CLINICAL SCIENCE

Lifetime risk of rheumatoid arthritis-associated interstitial lung disease in MUC5B mutation carriers

Antti Palomäki, Jukka Koskela, Kari K Eklund, Matti Pirinen, FinnGen, Samuli Ripatti, Tarja Laitinen, Nina Mars

ABSTRACT

Objectives To estimate lifetime risk of developing rheumatoid arthritis-associated interstitial lung disease (RA-ILD) with respect to the strongest known risk factor for pulmonary fibrosis, a MUC5B promoter variant.

Methods FinnGen is a collection of epidemiological cohorts and hospital biobank samples, integrating genetic data with up to 50 years of follow-up within nationwide registries in Finland. Patients with RA and ILD were identified from the Finnish national hospital discharge, medication reimbursement and cause-of-death registries. We estimated lifetime risks of ILD by age 80 with respect to the common variant rs35705950, a MUC5B promoter variant.

Results Out of 293,972 individuals, 1965 (0.7%) developed ILD by age 80. Among all individuals in the dataset, MUC5B increased the risk of ILD with a HR of 2.44 (95% CI: 2.22 to 2.68). Out of 6869 patients diagnosed with RA, 247 (3.6%) developed ILD. In patients with RA, MUC5B was a strong risk factor of ILD with a HR similar to the full dataset (HR: 2.27, 95% CI: 1.75 to 2.95). In patients with RA, lifetime risks of ILD were 16.8% (95% CI: 13.1% to 20.2%) for MUC5B carriers and 6.1% (95% CI: 5.0% to 7.2%) for MUC5B non-carriers. The difference between risks started to emerge at age 65, with a higher risk among men.

Conclusion Our findings provide estimates of lifetime risk of RA-ILD based on MUC5B mutation carrier status, demonstrating the potential of genomics for risk stratification of RA-ILD.

INTRODUCTION

Interstitial lung disease (ILD) is one of the most common extra-articular manifestations of rheumatoid arthritis (RA). The cumulative risk of developing clinical ILD during the RA disease course has varied in different studies, ranging from 5.0% to 7.7% in long-term follow-up studies of RA cohorts to up to 10% in a study using death records. Even higher estimates for subclinical radiographic findings consistent with ILD have been observed in patients with RA, ranging from 19% to 33%. Although the RA-ILD course can vary, the disease is associated with significantly increased mortality compared with patients with RA without ILD.

Clinical risk factors for RA-ILD include older age, male gender, tobacco smoking, high levels of anticitrullinated protein antibodies and disease activity. The strongest known genetic risk factor for idiopathic pulmonary fibrosis (IPF) is the common variant rs35705950, a promoter variant near the MUC5B gene. A recent case-control study has demonstrated that the MUC5B promoter variation is associated with an increased risk of ILD among patients with RA. The aim of this study was to evaluate the lifetime risk of ILD in patients with RA, comparing the risk to the population, and estimate how the MUC5B promoter variant modifies these risks in the real-world setting.

METHODS

FinnGen is a collection of prospective epidemiological and disease-based cohorts, and hospital biobank samples. The unique personal identification number links the genotypes to multiple nationwide registries, and cases were identified through the national hospital discharge registry (starting from 1968) including both inpatient and outpatient data, the national death registry (1969–) and the medication reimbursement registry (1964–).

RA was defined as patients having medication reimbursement for inflammatory rheumatic diseases (code 202), with an additional requirement of two contacts with the International Classification of Diseases, Tenth Revision (ICD-10) codes for idiopathic pulmonary fibrosis (IPF) is the common variant rs35705950, a promoter variant near the MUC5B gene.
beginning with M05 (seropositive RA) or M06 (seronegative RA). In our recent validation study of RA diagnoses in Finnish biobank patients (unpublished), this combination resulted in a positive predictive value of 0.87 compared with chart review. Negative predictive value for any RA diagnosis was 1.0. Those without RA who had other inflammatory rheumatic diseases or inflammatory bowel disease were excluded.

