



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## The role of host genetic factors in respiratory tract infectious diseases

**Citation for published version:**

Patari, I, Gelemanovi, A, Kirin, M, Koli, I, Theodoratou, E, Baillie, KJ, de Jong, MD, Rudan, I, Campbell, H & Polašek, O 2015, 'The role of host genetic factors in respiratory tract infectious diseases: systematic review, meta-analyses and field synopsis' *Scientific Reports*, vol. 5, pp. 16119. DOI: 10.1038/srep16119

**Digital Object Identifier (DOI):**

[10.1038/srep16119](https://doi.org/10.1038/srep16119)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Scientific Reports

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# SCIENTIFIC REPORTS



OPEN

## The role of host genetic factors in respiratory tract infectious diseases: systematic review, meta-analyses and field synopsis

Received: 20 July 2015

Accepted: 09 October 2015

Published: 03 November 2015

Inga Patarčić<sup>1,\*</sup>, Andrea Gelemanović<sup>1,\*</sup>, Mirna Kirin<sup>1</sup>, Ivana Kolčić<sup>1</sup>, Evropi Theodoratou<sup>2</sup>, Kenneth J. Baillie<sup>3,4</sup>, Menno D. de Jong<sup>5</sup>, Igor Rudan<sup>2</sup>, Harry Campbell<sup>2</sup> & Ozren Polašek<sup>1,2</sup>

Host genetic factors have frequently been implicated in respiratory infectious diseases, often with inconsistent results in replication studies. We identified 386 studies from the total of 24,823 studies identified in a systematic search of four bibliographic databases. We performed meta-analyses of studies on tuberculosis, influenza, respiratory syncytial virus, SARS-Coronavirus and pneumonia. One single-nucleotide polymorphism from *IL4* gene was significant for pooled respiratory infections (rs2070874; 1.66 [1.29–2.14]). We also detected an association of *TLR2* gene with tuberculosis (rs5743708; 3.19 [2.03–5.02]). Subset analyses identified *CCL2* as an additional risk factor for tuberculosis (rs1024611; OR = 0.79 [0.72–0.88]). The *IL4-TLR2-CCL2* axis could be a highly interesting target for translation towards clinical use. However, this conclusion is based on low credibility of evidence - almost 95% of all identified studies had strong risk of bias or confounding. Future studies must build upon larger-scale collaborations, but also strictly adhere to the highest evidence-based principles in study design, in order to reduce research waste and provide clinically translatable evidence.

Infectious diseases are characterized by a number of unique features - they rely on a single agent as a cause, they can be transmitted from one person to another and cause epidemics, and they have had a strong impact on human evolution<sup>1,2</sup>. In addition, infectious diseases can be eradicated, but also new ones may emerge, creating a dynamic stage for human-infection interplay<sup>3</sup>.

Even the most basic insight into the infectious disease occurrence and outcomes suggests strong inter-individual differences, commonly attributable to the host genetics profile. Adoption, twin and heritability studies provided the first line of evidence that corroborate this<sup>4,5</sup>, leading to an increased interest in understanding of genetic background of infectious disease in the last decade of the 20<sup>th</sup> century<sup>6,7</sup>. As a result, over 4,000 candidate-gene studies were published between 2001 and 2010, often focusing on respiratory infections and especially on tuberculosis<sup>8</sup>. However, these studies often provided conflicting results<sup>9,10</sup>. Some of the problems of such studies included low study power, high risk of publication bias, differences in study designs, especially cases and controls recruitment schemes, which altogether led to often bias and confounding, genotyping inaccuracies, and substantial problems in phenotype definition, especially in the case of tuberculosis<sup>11–13</sup>. The entire field was plagued by very unfavourable effect-to-bias ratios, a situation in which the magnitude of bias exceeds that of the sought effect<sup>14</sup>. Not even the use of

<sup>1</sup>Department of Public Health, University of Split School of Medicine, Split, Croatia. <sup>2</sup>Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. <sup>3</sup>Roslin Institute, University of Edinburgh, Midlothian, UK. <sup>4</sup>Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, UK. <sup>5</sup>Department of Medical Microbiology, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to O.P. (email: ozren.polasek@mefst.hr)

hypothesis-free approaches, like genome-wide association studies<sup>15</sup>, managed to spark a progress of the field, indicated as a lack of replication of results from such studies<sup>10,16–18</sup>. Unfortunately, insights derived from the fields of clinical infectology, microbiology, immunology, epidemiology, as well as clinical, evolutionary and population genetics remained largely isolated from one another<sup>19–21</sup>, preventing systematic understanding of the entire field.

The interest in host factors associated with respiratory infectious diseases stems from their ability to cause epidemics and pandemics. The infamous Spanish influenza of 1918 has been implicated as the largest ever recorded pandemic, causing an estimated 25–100 millions of deaths<sup>22,23</sup>. On the other hand, tuberculosis is an example of highly adaptive pathogen, which managed to coevolve with humans, possibly as far as the initial waves of human migrations out of Africa<sup>24–26</sup>. The 2013 estimates suggested that 9 million of people suffered from it globally, with 1.5 million of attributable deaths annually and possibly as much as a third of all living humans being latent carriers<sup>27</sup>. Pneumonia remains to be one of the main childhood mortality causes, with an estimated 120 million episodes and 1.3 million of lethal outcomes globally<sup>28</sup>. Similarly, respiratory syncytial virus (RSV) is attributable for 33.8 million episodes of newly diagnosed acute lower respiratory infections worldwide in children under 5 years, with at least 3.4 million episodes of severe cases requiring hospital admission<sup>29</sup>. Finally, SARS-Coronavirus outbreak in 2003 was an example of how quickly a novel respiratory pathogen can spread on a global scale<sup>30,31</sup>.

Understanding the host genetic side could be an invaluable tool in clinical medicine, especially because susceptibility to an infectious agent lies at least partly hidden or masked in inborn errors or immune response<sup>32,33</sup>. This renders infectious diseases a high-ranking research priority, having in mind the mobility of modern human population and the ever changing pathogen nature. Therefore, the aim of this study was to provide a field synopsis and systematic review with meta-analyses of host genetic susceptibility to respiratory infectious diseases. A special focus was set on the comparative analysis of different clinical presentations and pathogens, aiming to increase the understanding of both shared and specific disease pathways that could have been identified on the basis of previously published candidate gene studies. We also provide a critical overview of the entire field and generalized framework for improvement of future studies in infect-omics - application of genomics, proteomics metabolomics or other omics technologies in understanding of the infectious disease development, progression and outcome.

