Gut microbiome and atrial fibrillation—results from a large population-based study

Joonatan Palmu,a,b,x Christin S. Börschel,c,d,x Alfredo Ortega-Alonso,a,e,f Lajos Markó,g,h,i,j Michael Inouye,k,l Pekka Jousilahti,a Rodolfo A. Salido,t Karenina Sanders,1 Caisonia Brennan,1 Gregory C. Humphrey,1 Jon G. Sanders,t Veirka Salomaa,a Aki S. Havulinna,4,5 Sofia K Forslund,5,6 Rob Knight,s,t,u Leo Lahti,v Teemu Niiranen,a,b,x and Renate B. Schnabelc,d,x,

* Corresponding author. Department of Cardiology, University Heart and Vascular Centre Hamburg-Eppendorf, Martinistr. 52, 20246, Hamburg, Germany.
E-mail address: r.schnabel@uke.de (R.B. Schnabel).
Both authors contributed equally.

Summary

Background Atrial fibrillation (AF) is an important heart rhythm disorder in aging populations. The gut microbiome composition has been previously related to cardiovascular disease risk factors. Whether the gut microbial profile is also associated with the risk of AF remains unknown.

Methods We examined the associations of prevalent and incident AF with gut microbiota in the FINRISK 2002 study, a random population sample of 6763 individuals. We replicated our findings in an independent case–control cohort of 138 individuals in Hamburg, Germany.

Findings Multivariable-adjusted regression models revealed that prevalent AF (N = 116) was associated with nine microbial genera. Incident AF (N = 539) over a median follow-up of 15 years was associated with eight microbial genera with false discovery rate (FDR)-corrected P < 0.05. Both prevalent and incident AF were associated with the genera Enorma and Bifidobacterium (FDR-corrected P < 0.001). AF was not significantly associated with bacterial diversity measures. Seventy-five percent of top genera (Enorma, Paraprevotella, Odoribacter, Collinsella, Barnesiella, Alistipes) in Cox regression analyses showed a consistent direction of shifted abundance in an independent AF case–control cohort that was used for replication.

Abbreviations: AF, Atrial fibrillation; BMI, Body mass index; FDR, False discovery rate; LPS, Lipopolysaccharides; PCoA, Principal coordinates; SCFA, Short-chain fatty acid; SD, Standard deviation

www.thelancet.com Vol 91 May, 2023 1
Interpretation Our findings establish the basis for the use of microbiome profiles in AF risk prediction. However, extensive research is still warranted before microbiome sequencing can be used for prevention and targeted treatment of AF.

Funding This study was funded by European Research Council, German Ministry of Research and Education, Academy of Finland, Finnish Medical Foundation, and the Finnish Foundation for Cardiovascular Research, the Emil Aaltonen Foundation, and the Paavo Nurmi Foundation.

Copyright © 2023 The Author(s). 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/).

Keywords: Atrial fibrillation; Gut microbiome; Metagenomics; Epidemiology

Introduction Atrial fibrillation (AF) is a complex disease, and the majority of cases occur after the age of 60 years.1 The exact mechanisms of AF development and perpetuation remain unclear. Classical cardiovascular risk factors explain only slightly more than 50% of AF risk. Many of these established AF risk factors, including age, sex, hypertension,2 obesity,3 prevalent ischemic heart disease4,5 and heart failure,6 have been shown to be associated with an altered composition and function of the gut microbiome. Furthermore, diverse bacterial species have been described in atherosclerotic plaques7 and a reduced diversity of gut microbiome has been observed in heart failure.8,9

For AF, however, it remains unknown to which extent the gut microbial profile is related to the disease.10,11 The gut microbiome or its products can act on downstream targets, as has been shown for metabolites such as short-chain fatty acids (SCFA, a major product of microbial dietary fibre degradation), trimethylamine N-oxide, and lipopolysaccharides (LPS).12–14 Recently, a prospective association of trimethylamine N-oxide with AF was reported in Norwegian individuals.15 In canine experiments, trimethylamine N-oxide injected in ganglionated plexi, which are central components of the autonomic cardiac system, induced enhanced electrical excitability, atrial electrical remodelling, and prolonged induced AF,16 resulting in fibrosis and cardiac dysfunction.17 LPS may act through inflammatory pathways. In AF patients, LPS was associated with higher incidence of major adverse cardiovascular events, and adherence to a Mediterranean diet appeared to lower LPS concentrations and outcomes.18 The literature shows, that to date, most of the research has focused on more easily measurable gut microbiome-produced metabolites.10 Even without direct measurement of circulating metabolites, a more detailed assessment of the whole microbiome could help the research community to identify potential gut microbial species and metabolic pathways in human AF.