ILD cases were identified with J84, M05.1/J99.0 (ICD-10), 515, 516 (ICD-9) or 484.99 or 517.01 (ICD-8) with following criteria: (1) the first and only record in the death registry or (2) after the initial diagnosis, a second contact (or death due to ILD) was required within 5 years, that is, we excluded individuals with no further healthcare contacts with ILD within 5 years. No exclusions were made based on temporality of RA and ILD. For both RA and ILD, age at onset was defined as age at first registered diagnosis.

For MUC5B (mucin 5B, oligomeric mucus/gel-forming), we studied carriers of the minor allele for the promoter variant rs35705950 (G>T) with minor allele frequency 0.1 (no enrichment compared with non-Finnish Europeans12) and mean INFO 0.948 indicating high imputation quality. Individuals homozygous for the variant were analysed jointly with the heterozygotes.

Start of follow-up was set at birth, with follow-up ending at the first record of the endpoint of interest, death, or at the end of follow-up on 31 December 2019, whichever came first. Using the Cox proportional hazards model, we estimated adjusted HRs and 95% CIs (CI). With age as time scale, all regression models were stratified by sex, adjusted for 10 principal components of ancestry, FinnGen genotyping array and cohort. We report cumulative incidences with 95% CIs by age 80. We used R V3.6.3. Detailed information on genotyping, disease definitions and analyses are provided in online supplemental methods.

**Patient and public involvement**

This study was carried out without direct patient and public involvement.

**RESULTS**

Among 293,972 individuals (mean age at the end of follow-up: 59.8, SD: 17.3, 56.4% women), we identified 1,963 patients (1,172 men, 793 women) diagnosed with ILD by end of follow-up. Out of 6,869 patients with RA (mean age at onset: 49.4, SD: 14.9, 71.1% women), 247 (3.6%) had been diagnosed with ILD. Out of these 247 individuals, 20 (8.1%) had been diagnosed with ILD >1 year before the earliest record of RA, 36 (14.6%) within a year prior to or after the earliest record of RA and 191 (77.3%) >1 year after. Out of patients without RA, 19.3% were MUC5B carriers, and out of patients with RA, 20.9%. Among all individuals in the dataset, the MUC5B promoter variant rs35705950 was associated with ILD with a HR of 2.44 (2.22–2.68, p=3.87×10⁻⁷), and among patients with RA, with a HR of 2.27 (1.75–2.95, p=8.15×10⁻⁸). In a formal test for interaction by introducing an interaction term in the regression model, we found no evidence of an interaction between MUC5B and RA (p=0.16). These interaction tests indicate that the effect of MUC5B is similar in the population and in patients with RA.

Next, we quantified the lifetime risk of ILD for four groups: (1) MUC5B non-carriers in the population, (2) MUC5B carriers in the population, (3) MUC5B non-carriers with RA and (4) MUC5B carriers with RA (figure 1, table 1). The corresponding lifetime risks were (1) 1.5% (95% CI: 1.3% to 1.6%), (2) 4.4% (95% CI: 4.1%–4.8%), 3) 6.1% (95% CI: 5.0%–7.2%) and (4) 16.8% (95% CI: 13.1%–20.2%). In sex-specific analyses, the lifetime risk was 20.9% (95% CI: 14.1%–27.1%) in men with RA who are MUC5B carriers, and the corresponding lifetime risk in women was 14.5% (95% CI: 10.2%–18.6%). Accounting for competing risks (non-ILD causes of death) yielded marginally lower estimates of lifetime risks, particularly in men (online supplemental table 1).