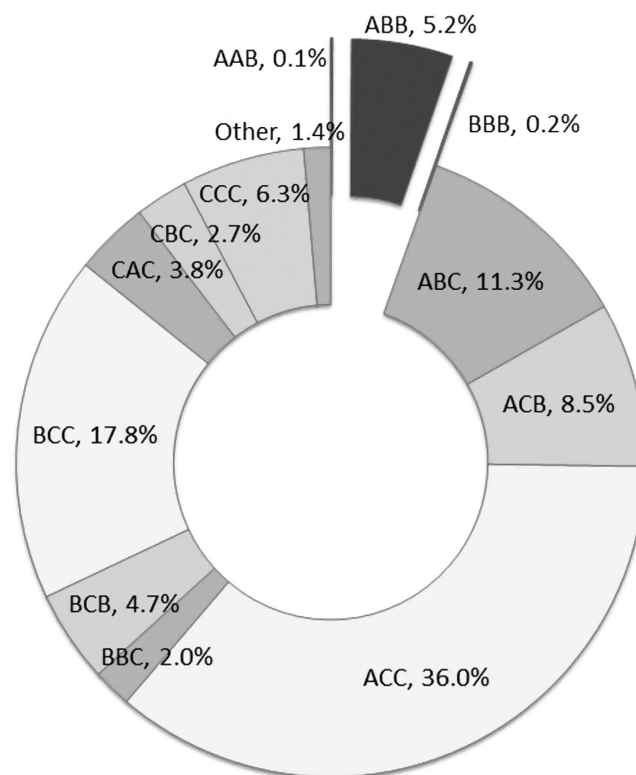
## Results

A total of 2,209 data points were extracted from 386 studies. The majority of data addressed tuberculosis (1,417 data points; 64.1%), followed by RSV (285; 12.9%), SARS-Coronavirus (198; 9.0%) and influenza (84; 3.8%), pneumonia (172; 7.8%), while the rest were on other respiratory infections such as otitis media and bronchitis (53; 2.4%). There were 274 different genes in the data, with most frequent results for *TNFA* (123; 5.6%), *IL10* (117; 5.3%) and *SLC11A1* (92; 4.2%). A total of 949 distinctive markers were recorded in the database, but only 89 met the inclusion criteria of at least four data points available per marker, and were used in the subsequent meta-analysis. Out of 1,963 data points with all three CSI score domains scored, only 107 data points (5.5%) had credible CSI score (having only A or B grades), while the remaining 94.5% of data points had at least one C grade and were thus considered to have weak credibility (Fig. 1). Disease-specific CSI profiles indicated better quality of studies on pneumonia, influenza and tuberculosis, while CSI scores were lower in case of SARS-Coronavirus and respiratory syncytial virus (RSV) (Supplementary Figure 1).

Among investigated data points, a total of 1,952 were attributable to infectious disease susceptibility (88.4%), 172 to disease severity (by comparing severe vs. mild disease cases; 7.8%) and 38 to mortality (1.7%), while the remaining studies were based on comparison of tuberculosis skin-test positivity results with controls. We performed a total of 515 disease susceptibility meta-analysis in all four genetic models, with a total of 86 nominally significant results (Supplementary Tables 3–6). However, only two genes retained noteworthiness for the mid/low BFDP level (Table 1). The first one was *IL4* gene, for pooled result in allelic model (rs2070874; OR = 0.79, 95% CI [0.70–0.89]; Supplementary Figure 2), and pooled result in recessive model (rs2070874; 1.66 [1.29–2.14]; Supplementary Figure 3). The second one was *TLR2* gene for tuberculosis in allelic model (rs5743708; 3.19 [2.03–5.02]; Supplementary Figure 4). We did not detect strong traces of publication bias in all meta-analyses (plots available in RISEdb website).

Subset analysis of methodologically better studies (with controls of known exposure to the pathogen) revealed five nominally significant results, with a single noteworthy result (Supplementary Table 7) - *CCL2* gene (rs1024611; OR = 0.79 [0.72–0.88]; Table 1). There were no significant results for severity models (Supplementary Table 7), while we did not perform mortality meta-analysis due to lack of at least four published studies for any marker.

We identified a total of 19 family and linkage-based studies, which were not included in the meta-analysis due to methodological restraints. These studies covered either broad genome regions on chromosomes 3–7, 10–12, 15, 17, 19, 20, X, or specific genes such as *SLC11A1* (*NRAMP1*), *IL8*, *CCR5*, *IL12RB1*, *SP-A*, *SP-D*, *SLC22A4*, *SCL22A5*, *RAD50*, *FBOX11* and *EVII* gene (Supplementary Table 8). We also identified 11 GWAS studies; seven were on tuberculosis, three on otitis media and one on influenza. Two genome-wide significant results were recorded for tuberculosis, belonging to *WT1* gene and *ASAP1* gene (Supplementary Table 9). We also identified 54 published meta-analyses, of which the majority addressed genetic susceptibility to tuberculosis. However, these meta-analyses investigated



**Figure 1.** Quality score assessment, based on confounding, selection bias and information bias risk, based on 1,963 data points from RISE database (only data points with all three CSI domain scores were analysed here).

Gene	Disease; genetic model	rs code	Heterozygote	Risk allele	N studies	N cases (alleles)	N Controls (alleles)	OR [95% CI]	P	I <sup>2</sup> [95% CI]	Venice score	BFDP (med/low)	BFDP (very low)
IL4	Pooled result; allelic	rs2070874	CT	T	6	2908	6422	0.79 [0.70–0.89]	7.68E-05	0.06 [0.00–0.76]	BAC	0.063	0.779
IL4	Pooled result; recessive	rs2070874	CT	T	5	1059	2210	1.66 [1.29–2.14]	8.68E-05	0.00 [0.00–0.39]	BAC	0.128	0.886
TLR2	Tuberculosis; allelic	rs5743708	AG	A	6	3262	3124	3.19 [2.03–5.02]	4.91E-07	0.00 [0.00–0.63]	BAC	0.087	0.834
CCL2	Tuberculosis; heterozygote advantage, in exposed controls	rs1024611	AG	G	4	3066	3556	0.79 [0.72–0.88]	1.26E-05	0.00 [0.00–0.83]	BAC	0.014	0.426

**Table 1.** Significant and noteworthy results of meta-analyses (the entire set of results is available in Supplementary Tables 3, 4, 5 and 6). Abbreviations used: TBC – tuberculosis, BFDP - Bayesian false-discovery probability.

infection susceptibility in much wider set of clinical appearances, often with manifestations not restricted to respiratory tract, and were thus not directly comparable to this study.

## Discussion

This is the first extensive, systematic review and meta-analysis of all published studies that were addressing host genetic factors implicated in five common respiratory tract infectious diseases - tuberculosis, influenza, respiratory syncytial virus (RSV), SARS-Coronavirus and pneumonia. Until now, majority of published studies and meta-analyses were based on one pathogen - one polymorphism approach, focusing on pathogen susceptibility. Only a handful of published studies compared different respiratory tract diseases. In addition, majority of previously published meta-analyses addressed tuberculosis, and none was found for SARS-Coronavirus, RSV and influenza. Therefore, by developing this entire resource we

enabled not only disease-specific analyses, but also analyses of the shared (pooled) infectious disease mechanisms in the respiratory system - all in the stringent evidence-based manner.