A small series of articles based on a case–control study of 50 Chinese patients hospitalized with AF was published recently.19–21 The results suggested that different gut microbiome signatures in AF exist. Changes in microbial diversity and the predominant microbiome pattern have been seen in paroxysmal AF. Current advances in metagenomics with comprehensive microbiome characterization permit the examination of the gut microbiome in relation to AF at scale. In the population-based FINRISK 2002 cohort of 6763...
individuals with >15 years of follow-up, we examined how the prevalence and long-term incidence of AF was associated with the compositional profile and functional potential of the gut microbiome and qualitatively compared our findings with published gut microbiome associations with other cardiovascular diseases.22

Methods

Study sample

The Finnish Institute for Health and Welfare has performed population surveys every five years since 1972 to monitor the development of cardiovascular risk factors in the Finnish population.23 A random population sample of 13,437 individuals aged 25–74 years from six geographic regions was invited to participate in the FINRISK 2002 study.24 Out of all invited individuals, 8799 (65.5%) participated in FINRISK 2002. In the cross-sectional sample, we excluded 1568 participants who did not provide stool samples, 20 participants due to low total read count (N < 50,000), and 448 participants due to missing relevant covariates resulting in a final sample of 6763 individuals. Of the 448 participants missing relevant covariates, 286 did not provide information on alcohol consumption, 115 did not grant permission to registry follow-up, 30 did not provide smoking status, and 17 had other missing covariates. In the prospective analyses, we additionally excluded 116 individuals with prevalent AF for a final longitudinal sample of 6647 participants.

Health examination

The participants completed a questionnaire on socio-demographic information, lifestyles, medications, and medical history at home. In the current study questionnaire information was used to define smoking status, alcohol consumption, physical activity, food choices, and as one criterion for the definition of diabetes. Physical examinations were performed at a local study site by trained staff. The participants underwent measurements for height and weight. A nurse drew venous blood samples for analysis of routine biomarkers and measured sitting blood pressure two times on the right arm using a mercury sphygmomanometer and a 14 × 40 cm sized cuff after a 5-min rest. The health examinations were performed in 2002.

Stool sampling and storage

Stool samples were collected at home after the physical examination in 50 ml Falcon tubes and were mailed to Finnish Institute for Health and Welfare using prepaid packages. The samples were then frozen in −20 °C until they underwent metagenomic sequencing in 2017.

Stool DNA extraction and library preparation

Microbiome analysis was performed at the University of California San Diego using whole-genome untargeted shallow shotgun metagenomic sequencing against mapped reference databases, following a previously published protocol.25 In brief, Illumina-compatible libraries were prepared from isolated DNA, normalized to 5 ng input per sample, and sequenced using Hi-Seq 4000 for paired-end 150 bp reads. Sequence reads were mapped against taxonomy using SHOGUN v1.0.5 against NCBI RefSeq database (version 82; May 8, 2017).26 Functional profiles were calculated from a combination of observed and predicted Kyoto Encyclopedia of Genes and Genomes Orthology group (KO) annotations from the RefSeq genomes following the predicted parameters of the SHOGUN tool.26

Variables and covariates

Body mass index (BMI) was calculated as kg/m². Smoking was defined as current daily smoking. We used self-reported average alcohol consumption (grams per week) during the last 12 months. Information on medication use was retrieved from the Finnish National Drug Purchase Register, which captures all reimbursed prescription drug purchases in Finland. Antihypertensive medication use was defined as a drug purchase occurring during the four months preceding the study baseline under following Anatomical Therapeutic Chemical classification code classes: diuretics (C03), beta-blockers (C07), calcium channel blockers (C08), and renin–angiotensin system inhibitors (C09). Prevalent diabetes was defined as self-reported diabetes, a previous diagnostic code (ICD-10 codes E10-E14 or ICD-8/9 code 250) indicating diabetes in the nationwide Care Register for Health Care, which includes hospital discharges and specialist outpatient visits, three prior diabetes medication purchases (ATC code class A10), or special reimbursement code for diabetes medications in the Drug Reimbursement Register. Heart failure was defined using a previous diagnostic code indicating heart failure in the nationwide Care Register for Health Care (ICD-10 codes I10, I110, I130, I132; ICD-9 codes 4029B, 404, 4148, and 428; ICD8 codes 42,700, 42,710, and 428) or special reimbursement code for heart failure in the Drug Reimbursement Register. AF was defined using ICD-10 code 148, ICD-8 code 4273, ICD-9 code 42,792, in the nationwide Care Register for Health Care, or Causes-of-Death Registers, or special reimbursement for drone-darone medication in the Drug Reimbursement Register before 31 December 2017.