Lastly, we observed an association between MUC5B and risk of RA (HR: 1.10, 1.04–1.17, p=0.0009), with a somewhat larger association in men (HR: 1.17, 1.05–1.30, p=0.005) than in women (HR: 1.08, 1.01–1.16, p=0.04). The effects remained similar when excluding all 1,172 men with ILD (HR: 1.13 in men, 1.01–1.26, p=0.03) and all 793 women with ILD (HR: 1.05 in women, 0.98–1.13, p=0.19). This observation was replicated in UK Biobank (1911 RA cases; see online supplemental methods for details) with a HR of 1.15 (1.03–1.28, p=0.01).

**DISCUSSION**

In this large observational cohort study, we demonstrate that a combination of RA and MUC5B variation confers a 10-fold elevated risk of ILD compared with the population. Every sixth

---

**Figure 1** Lifetime risk of interstitial lung disease in the population for MUC5B carriers and non-carriers with respect to diagnosis of RA. The risks are shown for men and women both combined and individually. MUC5B=carriers of the minor allele for the promoter variant rs35705950. Sample size: 293,972 (128,233 men and 165,739 women), RA, rheumatoid arthritis.
patient with RA carrying the MUC5B risk allele was diagnosed with ILD by age 80, and the risk rapidly increased after age 65. A case–control study by Juge and colleagues recently demonstrated enrichment of MUC5B carriers in patients with RA-ILD, with supporting evidence from gene expression in lung parenchyma and high-resolution imaging. Using large-scale biobank data, we now show how this finding translates to lifetime risks and demonstrate the potential of genomics for risk stratification of RA-ILD and early identification of patients.

Prevalence of RA-ILD shows high variability in the literature depending on the population, diagnostic methods and disease definitions used. Our lifetime risks compare well with previous estimates of clinically significant disease, reported to occur in up to 5%–10% of patients with RA. We show that the effect of MUC5B is similar in the population and in patients with RA, but as both MUC5B and RA are important risk factors of ILD, patients with RA who are MUC5B carriers are at a much higher risk of ILD than MUC5B carriers without RA. The common variant rs35705950 in the MUC5B promoter is strongly associated with upregulation of MUC5B expression in the lungs, and the general association between the variant and ILD has been widely replicated. In addition, evidence from fine-mapping indicates that rs35705950 might be a causal variant: Bayesian fine-mapping analyses of genome-wide association study (GWAS) results can be used for defining variant sets (credible sets), that with high probability contain one or several causal variants. Several sources report rs35705950 as the only variant in the credible sets for the locus on ILD and IPF.

We were unable to account for some important risk factors, such as smoking and disease activity, and did not consider other common or rare genetic risk factors, all of which are likely to further contribute to the risk. We did not have information about histological or radiological patterns of ILD. The study was limited to individuals of European ancestry, but MUC5B may be a relevant risk factor also in other populations, although many have allele frequencies that are much lower. With a prevalence of 2.3% for RA and 0.7% for ILD, our sample is slightly enriched in cases, which may affect our estimates. Although ILD was identified through healthcare registries, recurring healthcare encounters were required to reduce the proportion of false positives in our study, and the long-term risk of ILD in patients with RA was in line with previous studies. Patients with RA might be exposed to more chest imaging as part of their standard care and due to increased awareness for the risk of ILD particularly during recent years, which could overestimate the risk difference between patients with and without RA. We also observed a modest association between MUC5B and RA, which was replicated in UK Biobank. This association was not detected in a previous study with a smaller sample size by Juge and colleagues. This tentative finding, which was clearer in men, requires further replication with consideration of other important risk factors, such as smoking. As the effects remained similar when excluding all patients with ILD, we propose that the temporal sequence of ILD and RA is unlikely to impact the association.

In conclusion, the MUC5B promoter variant is a common risk factor for ILD in patients with RA and confers a significantly elevated lifetime risk of ILD. This study demonstrates the potential of genomics for risk stratification of RA-ILD and highlights the importance of genetic predisposition on the development of RA-ILD. Studies are needed to further investigate the interaction of clinical and genetic risk factors in the development of RA-ILD, and the impact of MUC5B on outcomes of RA-ILD.