In total, we found two results that withstood multiple testing correction: *IL4* in pooled model and *TLR2* for tuberculosis. Additionally, we established a role of *CCL2* in tuberculosis, by including only a subset of methodologically better studies, where controls were of known exposure.

*IL4* has previously been described to have a pivotal role in shaping the nature of immune response, by promoting and stimulating both T-cell and B-cell differentiation<sup>34</sup>. It provides a balance between Th1 and Th2 response, and therefore alteration of its function may substantially affect immune response<sup>34</sup>. Most commonly reported such alteration is associated with increased risk of atopy and allergies<sup>35</sup>. Furthermore, *IL4* seems to have a direct role in infectious disease outcome - *IL4*-deficient mice were more susceptible to *Legionella pneumophila* and had increased mortality rates compared to controls<sup>36</sup>. This was the single result that was significant in pooled analyses, and marginal in disease-specific meta-analyses. This reflects the need for studies with more statistical power, but also suggests that this is one of the most likely translatable results towards clinical application across the wider spectrum of respiratory infectious diseases.

*TLR2* is a member of large family of genes involved in pathogen recognition and signalling cascade of innate immune response<sup>37</sup>. Previous studies have suggested its role in the susceptibility to tuberculosis<sup>38–40</sup>, but also in other bacteria such as *Mycoplasma*, *Mycobacterium*, *Neisseria* and other Gram-positive bacteria<sup>41</sup>. Interestingly, despite being closely linked to *CD14* mechanisms<sup>37</sup>, the results of this study showed that *CD14* had no association with tuberculosis, while *TLR2* had a very strong result, with odds ratio of 3.2. This finding could point towards a specific role of *TLR2* in tuberculosis, with weaker association or even lack of such association in other elements of the immune response. Interestingly, *TLR2* effects seem to be very context-specific, as the absence of *TLR2* in knock-out mice did not affect their clinical outcomes, nor did it seem to have a role in secondary infections<sup>42</sup>. This suggests that the exact mechanisms of its function are still elusive, and that more focused studies are needed, taking into account the exact pathophysiological processes in patients with tuberculosis.

*CCL2* gene has been associated with diseases that include monocytic infiltrations<sup>43</sup>, migration and retention of monocytes in particular locations<sup>44</sup>, and consequently in granuloma formation<sup>45–47</sup>. As such, its role in tuberculosis has been described previously<sup>48,49</sup>, with notable ethnic-specific differences<sup>50</sup>. In this study, *CCL2* gene was significant only in a subset analysis, where studies with exclusively exposed controls were sub-selected. This clearly demonstrates that methodological improvements can have positive effect on gene discovery. Very complex genetic architecture of infectious diseases development, coupled with errors and omissions in primary study designs, might explain this accumulation of large amount of poorly usable evidence. Therefore, one of the main conclusions of this study is that the best way forward is through improvement of primary studies and reduction of research waste.

An overview of family and linkage-based studies suggested frequent differences in study designs, marked by dissimilarities in phenotype definitions, marker types and analytic approaches. Nevertheless, the association of tuberculosis with *SLC11A1* was replicated in two studies<sup>51,52</sup>, with a notable exception in the third one (Supplementary Table 8)<sup>53</sup>. There was no overlap of candidate-gene studies and GWAS<sup>10,54</sup>, thus prohibiting comparative analysis. All GWAS studies provided a two significant and replicated result, located in an intergenic region near *WT1* gene (rs2057178)<sup>16,55</sup> and *ASAP1* gene (rs10956514)<sup>56</sup>. Other reported signals were not replicated in independent cohorts (Supplementary Table 9). This is likely a consequence of largely underpowered study designs (with sample sizes rarely exceeding 1,000 cases or controls), but more importantly methodological problems that were carried over from the candidate-gene study designs. Up-scaling the study in terms of genetic resolution (from single marker or just a few markers to thousands and hundreds of thousands of markers in GWAS), increase in the sample size and use of more advanced analytic methods will likely not provide a substantial step forward before we manage to overcome methodological and study design limitations. In addition, we are faced with a high level of genetic complexity<sup>57</sup>, and duality between rare, Mendelian variants, which may confer complete immunity, but have very low prevalence in a population<sup>58</sup>, or common variants, which usually have much lower explained variance, and therefore have a limited capacity for clinical intervention. The principal step forward is to create larger-scale collaborations and enable data sharing, built upon the common framework for analysis, based on aligned protocols, larger-scale harmonizing and collaborative efforts<sup>59</sup>.

The results of this study fit into previously implied mechanisms of respiratory infections, where *CCL7-CCL2-CCR2* axis was described to have a critical role in *IL4* production and immune response and regulation in both fungal<sup>60</sup> and viral infections<sup>61</sup>, as well as atherosclerosis<sup>62</sup> and tumours<sup>63</sup>. Despite differences described from *in vitro* and *in vivo* studies<sup>64</sup>, this could be the most interesting translatable mechanism towards clinical medicine. In addition, *WT1* gene from GWAS studies<sup>16,55</sup>, implied in susceptibility to tuberculosis, has been described to be under control of *CCL2*, which seems to be having target-responsive cytotoxic activity on *WT1*-specific mechanisms<sup>65</sup>. This mechanism could, at least partly, explain the relationship between some infections and tumours<sup>66</sup>. However, to clarify their relationship and provide a step forward, we must firstly be able to derive much better evidence from primary studies.

Primary study design was one of the main barriers for the further field development, with almost 95% of all extracted data points assessed as having strong risk of bias or confounding. The improvements must be embodied into evidence-based principles, and have to be undertaken in almost all domains of the field, including better clinical definitions and phenotyping, improved controls selection and diagnostics, appropriate statistical analysis, favourable study design, the use of novel molecular technologies