Validation cohort and data collection

For external validation, we used a case–control study of patients with AF and limited risk factor burden compared to matched controls specifically collected for the examination of the gut microbiome between October 2019 and March 2020 at the University Clinic Hamburg-Eppendorf, Germany. Cases and controls were matched based on age, sex, cardiovascular risk factors.
factors, and medication. N = 64 of patients with AF and 74 of the controls were finally available for analysis. The OMNIgene.GUT DNA Stabilisation Kit (DNA Genotek) was used. After aliquoting and freezing at −80 °C samples were shipped to the Max-Delbrück-Center, Berlin. In total, the cohort comprised 138 individuals with microbiome data available, 64 with an AF diagnosis. Microbial DNA was extracted from stool samples and shotgun sequenced, filtered and quality controlled, then mapped using NGLESS28 to the mOTU taxonomic space v2.5.29 Prior to analysis, samples (represented by reads mapping to mOTU marker genes) were rarefied to this count from the smallest sample (1884), calculating alpha diversity in this process, using the RTK software.10 Features identified in FINRISK were assessed in the replication cohort for (direction of) effects (using the Cliff’s delta nonparametric measure) by comparing rarefied abundances of AF and control samples in this cohort. We did not differentiate between paroxysmal, persistent and permanent forms of AF.

Statistical methods
We used R version 3.6.3 for all statistical analyses. The source code for the analyses is available at https://doi.org/10.5281/zenodo.4312841. Unless otherwise noted, we adjusted the analyses for age, sex, BMI, systolic blood pressure, smoking, alcohol consumption, diabetes mellitus, heart failure, antihypertensive medication use, and total cholesterol. Alcohol consumption was log (x+1)-transformed to reduce the skewness of the lower tail bound distribution. We also assessed the characteristics of the study sample versus those that were excluded due to missing covariates. We calculated alpha diversity (Shannon index as a measure of mean species diversity as variation and richness in the sample) using species-level data with the R package vegan.31 We studied the association between prevalent AF and alpha diversity using logistic regression where prevalent AF was the dependent variable. With N = 6763, 539 incident cases, and alpha set at 0.05, we had 80% and 90% powers to detect odds ratios of 1.13 and 1.16.32 We calculated the dissimilarity matrix (beta diversity indicating the variation in taxonomic abundance profiles between samples) and Principal Coordinates Analysis (PCoA) using Bray–Curtis dissimilarity on compositional microbial species-level abundance using R packages vegan.31.37 We further studied common microbial genera prevalent in at least 1% of the sample population with a relative abundance over 0.1%. We examined associations of prevalent AF or incident AF (prevalent cases excluded) with the common microbial genera using DESeq2 with the Benjamini–Hochberg correction (FDR).33 In DESeq2, microbiome composition is used as the outcome, instead of the exposure variable. The development of AF was assessed in subset of participants without AF at baseline Cox regression models with Breslow approximation for centered log-ratio transformed (CLR) microbial abundances.34 We also performed functional analyses using log (x+1) transformed KEGG Orthology (KO) groups using Cox regression models with the Benjamini–Hochberg correction.

We performed two additional sensitivity analyses. In the first sensitivity analysis, we introduced two additional covariates: leisure time activity and healthy food choices.35 In the second sensitivity analysis, we limited the follow-up to 7.5 years which is approximately half of the total follow-up time. We also performed sparse Partial Least Squares Discriminant Analysis (sPLS-DA) using mixOmics library under R version 4.1.2.36 We tuned optimal values for the sparsity parameters using k-fold cross validation.

Ethics
FINRISK 2002 study and the case–control validation study complies with the Declaration of Helsinki. Helsinki. Informed, written consent was obtained from all participants. The Coordinating Ethics Committee of the Helsinki and Uusimaa University Hospital District approved the FINRISK 2002 study. The case–control validation study was approved by Ärztekammer Hamburg (PV5705).

Role of funders
Funders had no role in the in the study design, data collection, data analyses nor interpretation or writing of the report.

Results
The baseline characteristics of the cross-sectional and prospective samples are shown in Table 1. The characteristics of the study sample and of individuals with missing covariates are reported in Supplemental Table S1. With the large study sample size even small between-group differences were significant using chi-square test and ANOVA. However, the number of individuals with missing data was low and the absolute/clinical between-group differences were small. We defined common genera as genera that were present in the stool samples of at least 1% of study participants; we used a cut-off value of over 0.1% relative abundance to define the presence of a genus in a stool sample. We observed 91 common microbial genera (Supplementary Table S2).

Gut microbiome alpha diversity was not associated with prevalent AF in the age- and sex-adjusted (odds ratio [OR] 1.00; 95% confidence interval [CI] 0.83–1.20; P = 0.98) or in the multivariable-adjusted models (OR 1.04; 95% confidence interval [CI] 0.86–1.26; P = 0.71). Gut microbiome beta diversity was not associated with AF in age- and sex-adjusted (R² = 0.024%; P = 0.05) and multivariable-adjusted models (R² = 0.020%; P = 0.12).