### Table 1  Data characteristics, and effect of RA and MUC5B on risk of ILD

<table>
<thead>
<tr>
<th></th>
<th>Individuals without RA</th>
<th>Individuals with RA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-carriers of MUC5B</td>
<td>Carriers of MUC5B</td>
</tr>
<tr>
<td></td>
<td>promoter variant</td>
<td>promoter variant</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>231 860</td>
<td>55 243</td>
</tr>
<tr>
<td><strong>ILD cases</strong></td>
<td>1007</td>
<td>71</td>
</tr>
<tr>
<td><strong>ILD cases in men/women</strong></td>
<td>600/407</td>
<td>461/250</td>
</tr>
<tr>
<td><strong>Age at ILD onset, men/women (mean (SD))</strong></td>
<td>66.9 (10.4)/63.0 (13.3)</td>
<td>67.9 (8.3)/65.3 (11.1)</td>
</tr>
<tr>
<td><strong>Risk of ILD in women and men</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime risk, % (95% CI)</td>
<td>1.5 (1.3–1.6)</td>
<td>4.4 (4.1–4.8)</td>
</tr>
<tr>
<td>HR (95% CI)</td>
<td>Reference</td>
<td>2.49 (2.25–2.75)</td>
</tr>
<tr>
<td>P value –</td>
<td>1.24×10^{-71}</td>
<td>2.67×10^{-74}</td>
</tr>
<tr>
<td><strong>Risk of ILD in men</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime risk, % (95% CI)</td>
<td>1.7 (1.6–1.9)</td>
<td>5.6 (5.1–6.2)</td>
</tr>
<tr>
<td>HR (95% CI)</td>
<td>Reference</td>
<td>2.63 (2.31–2.98)</td>
</tr>
<tr>
<td>P value –</td>
<td>2.04×10^{-50}</td>
<td>6.81×10^{-43}</td>
</tr>
<tr>
<td><strong>Risk of ILD in women</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime risk, % (95% CI)</td>
<td>1.1 (1.0–1.3)</td>
<td>3.1 (2.6–3.5)</td>
</tr>
<tr>
<td>HR (95% CI)</td>
<td>Reference</td>
<td>2.26 (1.92–2.66)</td>
</tr>
<tr>
<td>P value –</td>
<td>1.46×10^{-22}</td>
<td>1.31×10^{-24}</td>
</tr>
</tbody>
</table>

ILD, interstitial lung disease; RA, rheumatoid arthritis.
Rheumatoid arthritis

United States); Yufeng Huang (Biogen, Cambridge, MA, United States); Joseph
Maranville (Celgene, Summit, NJ, United States); Bertil Myers Squibb, New York, NY, United
States); Shameek Biswas (Celgene, Summit, NJ, United States); Bertil Myers Squibb, New York, NY, United
States); Elmutaz Mohammed (Celgene, Summit, NJ, United States); Bertil Myers Squibb, New York, NY, United
States); Eeva Kangasniemi (Finnish Clinical Biobank Tampere, Tampere, Finland); Katri Pylkäs (University of Oulu, Oulu, Finland); Marta Kalio (Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland); Sirpa Soini (THL Biobank / The National Institute of Health and Welfare Helsinki, Finland); Jukka Partanen (Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland); Eero Punkkia (Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki, Finland); Raisa Serpi (Northern Finland Biobank Board of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland); Johanna Mäkelä (Finnish Clinical Biobank of Tampere / University of Turku / Hospital District of Southwest Finland, Turku, Finland); Siri Ranta (THL Biobank / The National Institute of Health and Welfare Helsinki, Finland); Teijo Kuopio (Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland); FinGen Teams: Administration: Anu Jalanko (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Hui-Yi Shen (Biomarkers for Rheumatoid Arthritis in Finland, HIFLE, University of Helsinki, Finland); Risto Kajanne (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Mervi Aavikko (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Analysis: Mitja Kuki (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Olli Harma (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Mark Daly (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Juha Kajakainen (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Arto Lehto (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Arto Mehtonen (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Samuli Ripatti (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland). 