and better reporting with larger sample sizes obtained through consortia development. Firstly, we need to improve phenotype definitions and criteria for diagnosis, especially in case of latent tuberculosis<sup>13,67</sup>. The use of proxy and extreme phenotypes, which was seen in HIV, with exposed uninfected subjects, long-term non-progressors, fast progressors and elite controllers<sup>68–70</sup>, might also provide an interesting step forward<sup>71</sup>. It was very common for authors to use the term “healthy” controls, but almost universally used dissimilar diagnostic process for classification of cases and controls. Thus, selection of controls and diagnostics process must be fundamentally improved and harmonized (Supplementary Table 10). A substantial effort must go towards inclusion of controls that were exposed to a pathogen, which is necessary for the disease development (but may not be sufficient). Data analysis must also be improved, by adhering to the best practices, including the use and report of odds ratios and confidence intervals as measures of association, and multiple testing correction<sup>72</sup>. Improvements can be made in the study design, where availability of finely phenotyped life-long cohorts and biobanks provides an opportunity to use this data in an understanding of life-long risk of developing respiratory infectious disease, such as tuberculosis. The use of novel molecular technologies, various *omics*, and their application to infectious disease development, progression and outcome (infect-*omics*) presents an unprecedented opportunity in this field<sup>73</sup>. However, these must be based on evidence-based, always include estimates of repeatability and focus on replication efforts in an unlinked population. Lastly, we must substantially improve results reporting. This refers to the structured manuscript preparation, reporting of all relevant information and manuscript preparation against a given set of criteria like STREGA<sup>74</sup>, STROBE-ME<sup>75</sup> or STREIS, where appropriate<sup>76</sup>. This also extends to the need to develop individual-data public repositories and creation of large-scale consortia that share common standard-operating procedures, similarly to these in chronic disease genetic epidemiology. We do not neglect the proposed finer scale analyses of various population subsets<sup>67</sup>, but we consider improvements in study design to have much stronger potential impact. All these suggestions are applicable not only for genomics, but also for any type of infect-*omics* studies, defined as application of various *-omics* technologies in understanding of the infectious disease development, progression and outcome. In addition, we need to start developing framework of systematic reviews, where we need to compare disease-specific genes and pathways to disease-specific genes, as the final outcome, clinical disease, might be attributable to different sets of genes that make a person more susceptible to e.g. pneumonia regardless on the pathogen that is causing it. Only by comparing such complementary results and insights will we be able to better understand and harness this information in clinical care.

These results must be viewed in light of numerous limitations. We reviewed a very diverse field with no unified reporting scheme, suggesting that the search might have missed some relevant articles and data sources. In order to accommodate for this problem we performed extensive search supplemented by hand-search procedure described in Materials and methods section. Existence of wider-scale unpublished data and occasionally our inability to obtain raw data present another possible source of bias. This study also suffers from the limited input, due to exclusion of HLA and KIR markers, haplotypes and complete omission of any information on pathogen, thus providing rather limited insight into the field and possibly missing out some of the key element for more in-depth understanding<sup>77</sup>. Most of these limitations have very negative impact on effect-to-bias ratio, where bias may be several orders of magnitude greater than the size effects. We attempted to contrast this by the use of the stringent analytic approach in an evidence-based manner, including multiple testing correction, careful application of inclusion criteria and quality assessment scoring, coupled with use of random-effect meta-analysis. The study also suffers from inability to adjust the analysis for basic covariates, such as age, sex and comorbidity information. We also used a novel score for quality assessment that was developed within this study and was not validated before. We thus provided basic agreement indices, but wider use of this score would require an in-depth reliability assessment and validation. If we turn back to biological processes, earlier primary studies ignored microbiomes, while some novel studies are suggesting that lungs are not sterile<sup>78,79</sup> and that respiratory tract microbiomes could have strong modifying effect on health and disease mechanisms. Despite all these limitations, we attempted to provide a field synopsis, supplemented by a generalized framework that could be useful not only for respiratory infectious disease genetics, but can be extended to all infectious diseases. However, these results are just the first step in an attempt to provide a reflection of just one side of the infectious disease story – the host. Only by combining the host genomics (and other *-omics*) information with pathogen *-omics* will we be able to truly advance this field and enable the development of personalized infectious disease medicine.

## Materials and Methods

**Protocol and registration.** The review strategy and details were registered prospectively in PROSPERO as the record 2014:CRD42014009072.

**Eligibility criteria.** We undertook an extensive and systematic search of all published studies related to host genetics implicated in development or outcome of respiratory tract infectious diseases. Instead of focusing on a single-disease approach, we aimed at delivering systematic evidence on all respiratory infections, including cases where a syndrome phenotype was reported, such as pneumonia, without the specific pathogen information.

**Information sources.** We used three bibliographic databases; PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>), Web of Science (<http://wok.mimas.ac.uk>) and Scopus (<http://www.scopus.com>). We also performed an additional search of the HuGe Literature Finder (<http://www.hugenavigator.net>), to identify articles that might have been missed in PubMed.

**Search.** Two of the joint first authors performed the search and study selection, which was supervised by the last author. The last author also resolved any discrepancies in scoring system or data extraction process. The search was performed on May 06, 2014, with an update performed on August 25, 2015. After duplicates removal, the search yielded a total of 24,823 articles. All identified articles were firstly checked for inclusion by reading the Title and Abstract. This led to identification and exclusion of a total of 23,677 studies. The remaining 881 articles were supplemented by hand search of their references, producing an additional set of 265 articles. In total, we used a set of 1,146 articles that were read in full and screened against inclusion criteria (Supplementary Figure 5).

**Study selection.** The study selection process was based on iterative steps. Firstly, the studies needed to report the number of cases and controls for every analysed genotype. Therefore, we initially excluded all of the family-based and linkage studies, as well as genome-wide association (GWAS) studies from this meta-analysis, due to inability to obtain complete set of GWAS results in the appropriate format. However, we retained these studies and provided their qualitative summary, in order to develop an inclusive field synopsis. Secondly, studies had to be based on biallelic single-nucleotide polymorphisms (SNP) or biallelic insertion-deletion marker type. Studies based on other marker types (including microsatellites or short tandem repeats), or those that reported aggregated haplotypes were excluded. We also excluded all studies that reported exclusively gene expression profiles without relevant genotyping information. We then limited the review to studies that were published during or after the year 2000 and were written in English language. Despite similarity of disease patterns and mechanisms, we excluded all animal and all *in vitro* studies and focused on susceptibility in previously healthy subjects. We also excluded studies that reported: any outcome in previously affected patients (either as a part of immunodeficiency syndromes or with any form of disease such as asthma, HIV/AIDS or other comorbidities; where reported), fungal infections (due to also very prevalent comorbidities), nosocomial infections (such as ventilator-related or hospital-acquired pneumonia), exclusively pathogen genotypes or host-pathogen interactions. Studies that reported non-respiratory infection sites (such as extra-pulmonary tuberculosis, meningitis or invasive pneumococcal disease) were also excluded, as the focus of this study was on respiratory infectious diseases. Lastly, we excluded all studies that were obviously reporting previously published results or were re-using the cases that has already been used elsewhere (when we managed to identify this), in order to remove redundant data or publications from meta-analysis. To ensure a systematic source of information, we additionally checked studies which provided insufficient information. A total of 142 e-mails were sent to authors of studies which we classified as having reported insufficient information, and asked for clarifications or raw data sharing. We also contacted all authors of identified GWAS studies to share their results. This entire process results in a total of 10 responses (4 rejections, 2 responses that did not meet the inclusion criteria and 4 cases of clarifications or shared raw data, which are listed in the Acknowledgements). In total, we used 386 articles in the data extraction process (Supplementary Figure 5).