Fig. 1 shows microbial diversity (Bray–Curtis
dissimilarity) as principal coordinate analysis for species-level bacterial abundances. The first three PCoA axes explained 31.2% of the variation in bacterial abundances. Ecological diversity measures did not accurately discriminate participants with AF status.

We observed prevalent AF having nine significant associations with common microbial genera with FDR-corrected $P < 0.05$ (Fig. 2, Table 2). The associations were positive for Eisenbergiella, Enorma, Enterobacter, and Kluyvera, and negative for Bacteroides, Bifidobacterium, Holdemanella, Parabacteroides, and Turicibacter. Therefore, specific genus-level gut microbial abundances have potential to identify individuals with AF.

We also performed sPLS-DA analysis to maximize the discrimination potential between general gut microbial composition and AF. The method did not improve discrimination compared to principal coordinate analysis. Therefore, we suggest that the potential link between gut microbiota and AF is mainly driven by specific gut microbial species rather than general gut microbial composition.

A total of 539 individuals developed AF over a median follow-up of 14.8 ± 3.0 years. There were no statistically significant differences in gut microbiome alpha or beta diversity between individuals who did and did not develop AF. We observed eight associations between incident AF and baseline common microbial genera with FDR-corrected $P < 0.05$ using DESeq2 (Table 3, Fig. 2). These associations were positive for Bifidobacterium, Enorma, Lactococcus, Mitsuokella, and Sellimonas, and negative for Tyzzerella, Hungatella, and Sanguibacteroides. Comparisons between top bacterial genera in each analysis were evaluated in the validation cohort (baseline characteristics see Supplementary Table S3) as the direction of effect contrasting AF cases and controls (Supplementary Table S4). Regression analysis revealed that 6 out of 8 shared top bacterial genera detected in both cohorts were shifted in the same direction (Fig. 3). The top hits DEseq2 models overlapped for Enorma, Eisenbergiella, and Bifidobacterium (Fig. 4). While ecological diversity measures did not accurately discriminate participants that developed AF.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall (N = 6763)</th>
<th>No prevalent AF (N = 6647)</th>
<th>Prevalent AF (N = 116)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years (SD)</td>
<td>49.2 (12.9)</td>
<td>48.9 (12.8)</td>
<td>62.9 (8.5)</td>
</tr>
<tr>
<td>Women, N (%)</td>
<td>3680 (54.4)</td>
<td>3646 (54.9)</td>
<td>34 (29.3)</td>
</tr>
<tr>
<td>Body mass index, kg/m² (SD)</td>
<td>26.9 (4.6)</td>
<td>26.9 (4.6)</td>
<td>29.2 (5.2)</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg (SD)</td>
<td>135.6 (20.2)</td>
<td>135.5 (20.2)</td>
<td>144.2 (21.3)</td>
</tr>
<tr>
<td>Diabetes mellitus, N (%)</td>
<td>371 (5.5)</td>
<td>357 (5.4)</td>
<td>14 (12.1)</td>
</tr>
<tr>
<td>Current smoker, N (%)</td>
<td>1594 (23.6)</td>
<td>1580 (23.8)</td>
<td>14 (12.1)</td>
</tr>
<tr>
<td>Antihypertensive medication, N (%)</td>
<td>1216 (18.0)</td>
<td>1134 (17.1)</td>
<td>82 (70.7)</td>
</tr>
<tr>
<td>Total cholesterol, mmol/l (SD)</td>
<td>5.6 (1.1)</td>
<td>5.6 (1.1)</td>
<td>5.6 (1.0)</td>
</tr>
<tr>
<td>Alcohol consumption, g (SD)</td>
<td>80.4 (122.0)</td>
<td>80.7 (122.1)</td>
<td>62.8 (113.8)</td>
</tr>
<tr>
<td>Prevalent atrial fibrillation, N (%)</td>
<td>116 (1.7)</td>
<td>0 (0.0)</td>
<td>116 (100.0)</td>
</tr>
<tr>
<td>Incident atrial fibrillation, N (%)</td>
<td>539 (8.0)</td>
<td>539 (8.1)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Heart failure, N (%)</td>
<td>94 (1.4)</td>
<td>70 (1.1)</td>
<td>24 (20.7)</td>
</tr>
</tbody>
</table>

Table 1: Baseline characteristics of the study cohort.

Fig. 1: Microbial diversity (Bray–Curtis dissimilarity) shown using principal coordinate analysis of species-level microbial abundances with prevalent atrial fibrillation cases denoted using red.
during follow-up, baseline genus level abundances have potential to identify individuals with extended follow-up.