He Xia (Sanofi, Paris, France); Beryl abuenca (University of Tampere, Tampere, Finland); Mervi Aavikko (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Susanna Lemmelä (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Tuomo Kiskinen (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Wei Zhou (Broad Institute, Cambridge, MA, United States); Mervi Aavikko (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Matti Peura (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Communication: Mari Kaunisto (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Data Management and IT Infrastructure: Elina Kälpläinen (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Tuuli Sistonen (THL Biobank / The National Institute of Health and Welfare Helsinki, Finland); Genotyping: Kati Donner (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Ingvar Järveläinen (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland). Sample Collection Coordination: Anu Loukola Helsinki Biobank / University of Helsinki, Finland). District and Hospital District of Helsinki and Uusimaa, Helsinki, Sample Logistics: Päivi Laito (THL Biobank / The National Institute of Health and Welfare Helsinki, Finland); Eija Laakkonen (Institute for Molecular Medicine Finland, HIFLE, University of Helsinki, Finland); Laila Lahdenmäki (THL Biobank / The National Institute of Health and Welfare Helsinki, Finland).
Rheumatoid arthritis

officer: Tero Jyrhämä (Institute for Molecular Medicine Finland, HIILIFE, University of Helsinki, Finland), FinnBB - Finnish biobank cooperative: Marco Hautoalahti, Laura Mustaniemi, Mirka Koiranen, Samuli Mattila, Tom Southerton.

Contributors AnttiP and NM conceived and designed the study. NM carried out the statistical and computational analyses with advice from SR and AnttiP. All authors provided critical inputs to interpretation of the data. The manuscript was written and revised by NM and AnttiP with comments from all of the coauthors. All coauthors have approved the final version of the manuscript.

Funding This work was supported by Turku University Hospital Research Foundation (to AnttiP), the Sigrid Juselius Foundation (to SR and AarnoP), Foundation and the Horizon 2020 Research and Innovation Programme [grant number 667301 (COSYN) to AarnoP], University of Helsinki HIILIFE Fellow grants 2017-2020 (to SR); Academy of Finland Center of Excellence in Complex Disease Genetics [grant number 312062 to SR, 312074 to AarnoP]; Academy of Finland (grant number 331671 to NM, grant number 285380 to SR, 334229 to JK and TL). This study is part of the FinnGen project funded by Business Finland (grant numbers HUS 4685/31/2016 and UH 4386/31/2016) and 12 pharma companies: AbbVie Inc, AstraZeneca UK Ltd, Biogen MA Inc, Celgene Corporation, Celgene International II Sàrl, Genentech Inc, Merck Sharp & Dohme Corp, Pfizer Inc, GlaxoSmithKline Intellectual Property Development Ltd, Sanofi US Services Inc, Maze Therapeutics Inc, Janssen Biotech Inc and Novartis AG.

Competing interests AnttiP has received a consulting fee from Pfizer and Abbvie, a lecture fee from Pfizer and Sanofi, and support for attending a meeting from Bristol-Myers-Squibb and Novartis. JK has received a consulting fee and a lecture fee from Pfizer. TL has received consulting fee and lecture fee from Pfizer. The other authors declare no competing interests.

Patient consent for publication Not required.


Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The Finnish biobank data can be accessed through the Fingene® services (https://site.fingene.fi/en/) managed by FINBB. The remaining data are available within the article, supplemental information or available from the authors upon request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, and build upon this work for any purpose, provided the original work is properly cited, a link to the license is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs Antti Palomäki http://orcid.org/0000-0002-8835-8116 Nina Mars http://orcid.org/0000-0002-7259-5993

REFERENCES