**Data collection process and data items.** A total of 48 pieces of information were extracted for every study (available in the on-line RISEdb – Respiratory Infection Susceptibility database, available at <http://www.prepare-europe.eu/risedb>). All the data were extracted and entered into a database by two authors, and checked by the third, in order to ensure data integrity. Each identified study was indexed with a unique identification code (RISEdb ID). Some studies provided more than one data point (which corresponds to one marker - one study data entry), and were consequently represented with more than one data entry in our database. In all instances the data were extracted in the raw format as genotype counts, or were calculated from the reported percentages and total sample sizes where number of cases and controls was not available. When more cases or control groups were available in one study, we either pooled them or retained multiple data points from the same study, especially in case of having discovery and replication datasets. In addition, all data points were classified according to the underlying research question into three categories: studies that investigated infectious disease susceptibility, disease severity or disease mortality. In addition, we defined a subset of methodologically better studies, where controls were of known exposure to the pathogen.

**Risk of bias in individual studies.** We developed a novel score for purposes of risk of bias assessment. As the methodological design defines, a study may be prone to three main domains of errors and bias – confounding, selection bias and information bias. These three domains were the basis for the score development, which was entitled Confounding-Selection-Information bias score - CSI (Supplementary Table 2). The elements of the score were developed on the basis of several previous studies<sup>80</sup> and assessment scores, including Venice criteria for assessing cumulative epidemiologic evidence in genetic associations<sup>81</sup>, Newcastle-Ottawa case-control scale<sup>82</sup>, Cochrane risk of bias tool<sup>83</sup>, and previously established quality scores in genetic epidemiology<sup>84</sup>. Neither of these tools provided a good fit for the risk of bias

assessment in observational studies in genetic epidemiology, and we therefore took elements of previous scales and derived a new one. In line with Venice criteria, we scored all three domains in three grades of credibility: high - graded as A; intermediate - B or weak - C. This scheme provided estimates ranging from the best AAA to worst CCC score, which was applied to every data point in RISE database (and not for the meta-analysis results, as the Venice criteria is intended to). Since information bias can affect several aspects of the study, we also divided information bias risk domain in three separate sub-scores, relating to cases and controls selection and genotyping procedures, with the worst of all three defining the overall information bias risk. All CSI scoring was done independently by two authors, and all discrepancies were settled by an agreement or input from a third author (there were 313 or 14.2% data points with initially dissimilar grades, providing good agreement with kappa coefficient of 0.717 and standard error for kappa of 0.011).

As an additional quality control check, we re-calculated Hardy-Weinberg test for every control set included in the analysis, using a two tailed chi-square test or an exact test, when appropriate. Those data points that failed HWE test at the level of  $P < 0.05$  were downgraded to C in the CSI score (there were 225 out of 2,209, or 10.2% of all analysed data points that failed HWE), in order to reflect possible methodological limitations of failed HWE (which indicates both the risk for selection and/or information bias). Where cases from a single study were classified as moderate or severe, we did not determine Selection domain of CSI score (a total of 246 data points). In such instances we assigned 0 to selection bias risk, as there was no selection process involved, and second-stage analysis was performed

**Summary measures and synthesis of results.** In the first step of meta-analysis we calculated all odds ratios anew from raw data, in order to provide a common analytic framework and reduce possible differences in data analysis. In order to better understand the genetic architecture underlying gene-disease association, we performed analyses using four separate genetic models: allelic, dominant, recessive and heterozygote advantage. Model selection was based on commonly used allelic model, two additional models that assume different inheritance effects (dominant and recessive), followed by the heterozygote advantage, as previous studies reported such results<sup>85–87</sup>. All calculations were therefore based on allele counts, rather than exact cases and controls counts. We compared the results of different models, in an attempt to provide broader insight into genetic architecture.

Minor allele frequencies in Homo sapiens Short Variation (GRCh38) data set were obtained from Ensembl Biomart on 08.12.2014<sup>88</sup>. We then performed a series of meta-analyses for all SNPs where four or more data points were available for a single SNP. We used random effects meta-analysis, assuming differences in designs and ethnic composition of individual studies. We also calculated fixed-effects model, in order to compare the results. We used  $I^2$  statistics and corresponding confidence intervals as a measure of heterogeneity.

In order to account for multiple testing as a result of numerous meta-analyses and different genetic models being performed, we used Bayesian false-discovery probability (BFDP)<sup>72</sup>. It was calculated for nominally significant results only, with a threshold of 0.2, using an online free version of the BFDP calculator (<http://faculty.washington.edu/jonno/BFDP.xls>). In addition, we calculated BFDP using two prior probabilities, with medium/low prior level (0.05 to E-03), consistent with a candidate gene; and very low prior level (E-04 to E-06), consistent with a random SNP. The advantage of this approach over the commonly used corrections is that it is not dependent on the number of tests performed, as it relies on odds ratios and confidence intervals for calculation. In addition, it provides conceptual improvement over the similar methods, such as false-positive report probability (FPRP)<sup>89</sup>. Due to substantial requirements for appropriate use of publication bias analysis methods<sup>90</sup>, we only examined publication bias visually.

All analyses were performed in R, version 2.15.3<sup>91</sup>. Meta-analysis was performed with package meta 3.2–0 and metabin, forest and funnel functions<sup>92</sup>. Hardy-Weinberg equilibrium testing was performed in HardyWeinberg\_1.5.2 package<sup>93</sup>, with HWExactMat or HWChisqMat functions, based on the cell expected frequencies.

**Risk of bias across studies.** For purpose of meta-analysis credibility assessment, we used Venice criteria<sup>81</sup>, in all instances of nominally significant results. Study power was assessed on the basis of sample sizes (with sample sizes of up to a 1,000 graded as C, 1,001–10,000 graded as B, and over 10,001 graded as A). Heterogeneity was based on  $I^2$  statistics (graded as A in case of 0–25%, B in case of 26–50% and C in case of  $I^2$  being over 50% ), while the third score domain was fixed as C (weak credibility), due to very prevalent risk of bias in primary studies.

## References

1. Fauci, A. S. & Morens, D. M. The perpetual challenge of infectious diseases. *The New England journal of medicine* **366**, 454–461, doi: 10.1056/NEJMra1108296 (2012).
2. Fumagalli, M. *et al.* Genome-wide identification of susceptibility alleles for viral infections through a population genetics approach. *PLoS genetics* **6**, e1000849, doi: 10.1371/journal.pgen.1000849 (2010).
3. Morens, D. M., Folkers, G. K. & Fauci, A. S. Emerging infections: a perpetual challenge. *The Lancet Infectious diseases* **8**, 710–719, doi: 10.1016/s1473-3099(08)70256-1 (2008).
4. Sorensen, T. I., Nielsen, G. G., Andersen, P. K. & Teasdale, T. W. Genetic and environmental influences on premature death in adult adoptees. *The New England journal of medicine* **318**, 727–732, doi: 10.1056/nejm198803243181202 (1988).