In sensitivity analyses we observed that half of the genera associated with incident AF remained significant and had relatively unchanged effect sizes when follow up was limited to 7.5 years (Supplementary Table S5). The eight genera associated with incident AF remained significant with similar effect sizes in sensitivity analyses when additional covariates for exercise and healthy food choices were included in the model (Supplementary Table S5). We also studied the association between the common microbial genera and incident AF using Cox proportional hazards models but observed no significant associations after FDR correction (Table 4).

We then studied the functional associations between KO groups and incident AF using Cox proportional hazard models. Based on prior literature, we focused on 479 KO groups that are associated with the production of SCFAs and 14 trimethylamines, both well-known gut microbiome products.39,40 In total, 288 of these 493 KO groups were detected in our baseline sample. We observed positive associations of AF with two KO groups (K15896, K15913) related to amino sugar and nucleotide sugar metabolism and with one KO group (K07271) linked to lipopolysaccharide biosynthesis (uncorrected P < 0.05; Supplementary Table S6). We also studied the associations between all available 6843 KO groups and incident AF observing FDR-corrected P > 0.05 for all associations (Supplementary Table S6).

Discussion
In a large, well-established population-based cohort we identified modest associations of prevalent and incident AF with the gut microbiome. The proportion of variance in microbial diversity measures explained by AF was low. We observed that prevalent AF was associated with nine genera and incident AF with eight genera using an FDR-corrected P value threshold of 0.05; *Enorma*, *Bifidobacterium*, and *Eisenbergiella* were among the top associations of prevalent and incident AF. *Enorma* also appeared among the top hits in Cox regression models for AF. Some plausible species and genera were identified in relation to AF, which are known in the context of established AF risk factors such as blood pressure control and heart failure. In a validation analysis, 75% of top genera (*Paraprevotella*, *Odoribacter*, *Collinsella*, *Enorma*, *Barnesiella*, *Alistipes*) in our Cox regression analyses showed a consistent direction of shifted abundance in an independent AF case–control cohort.

Differentially abundant genera in atrial fibrillation
At the genera level, the top association of prevalent AF was observed with *Enorma*. *Enorma* belongs to the family of *Coriobacteriaceae*, which were among the core

### Table 2: Top ten associations of prevalent atrial fibrillation (N = 116) with common genera using DESeq2 (N = 6763).

<table>
<thead>
<tr>
<th>Bacterial genus</th>
<th>Log2-fold change</th>
<th>SE</th>
<th>FDR-corrected P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enorma</td>
<td>0.91</td>
<td>0.15</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Holdemanella</td>
<td>−0.80</td>
<td>0.18</td>
<td>0.001</td>
</tr>
<tr>
<td>Eisenbergiella</td>
<td>0.81</td>
<td>0.19</td>
<td>0.001</td>
</tr>
<tr>
<td>Kluyvera</td>
<td>1.07</td>
<td>0.30</td>
<td>0.008</td>
</tr>
<tr>
<td>Parabacteroides</td>
<td>−0.47</td>
<td>0.16</td>
<td>0.039</td>
</tr>
<tr>
<td>Turicibacter</td>
<td>−0.74</td>
<td>0.25</td>
<td>0.039</td>
</tr>
<tr>
<td>Enterobacter</td>
<td>0.88</td>
<td>0.29</td>
<td>0.039</td>
</tr>
<tr>
<td>Bacteroides</td>
<td>−0.79</td>
<td>0.27</td>
<td>0.039</td>
</tr>
<tr>
<td>Bifidobacterium</td>
<td>−0.49</td>
<td>0.17</td>
<td>0.041</td>
</tr>
<tr>
<td>Desulfovibrio</td>
<td>0.56</td>
<td>0.20</td>
<td>0.051</td>
</tr>
</tbody>
</table>

The estimates are adjusted for age, sex, body mass index, systolic blood pressure, smoking, alcohol consumption, diabetes mellitus, heart failure, antihypertensive medication use, and total cholesterol. P values shown are adjusted for multiple testing using the Benjamini–Hochberg correction (FDR). Log2-fold changes were estimated using DESeq2. SE stands for standard error.
families related to heart failure. In a small, clinical case–control study with 20 heart failure patients the authors showed a significantly lower abundance of *Coriobacteriaceae* in diseased individuals. Enorma species also appeared among the top associations with incident AF in both, DESeq2 and Cox regression analyses even after adjustment for prevalent heart failure. An association between Enorma and hypertension, which is strongly related to AF, has also been reported. Further, *Coriobacteriaceae* have been positively correlated with total cholesterol, low-density cholesterol, and body mass index in healthy humans. A phase II study evaluating the safety and efficacy of a non-steroidal farnesoid X receptor agonist in non-alcoholic fatty liver disease was terminated early because short intervals of cardiac arrhythmia were recorded during Holter monitoring. In this study, the authors had measured a relative decrease in abundance of *Coriobacteriaceae* in the participants’ gut microbiome.

