Rastogi, M.R., 2014. Genetics of immune-mediated disorders: from genome-wide association to molecular mechanism. Curr. Opin. Immunol. 31, 51–67. https://doi.org/10.1016/j.coi.2014.09.007.
Stading, R., Chu, C., Couroucli, X., Lingappan, K., Moorthy, B., 2020. Molecular role of cytochrome P4501A enzymes in oxidative stress. Curr. Opin. Toxicol. 20-21, 77–84. https://doi.org/10.1016/j.cotox.2020.07.001.

Sterne, J.A.C., Brandram, M.J., Egger, M., 2001. Meta-Analysis in Stata. Systematic Reviews In Health Care: Meta-Analysis in Context, pp. 347–372.

Support, S. T. StataCorp, 2019. Citing Stata software, documentation, and FAQs. Stata Statistical Software: 2019, Release 16. StataCorp LLC, College Station, TX. Available online. https://www.stata.com/support/faqs/resources/citing-software-documentation-faqs/.

Tanimoto, K., Hirota, K., Fukazawa, T., Matsuo, Y., Nomura, T., Tanuza, N., Hirohashi, N., Bono, H., 2021. Inhibiting SARS-CoV-2 infection in vitro by suppressing its receptor, angiotensin-converting enzyme 2, via aryl-hydrocarbon receptor signal. Sci. Rep. 11 (1), 16629. https://doi.org/10.1038/s41598-021-96109-w.

Tian, L.X., Tang, X., Zhu, J.Y., Luo, L., Ma, X.Y., Cheng, S.W., Zhang, W., Tang, W.Q., Ma, W., Yang, X., et al., 2020. Cytochrome P450 1A1 enhances inflammatory responses and impeded phagocytosis of bacteria in macrophages during sepsis. Cell Commun Signal. 18 (1), 70. https://doi.org/10.1186/s12976-020-00523-S.

Torres, A., Cillóniz, C., Niederman, M.S., Mene´ndez, R., Chalmers, J.D., Wunderlin, R.G., van der Poll, T., 2021. Pneumonia. Nat Rev Dis Primers 7, 25. https://doi.org/10.1038/s41572-021-00259-0.

Wang, G., Xiao, B., Deng, J., Gong, L., Li, Y., Li, J., Zhong, Y., 2022 Feb 2. The role of cytochrome P450 enzymes in COVID-19 pathogenesis and therapy. Front. Pharmacol. (13) https://doi.org/10.3389/fphar.2022.791922. PMID: 35185562; PMCID: PMC8847594, 791922.

Wells, G.A., Shea, B., O’Connell, D., Peterson, J., Welch, V., Losos, M., Tugwell, P., 2001. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Non randomised Studies in Meta-Analyses. University of Ottawa.

WHO, 2022. The top 10 Causes of Death. Available online. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (accessed on 20 October 2021).

Zhao, J., Zhang, W., Shen, L., Yang, X., Liu, Y., Gai, Z., 2017. Association of the ACE, GSTM1, IL-6, NOS3, and CYP1A1 polymorphisms with susceptibility of mycoplasma pneumoniae pneumonia in Chinese children. Medicine 96, 15 e6642. https://doi.org/10.1097/md.0000000000006642.