5. Albright, F. S., Orlando, P., Pavia, A. T., Jackson, G. G. & Cannon Albright, L. A. Evidence for a heritable predisposition to death due to influenza. *Journal of Infectious Diseases* **197**, 18–24 (2008).
6. Cooke, G. S. & Hill, A. V. S. Genetics of susceptibility to human infectious disease. *Nature Reviews Genetics* **2**, 967–977 (2001).
7. Burgner, D., Jamieson, S. E. & Blackwell, J. M. Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *The Lancet Infectious diseases* **6**, 653–663, doi: 10.1016/s1473-3099(06)70601-6 (2006).
8. Rowell, J. L. *et al.* Trends in population-based studies of human genetics in infectious diseases. *PLoS ONE* **7**, e25431, doi: 10.1371/journal.pone.0025431 (2012).
9. Heitzeneder, S., Seidel, M., Forster-Waldl, E. & Heitger, A. Mannan-binding lectin deficiency - Good news, bad news, doesn't matter? *Clinical immunology (Orlando, Fla.)* **143**, 22–38, doi: 10.1016/j.clim.2011.11.002 (2012).
10. Chapman, S. J. & Hill, A. V. S. Human genetic susceptibility to infectious disease. *Nature Reviews Genetics* **13**, 175–188 (2012).
11. Colhoun, H. M., McKeigue, P. M. & Davey Smith, G. Problems of reporting genetic associations with complex outcomes. *Lancet* **361**, 865–872 (2003).
12. Affandi, J. S., Price, P. & Waterer, G. Can immunogenetics illuminate the diverse manifestations of respiratory infections? *Therapeutic advances in respiratory disease* **4**, 161–176, doi: 10.1177/1753465810371484 (2010).
13. Barry, C. E., 3rd. *et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature reviews Microbiology* **7**, 845–855, doi: 10.1038/nrmicro2236 (2009).
14. Ioannidis, J. P. *et al.* Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* **383**, 166–175, doi: 10.1016/s0140-6736(13)62227-8 (2014).
15. Hill, A. V. S. Aspects of genetic susceptibility to human infectious diseases. *Annual Review of Genetics* **40**, 469–486, (2006).
16. Thye, T. *et al.* Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nature genetics* **44**, 257–259, doi: 10.1038/ng.1080 (2012).
17. Thye, T. *et al.* Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nature genetics* **42**, 739–741, doi: 10.1038/ng.639 (2010).
18. Allen, E. K., Manichaikul, A. & Sale, M. M. Genetic contributors to otitis media: agnostic discovery approaches. *Current allergy and asthma reports* **14**, 411, doi: 10.1007/s11882-013-0411-7 (2014).
19. Casanova, J. L. & Abel, L. The genetic theory of infectious diseases: A brief history and selected illustrations. *Annual Review of Genomics and Human Genetics* **14**, 215–243 (2013).
20. Quintana-Murci, L., Alcais, A., Abel, L. & Casanova, J. L. Immunology in natura: clinical, epidemiological and evolutionary genetics of infectious diseases. *Nature immunology* **8**, 1165–1171, doi: 10.1038/ni1535 (2007).
21. Manry, J. & Quintana-Murci, L. A genome-wide perspective of human diversity and its implications in infectious disease. *Cold Spring Harbor Perspectives in Biology* **3**, a012450. doi: 10.1101/cshperspect.a012450 (2013).
22. Patterson, K. D. & Pyle, G. F. The geography and mortality of the 1918 influenza pandemic. *Bulletin of the history of medicine* **65**, 4–21 (1991).
23. Johnson, N. P. & Mueller, J. Updating the accounts: global mortality of the 1918–1920 “Spanish” influenza pandemic. *Bulletin of the history of medicine* **76**, 105–115 (2002).
24. Le Chevalier, F., Cascioferro, A., Majlessi, L., Herrmann, J. L. & Brosch, R. Mycobacterium tuberculosis evolutionary pathogenesis and its putative impact on drug development. *Future microbiology* **9**, 969–985, doi: 10.2217/fmb.14.70 (2014).
25. Gagneux, S. Host-pathogen coevolution in human tuberculosis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **367**, 850–859, doi: 10.1098/rstb.2011.0316 (2012).
26. Hershberg, R. *et al.* High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS biology* **6**, e311, doi: 10.1371/journal.pbio.0060311 (2008).
27. World Health Organization. *Global tuberculosis report 2014*, <http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809\_eng.pdf?ua=1> (2014) (Date of access: 11/12/2014).
28. Walker, C. L. *et al.* Global burden of childhood pneumonia and diarrhoea. *Lancet* **381**, 1405–1416, doi: 10.1016/s0140-6736(13)60222-6 (2013).
29. Nair, H. *et al.* Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *Lancet* **375**, 1545–1555, doi: 10.1016/s0140-6736(10)60206-1 (2010).
30. Fouchier, R. A. *et al.* Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* **423**, 240, doi: 10.1038/423240a (2003).
31. Poon, L. L., Guan, Y., Nicholls, J. M., Yuen, K. Y. & Peiris, J. S. The aetiology, origins, and diagnosis of severe acute respiratory syndrome. *The Lancet. Infectious diseases* **4**, 663–671, doi: 10.1016/s1473-3099(04)01172-7 (2004).
32. Alcais, A. *et al.* Life-threatening infectious diseases of childhood: Single-gene inborn errors of immunity? *Ann NY Acad Sci* **1214**, 18–33 (2010).
33. Alcais, A., Abel, L. & Casanova, J. L. Human genetics of infectious diseases: Between proof of principle and paradigm. *Journal of Clinical Investigation* **119**, 2506–2514 (2009).
34. Choi, P. & Reiser, H. IL-4: role in disease and regulation of production. *Clinical and experimental immunology* **113**, 317–319 (1998).
35. Hershey, G. K., Friedrich, M. F., Esswein, L. A., Thomas, M. L. & Chatila, T. A. The association of atopy with a gain-of-function mutation in the alpha subunit of the interleukin-4 receptor. *The New England journal of medicine* **337**, 1720–1725, doi: 10.1056/nejm199712113372403 (1997).
36. Newton, C. *et al.* Induction of interleukin-4 (IL-4) by Legionella pneumophila infection in BALB/c mice and regulation of tumor necrosis factor alpha, IL-6, and IL-1beta. *Infection and immunity* **68**, 5234–5240 (2000).
37. Van Bergenhenegouwen, J. *et al.* TLR2 & Co: a critical analysis of the complex interactions between TLR2 and coreceptors. *Journal of leukocyte biology* **94**, 885–902, doi: 10.1189/jlb.0113003 (2013).
38. Ogus, A. C. *et al.* The Arg753Gln polymorphism of the human toll-like receptor 2 gene in tuberculosis disease. *The European respiratory journal* **23**, 219–223 (2004).
39. Saraav, I., Singh, S. & Sharma, S. Outcome of Mycobacterium tuberculosis and Toll-like receptor interaction: immune response or immune evasion? *Immunology and cell biology* **92**, 741–746, doi: 10.1038/icb.2014.52 (2014).
40. Underhill, D. M., Ozinsky, A., Smith, K. D. & Aderem, A. Toll-like receptor-2 mediates mycobacteria-induced proinflammatory signaling in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 14459–14463 (1999).
41. Oliveira-Nascimento, L., Massari, P. & Wetzler, L. The role of TLR2 in infection and immunity. *Front Immunol* **3**, 79, doi: 10.3389/fimmu.2012.00079 (2012).
42. McBride, A., Bhatt, K. & Salgame, P. Development of a secondary immune response to Mycobacterium tuberculosis is independent of Toll-like receptor 2. *Infection and immunity* **79**, 1118–1123, doi: 10.1128/iai.01076-10 (2011).
43. Xia, M. & Sui, Z. Recent developments in CCR2 antagonists. *Expert opinion on therapeutic patents* **19**, 295–303, doi: 10.1517/13543770902755129 (2009).
44. Paiva, C. N. *et al.* CCL2/MCP-1 controls parasite burden, cell infiltration, and mononuclear activation during acute Trypanosoma cruzi infection. *Journal of leukocyte biology* **86**, 1239–1246, doi: 10.1189/jlb.0309187 (2009).