In DESeq2 analyses we found Eisenbergiella borderline differentially abundant in incident AF and significantly related to prevalent AF. This finding is in line with reports showing how Eisenbergiella is more abundant in normotensive persons. In addition, the abundance of this genus is different in coronary artery disease patients. Which is a strong predictor and established risk factor of AF.

We observed a possible negative association of prevalent AF with *Bifdobacterium*, and a positive relation for incident AF. This observation may reflect different disease stages with higher impact of concurrent conditions such as heart failure in individuals with AF at baseline. In heart failure patients, the genus *Bifdobacterium* is depleted. *Bifdobacterium* has been positively correlated with ejection fraction and negatively with the cardiac stress marker N-terminal pro B-type natriuretic peptide. This prior knowledge may help explain the lower abundance observed in prevalent AF with a link to heart failure. *Bifdobacterium* belongs to the most abundant intestinal bacteria and is an integral part of most probiotics. Treatments with probiotics containing *Bifdobacterium* have been suggested to improve the atherogenic lipid profile. Furthermore, *Bifdobacterium* has been related to favourable modulation of blood pressure.

In Cox regression analysis for incident AF, the genus *Odoribacter* was among the most differentially abundant bacteria. It comprises common species of the human intestinal microbiome isolated from faeces. One of the

<table>
<thead>
<tr>
<th>Bacterial genus</th>
<th>Log2-fold change</th>
<th>SE</th>
<th>FDR-corrected P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sellimonas</td>
<td>1.02</td>
<td>0.09</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mitsuokella</td>
<td>1.02</td>
<td>0.13</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Enorma</td>
<td>0.46</td>
<td>0.07</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Tyzzerella</td>
<td>−0.32</td>
<td>0.05</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><em>Bifdobacterium</em></td>
<td>0.37</td>
<td>0.08</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lactoacoccus</td>
<td>0.47</td>
<td>0.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hungatella</td>
<td>−0.20</td>
<td>0.06</td>
<td>0.016</td>
</tr>
<tr>
<td><em>Sanguibacteroides</em></td>
<td>−0.26</td>
<td>0.08</td>
<td>0.022</td>
</tr>
<tr>
<td>Lactobacillus</td>
<td>0.27</td>
<td>0.10</td>
<td>0.051</td>
</tr>
<tr>
<td>Eisenbergiella</td>
<td>−0.24</td>
<td>0.09</td>
<td>0.084</td>
</tr>
</tbody>
</table>

The estimates are adjusted for age, sex, body mass index, systolic blood pressure, smoking, alcohol consumption, diabetes mellitus, heart failure, antihypertensive medication use, and total cholesterol. P values shown are adjusted for multiple testing using the Benjamini–Hochberg correction (FDR). Log2-fold changes were estimated using DESeq2. SE stands for standard error.

Table 3: Top ten associations of incident atrial fibrillation (N = 539) with common genera using DESeq2 (N = 6647).
top associations in both prospective analyses was observed with a related taxon of *Sanguibacteroides*.52 In our cohort, individuals who developed AF had lower abundances of *Odoribacter* and *Sanguibacteroides*.A higher relative abundance of the *Odoribacter* genus has been related to lower blood pressure in pregnant women.53 Blood pressure is a strong AF risk factor.54 Bacteria of this genus produce the SCFA acid butyrate, a signalling molecule in blood pressure control.55 SCFAs belong to the most abundant gut microbiome-derived physiologic modulators. They interact with G-protein coupled receptor pathways including renin secretion and sympathetic activation,56 which are central to blood pressure regulation. Further, *Odoribacter* abundance positively correlates with isobutyric acid and an unfavourable lipid profile.57 On the other hand, abundance of

<table>
<thead>
<tr>
<th>Bacterial genus</th>
<th>HR</th>
<th>95% confidence interval</th>
<th>P value</th>
<th>FDR-corrected P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odoribacter</td>
<td>0.917</td>
<td>0.851–0.982</td>
<td>0.009</td>
<td>0.697</td>
</tr>
<tr>
<td>Solobacterium</td>
<td>1.093</td>
<td>1.021–1.164</td>
<td>0.015</td>
<td>0.697</td>
</tr>
<tr>
<td>Sanguibacteroides</td>
<td>0.943</td>
<td>0.893–0.995</td>
<td>0.026</td>
<td>0.804</td>
</tr>
<tr>
<td>Collinsella</td>
<td>1.041</td>
<td>0.982–1.100</td>
<td>0.186</td>
<td>0.818</td>
</tr>
<tr>
<td>Enorma&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.069</td>
<td>0.967–1.172</td>
<td>0.201</td>
<td>0.818</td>
</tr>
<tr>
<td>Barnesiella</td>
<td>0.96</td>
<td>0.913–1.008</td>
<td>0.095</td>
<td>0.818</td>
</tr>
<tr>
<td>Paraprevotella</td>
<td>0.962</td>
<td>0.914–1.011</td>
<td>0.12</td>
<td>0.818</td>
</tr>
<tr>
<td>Alistipes</td>
<td>0.942</td>
<td>0.871–1.013</td>
<td>0.101</td>
<td>0.818</td>
</tr>
<tr>
<td>Enterococcus</td>
<td>0.941</td>
<td>0.857–1.024</td>
<td>0.15</td>
<td>0.818</td>
</tr>
<tr>
<td>Clostridium</td>
<td>1.087</td>
<td>0.967–1.207</td>
<td>0.174</td>
<td>0.818</td>
</tr>
</tbody>
</table>

The hazard ratios (HR) are adjusted for age, sex, body mass index, systolic blood pressure, smoking, alcohol consumption, diabetes mellitus, heart failure, antihypertensive medication use, and total cholesterol. Microbial abundances were transformed using centered log-ratio transformation. We used Breslow approximation for ties in Cox regression model. P values shown are adjusted for multiple testing using the Benjamini-Hochberg correction (FDR).<sup>a</sup>Overlap with top genera of DESeq2 analyses.<sup>b</sup>Overlap with top genera in prevalent atrial fibrillation.

Table 4: Top ten common genera associated with incident atrial fibrillation in Cox regression analyses (N = 6923).
succinate-metabolizing Odoribacteraceae was lower in obesity, with a significant variation in the gut microbiome under Mediterranean diet.¹⁹

Results from previous reports in context with the current data
Prior studies have been small and inconsistent often lacking robust validation. In 50 Chinese AF patients, the genus Dorea was among the ten genera with the highest differences compared to controls and was more abundant in AF patients.¹⁹ Dorea was less abundant in heart failure in a prior study²⁰ and has been associated with atherosclerotic cardiovascular disease.²⁰ However, in our study neither prevalent nor incident AF was associated with Dorea. Whether this is a spurious finding or related to differences in ethnicity or exposome remains unclear. In the same study by Zuo K et al., Bifidobacterium was also more abundant in AF whereas we observed a negative association in our data. Another Chinese case-control study reported a gut microbial shift towards Bacteroides besides Prevotella as the dominant enterotypes. This change in microbial composition was already seen in paroxysmal AF and accentuated in more persistent AF and with longer duration of AF.²⁰,²¹ Bacteroides were among the top ten genera for prevalent AF in our current study and thus may demonstrate a certain consistency of findings.

Further overlap in results of our and other prior AF or heart failure studies that usually had samples sizes of <100 participants and used less deep 16S ribosomal RNA sequencing for analysing bacterial populations was only minor.¹¹ In our study, besides Enorma, there was further partial overlap of the most strongly associated genera such as Parabacteroides in relation to prevalent and Sellimonas bacteria in relation to incident AF. For Sellimonas species a negative association with sodium and blood pressure has been demonstrated.⁴¹ Thus, a minor shift in underlying faecal microbial composition related to the arrhythmia can be assumed.

Whereas our data are not intended to fully elucidate the mechanisms between changes in gut microbiome composition and AF, the aspect of its modifiability is attractive. In our results, a major pathophysiological axis to hypertension evolved. The majority of the most abundant genera in relation to prevalent and incident AF (>70%) have been shown to correlate with blood pressure indices in the FINRISK cohort.⁴² The gut–immune interaction can be modulated by salt-intake and thus address salt-sensitive hypertension. It can be speculated that small changes achieved through dietary, pre- or probiotic provisions or pharmacologic interventions on the gut microbiome could produce relevant effects over a lifetime. Since blood pressure carries a high attributable risk, even lowering it by a few mm Hg may decrease AF incidence.⁴³ Further, a Mediterranean diet can increase the abundance of Bifidobacteria and could thus help restore microbial dysbiosis seen in AF towards a more favourable enterotype. In a posthoc analysis of the Prevención Con Dieta Mediterránea Trial (PREDIMED), olive oil in context with a Mediterranean dietary pattern reduced AF incidence.⁴⁴ However, more evidence is needed to understand whether the microbiome composition represents a modifiable risk factor or risk marker of lifestyle components and comorbidities in AF patients.