45. Hussain, R., Ansari, A., Talat, N., Hasan, Z. & Dawood, G. CCL2/MCP-1 genotype-phenotype relationship in latent tuberculosis infection. *PLoS One* **6**, e25803, doi: 10.1371/journal.pone.0025803 (2011).
46. Deshmane, S. L., Kremlev, S., Amini, S. & Sawaya, B. E. Monocyte chemoattractant protein-1 (MCP-1): an overview. *Journal of interferon & cytokine research: the official journal of the International Society for Interferon and Cytokine Research* **29**, 313–326, doi: 10.1089/jir.2008.0027 (2009).
47. Siveke, J. T. & Hamann, A. T helper 1 and T helper 2 cells respond differentially to chemokines. *Journal of immunology* (Baltimore, Md.: 1950) **160**, 550–554 (1998).
48. Singh, B., Chitra, J. & Selvaraj, P. CCL2, CCL3 and CCL4 gene polymorphisms in pulmonary tuberculosis patients of South India. *International journal of immunogenetics* **41**, 98–104, doi: 10.1111/iji.12085 (2014).
49. Mishra, G., Poojary, S. S., Raj, P. & Tiwari, P. K. Genetic polymorphisms of CCL2, CCL5, CCR2 and CCR5 genes in Sahariya tribe of North Central India: an association study with pulmonary tuberculosis. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **12**, 1120–1127, doi: 10.1016/j.meegid.2012.03.018 (2012).
50. Feng, W. X. *et al.* CCL2-2518 (A/G) polymorphisms and tuberculosis susceptibility: a meta-analysis. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease* **16**, 150–156, doi: 10.5588/ijtld.11.0205 (2012).
51. Greenwood, C. M. *et al.* Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. *American journal of human genetics* **67**, 405–416, doi: 10.1086/303012 (2000).
52. Cervino, A. C., Lakiss, S., Sow, O. & Hill, A. V. Allelic association between the NRAMP1 gene and susceptibility to tuberculosis in Guinea-Conakry. *Annals of human genetics* **64**, 507–512 (2000).
53. El Baghdadi, J. *et al.* Variants of the human NRAMP1 gene and susceptibility to tuberculosis in Morocco. *International Journal of Tuberculosis and Lung Disease* **7**, 599–602 (2003).
54. Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious disease in human populations. *Nature reviews. Genetics* **15**, 379–393, doi: 10.1038/nrg3734 (2014).
55. Chimusa, E. R. *et al.* Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. *Human molecular genetics* **23**, 796–809, doi: 10.1093/hmg/ddt462 (2014).
56. Curtis, J. *et al.* Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. *Nat Genet* **47**, 523–527, doi: 10.1038/ng.3248 (2015).
57. Casanova, J. L. & Abel, L. Human genetics of infectious diseases: a unified theory. *The EMBO journal* **26**, 915–922, doi: 10.1038/sj.emboj.7601558 (2007).
58. Picard, C., Casanova, J. L. & Abel, L. Mendelian traits that confer predisposition or resistance to specific infections in humans. *Current opinion in immunology* **18**, 383–390, doi: 10.1016/j.coi.2006.05.005 (2006).
59. Brouwer, M. C. *et al.* Host genetic susceptibility to pneumococcal and meningococcal disease: a systematic review and meta-analysis. *The Lancet Infectious Diseases* **9**, 31–44 (2009).
60. Szymczak, W. A. & Deepe, G. S., Jr. The CCL7-CCL2-CCR2 axis regulates IL-4 production in lungs and fungal immunity. *Journal of immunology* (Baltimore, Md.: 1950) **183**, 1964–1974, doi: 10.4049/jimmunol.0901316 (2009).
61. Ladinig, A. *et al.* Cytokine profiles in pregnant gilts experimentally infected with porcine reproductive and respiratory syndrome virus and relationships with viral load and fetal outcome. *Veterinary research* **45**, 113, doi: 10.1186/s13567-014-0113-8 (2014).
62. Walch, L., Massade, L., Dufilho, M., Brunet, A. & Rendu, F. Pro-atherogenic effect of interleukin-4 in endothelial cells: modulation of oxidative stress, nitric oxide and monocyte chemoattractant protein-1 expression. *Atherosclerosis* **187**, 285–291, doi: 10.1016/j.atherosclerosis.2005.09.016 (2006).
63. Wang, H., Nemoto-Sasaki, Y., Kondo, T., Akiyama, M. & Mukaida, N. Potential involvement of monocyte chemoattractant protein (MCP)-1/CCL2 in IL-4-mediated tumor immunity through inducing dendritic cell migration into the draining lymph nodes. *International immunopharmacology* **3**, 627–642, doi: 10.1016/s1567-5769(02)00251-5 (2003).
64. Fiorina, P. *et al.* Phenotypic and functional differences between wild-type and CCR2<sup>-/-</sup> dendritic cells: implications for islet transplantation. *Transplantation* **85**, 1030–1038, doi: 10.1097/TP.0b013e31816843a0 (2008).
65. Asai, H. *et al.* Co-introduced functional CCR2 potentiates *in vivo* anti-lung cancer functionality mediated by T cells double gene-modified to express WT1-specific T-cell receptor. *PLoS One* **8**, e56820, doi: 10.1371/journal.pone.0056820 (2013).
66. de Martel, C. *et al.* Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The Lancet. Oncology* **13**, 607–615, doi: 10.1016/s1470-2045(12)70137-7 (2012).
67. Abel, L., El-Baghdadi, J., Bousfiha, A. A., Casanova, J. L. & Schurr, E. Human genetics of tuberculosis: a long and winding road. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130428, doi: 10.1098/rstb.2013.0428 (2014).
68. van Manen, D., van 't Wout, A. B. & Schuitemaker, H. Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. *Retrovirology* **9**, 70, doi: 10.1186/1742-4690-9-70 (2012).
69. Aouizerat, B. E., Pearce, C. L. & Miaskowski, C. The search for host genetic factors of HIV/AIDS pathogenesis in the post-genome era: progress to date and new avenues for discovery. *Current HIV/AIDS reports* **8**, 38–44, doi: 10.1007/s11904-010-0065-1 (2011).
70. An, P. & Winkler, C. A. Host genes associated with HIV/AIDS: advances in gene discovery. *Trends in genetics: TIG* **26**, 119–131, doi: 10.1016/j.tig.2010.01.002 (2010).
71. Rietveld, C. A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13790–13794, doi: 10.1073/pnas.1404623111 (2014).
72. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *American journal of human genetics* **81**, 208–227, doi: 10.1086/519024 (2007).
73. Polasek, O. Future of biobanks - bigger, longer, and more dimensional. *Croatian medical journal* **54**, 496–500 (2013).
74. Little, J. *et al.* STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *PLoS medicine* **6**, e22, doi: 10.1371/journal.pmed.1000022 (2009).
75. Gallo, V. *et al.* STrengthening the Reporting of Observational studies in Epidemiology: Molecular Epidemiology STROBE-ME. An extension of the STROBE statement. *Journal of epidemiology and community health* **66**, 844–854, doi: 10.1136/jech-2011-200318 (2012).
76. Hollenbach, J. A. *et al.* A community standard for immunogenomic data reporting and analysis: proposal for a STrengthening the REporting of Immunogenomic Studies statement. *Tissue antigens* **78**, 333–344, doi: 10.1111/j.1399-0039.2011.01777.x (2011).
77. Kodaman, N., Sobota, R. S., Mera, R., Schneider, B. G. & Williams, S. M. Disrupted human-pathogen co-evolution: a model for disease. *Frontiers in genetics* **5**, 290, doi: 10.3389/fgene.2014.00290 (2014).
78. Marsland, B. J. & Gollwitzer, E. S. Host-microorganism interactions in lung diseases. *Nature reviews. Immunology* **14**, 827–835, doi: 10.1038/nri3769 (2014).
79. Schmidlin, P. R. *et al.* Shared microbiome in gums and the lung in an outpatient population. *The Journal of infection*, doi: 10.1016/j.jinf.2014.10.005 (2014).