Caveats and limitations
AF is a heterogeneous phenotype and associations may have been diluted due to the lack of subtype differentiation.⁴⁵ In the current study, self-reported information was used to define certain lifestyle-related covariates which could result in self-reporting bias. For the outcome, we relied on discharge diagnoses and coded data at in- and outpatient clinics. Paroxysmal, oligo- or asymptomatic AF may thus have been missed and weakened observed differences in associations. The gut microbiota composition was assessed only once, which may not accurately represent the longitudinal changes over time during the follow-up in this study. We could not replicate and identify potentially spurious findings derived from mass data interrogation because unfortunately, there is no comparable study with prospective data. This issue limited us to validating our main findings in a case-control sample with a comparatively small sample size. The extensive requirement of computational resources prevented us from studying the associations between KEGG Orthology groups and AF using DESeq2. The matching of the case-cohort validation sample was performed before the stool sample substudy. Not all participants were willing to return to the study center and participate in the substudy. Therefore, we observe differences between the sample sizes and proportions of the base characteristics that could potentially affect the analysis results. Observed differences are however, mostly modest and the case-control sample was used only for validation of the previously observed significant associations.

Strengths of the study are the large sample size with long-term follow-up and a maximum depth of information through shotgun sequencing that provide biologically plausible results which serve as hypothesis-generating. In addition, an overlap of the most strongly associated species can serve as internal validation because individuals with prevalent AF were excluded from prospective analyses. Furthermore, though agreement was substantial only for our Cox regression top drivers, our findings derived from mass data interrogation because unfortunately, there is no comparable replication studies. However, the reason behind the observed association between baseline gut microbiota and long-term development of AF is non-trivial and there may be other contributing factors that we were unable to adjust for in the current study. Although epidemiological study participation rates have been decreasing over the past five decades, FINRISK 2002 had a high overall health examination participation rate of 66% and 54% of all
invited individuals (80% of health examination participants) donated a stool sample. The invited population sample was randomly selected from the Finnish population register. We therefore think that FINRISK 2002 and the stool samples are fairly representative of the general population of Finnish people in 2002.

Conclusions
In conclusion, we observed a different microbiome composition in prevalent and incident AF compared to non-affected individuals with a number of genera and species which differed in abundance. Overall, the alpha and beta diversity of the gut microbiome did not meaningfully discriminate individuals with prevalent or incident disease. Machine learning methods optimized for binomial discrimination and use of species (or even higher) resolution level may be required to improve our understanding of the role between AF and gut microbiome. The shift of the bacterial composition towards a spectrum with similarities to the microbiome in hypertension and heart failure highlights a shared underlying pathophysiology. It still remains unknown whether modulating the intestinal microbiome and metabolism offers new approaches to primary, secondary or tertiary prevention of AF, and whether tracking the gut microbiome composition may help to guide lifestyle interventions and management in AF patients.

Contributors
Joonatan Pärmu and Christin S. Börnschel contributed to the data analysis, writing, and figures. Alfredo Ortega-Alons, Lajos Markó, Mike Inouye, Rodolfo A. Salido, Karinna Sanders, Caitriona Brennan, Gregory C Humphrey, Friederike Gutmann, Jon G Sanders, Dominik Linz, and Rob Knight contributed to the data preparation. Pekka Joutsilahti, Veikko Salomaa, and Aki S. Havulinna contributed to the data collection and study design. Leo Lahti data interpretation and study design. Teemu Niiranen and Renate B. Schnabel contributed to the study design, data interpretation, and writing. All authors read and approved the final version of the manuscript. Joonatan Pärmu and Teemu Niiranen verified underlying FINRISK 2002 data. Sofia K. Forsthul verified case-cohort validation study data.

Data sharing statement
The FINRISK 2002 data described in the manuscript are available from the Finnish Institute for Health and Welfare Biobank based on a written application as instructed on the website of the Biobank (https://thl.fi/en/web/thl-biobank/for-researchers/application-process). The phenotype data are not publicly available because they contain information that could compromise research participant privacy/consent. The metagenomic data are not publicly available because they contain information that could compromise research participant privacy/consent. The case-cohort validation study data can be obtained upon request via the Center for Microbiome Innovation and Clarity Genomics. VS has ongoing research collaboration with Bayer Ltd., unrelated to the present study.

Acknowledgements
RBS has received consulting fees from GenCirq Inc, DayTwo Ltd and JGilbert Consulting LLC, payment for lectures from BP Technology Ventures Inc and Hamilton College, support for attending meetings and/or travel from the Chilean Senate, the University of Melbourne, the Shenzhen Society of Science and Technology, the Norwegian Institute of Public Health and participated on a data safety monitoring board for Diversigen, GenCirq Inc, DayTwo Ltd, Cybèle Microbiome Inc, BioMeSense Inc, Micronoma Inc. He owns stocks or stock options from Biotia Technology Inc, Diversigen, GenCirq Inc Cybèle Microbiome Inc and Micronoma Inc. RBS has received lecture fees and advisory board fees from BMS/Pfizer outside this work. TN received payment for lectures or presentations from Servier. JGS occasionally perform bioinformatics consulting for research group in the field of microbiome science via the Center for Microbiome Innovation and Clarity Genomics. VS has ongoing research collaboration with Bayer Ltd., unrelated to the present study.

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.ehjim.2023.104583.

References