80. Sahoo, G. S., Little, J. & Higgins, J. P. Systematic reviews of genetic association studies. *Human Genome Epidemiology Network. PLoS medicine* **6**, e28, doi: 10.1371/journal.pmed.1000028 (2009).
81. Ioannidis, J. P. *et al.* Assessment of cumulative evidence on genetic associations: interim guidelines. *International journal of epidemiology* **37**, 120–132, doi: 10.1093/ije/dym159 (2008).
82. Wells G *et al.* *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*, <[http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp)> (2013) (Date of access: 18/11/2014).
83. Higgins, J. P. T. & S. G. *Cochrane handbook for systematic reviews of interventions version 5.1.0* <<http://www.cochrane-handbook.org/>> (2011) (Date of access: 18/11/2014).
84. Bogardus, S. T., Jr., Concato, J. & Feinstein, A. R. Clinical epidemiological quality in molecular genetic research: the need for methodological standards. *Jama* **281**, 1919–1926 (1999).
85. Chan, V. S. F. *et al.* Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. *Nature Genetics* **38**, 38–46 (2006).
86. Prugnolle, F. *et al.* Pathogen-driven selection and worldwide HLA class I diversity. *Current biology: CB* **15**, 1022–1027, doi: 10.1016/j.cub.2005.04.050 (2005).
87. Fumagalli, M. *et al.* Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *The Journal of experimental medicine* **206**, 1395–1408, doi: 10.1084/jem.20082779 (2009).
88. Kinsella, R. J. *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* **2011**, doi: 10.1093/database/bar030 (2011).
89. Lucke, J. F. A critique of the false-positive report probability. *Genetic epidemiology* **33**, 145–150, doi: 10.1002/gepi.20363 (2009).
90. Ioannidis, J. P. & Trikalinos, T. A. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *CMAJ: Canadian Medical Association journal* **176**, 1091–1096, doi: 10.1503/cmaj.060410 (2007).
91. R Core Team. *R: A language and environment for statistical computing*, <<http://www.R-project.org/>> (2013) (Date of access: 18/11/2014).
92. Schwarzer, G. *meta: Meta-Analysis with R*, <<http://CRAN.R-project.org/package=meta>> (2014) (Date of access: 18/11/2014).
93. Graffelman, J. *HardyWeinberg: Graphical tests for Hardy-Weinberg equilibrium. R package version 1.5.2*, <<http://CRAN.R-project.org/package=HardyWeinberg>> (2013) (Date of access: 18/11/2014).

## Acknowledgements

We would like to apologize to all authors whose studies we could not cite in this paper due to space limitations. We would also like to thank to all authors who provided additional information on their studies or provided raw data for purposes of this meta-analysis: Eileen Hoal and Marlo Möller (Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa; RISE0456), Riny Janssen and Hennie Hodemaekers (Center for Health Protection, National Institute for Public Health and the Environment, Bilthoven, the Netherlands; RISE0599), Meghan Baker and Megan Murray (Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; RISE0640) and Antony Payton (Centre for Integrated Genomic Medical Research, The University of Manchester, Manchester, UK) and Enitan D Carrol (Institute of Infection and Global Health, University of Liverpool, Liverpool, UK; RISE0880). The study was financed by the European Commission FP7 project PREPARE (602525) and Croatian Science Foundation grants 8875 and 8445.

## Author Contributions

O.P. and I.P. conceived the study idea; H.C., E.T., I.K., I.R., K.B. and M.J. further developed the idea and developed the current study scope; I.K., M.K., E.T. and O.P. developed and critically assessed the search strategy; I.P., A.G. and O.P. performed the search and extracted the data; I.P. and O.P. performed the data analysis; I.P., A.G. and O.P. drafted the manuscript; all authors approved the final manuscript version and are held accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Patarčić, I. *et al.* The role of host genetic factors in respiratory tract infectious diseases: systematic review, meta-analyses and field synopsis. *Sci. Rep.* **5**, 16119; doi: 10.1038/srep16119 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>