



OPEN

Atrial fibrillation and left atrial size and function: a Mendelian randomization study

Yordi J. van de Vegte^{1,4}, Joylene E. Siland^{1,4}, Michiel Rienstra¹ & Pim van der Harst^{1,2,3}✉

Atrial fibrillation (AF) patients have enlarged left atria (LA), but prior studies suggested enlarged atria as both cause and consequence of AF. The aim of this study is to investigate the causal association between AF and LA size and function. In the UK Biobank, all individuals with contoured cardiovascular magnetic resonance data were selected. LA maximal volume (LA max), LA minimal volume (LA min), LA stroke volume and LA ejection fraction were measured and indexed to body surface area (BSA). Two-sample Mendelian randomization analyses were performed using 84 of the known genetic variants associated with AF to assess the association with all LA size and function in individuals without prevalent AF. A total of 4274 individuals (mean age 62.0 ± 7.5 years, 53.2% women) were included. Mendelian randomization analyses estimated a causal effect between genetically determined AF and BSA-indexed LA max, LA min, and LA ejection fraction, but not between AF and LA stroke volume. Leave-one-out analyses showed that the causal associations were attenuated after exclusion of rs67249485, located near *PITX2* gene. Our results suggest that AF causally increases LA size and decreases LA ejection fraction. The AF risk allele of rs67249485, located near the *PITX2* gene, contributes strongly to these associations.

Atrial fibrillation (AF) is the most common cardiac arrhythmia worldwide, and many patients with AF develop an enlarged left atrium (LA)^{1,2}. LA enlargement is associated with poorer prognosis of AF ablation outcomes and AF recurrences³, but may also increase the risk of stroke, adverse cardiovascular outcomes and death^{4,5}.

LA enlargement is hypothesized to be a result of atrial remodeling, a persistent change in atrial structure or function^{6,7}. However, pressure and/or volume overload commonly seen in conditions as hypertension, structural heart disease, mitral valve disease and heart failure may also induce change in atrial structure or function. As a consequence of atrial remodeling, it might trigger AF episodes, and then a vicious circle starts where AF episodes might trigger further atrial remodeling^{8,9}. The degree of atrial remodeling can be assessed through measurement of LA volume with cardiovascular magnetic resonance (CMR) imaging¹⁰.

Co-existence of risk factors of AF and LA size and function makes it difficult to determine causality. Uncertainty exists if atrial remodeling is the cause or the consequence of AF. The hypothesis of a causal link between AF and LA volume may be tested by applying a Mendelian randomization approach (MR). Since AF associated genetic variants are randomly assigned at birth, a “naturally” randomized controlled trial can be performed, assuming that (1) genetic variants are reliably associated with AF, (2) genetic variants are independent of confounding factors and (3) genetic variants are only associated with LA volume through AF¹¹. In present study, we use a two-sample MR approach to study the potential causal association between AF and LA size and function in the UK Biobank.

Results

In the current study, 4274 individuals from the general population were included (mean age 62.0 ± 7.5 , 53.2% women). The mean body mass index (BMI) was 26.6 (SD 4.4) kg/m², the prevalence of hypertension and diabetes mellitus type II were 32.1% and 3.5%. Body surface area (BSA) indexed maximum LA volume (LA max), minimum LA volume (LA min) and LA stroke volume were 35.9 ± 10 ml/m², 14.1 ml/m² [Interquartile range (IQR) 10.9–17.9], and 21.1 ± 5.6 ml/m², respectively. LA ejection fraction (LA EF) was on average $59.4 \pm 8.3\%$. A total of 36 individuals (0.8%) developed AF during the median follow-up of 2.0 years (IQR 1.8–2.4). Additional

¹Department of Cardiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, P.O. Box 30.001, 9700 RB Groningen, The Netherlands. ²Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. ³Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, University of Utrecht, Utrecht, The Netherlands. ⁴These authors contributed equally: Yordi J. van de Vegte and Joylene E. Siland. ✉email: p.van.der.harst@umcg.nl

	Sample
No	4274
Age, y	62.0 ± 7.5
Sex, female, %	53.2
BMI, kg/m ²	26.6 ± 4.4
Diabetes mellitus type 2, %	3.5
Hypertension, %	32.1
Prevalent atrial fibrillation, %	0
Incident atrial fibrillation, %	0.8
LA max (ml)	67.0 ± 20.0
LA max, indexed (ml/m ²)	35.9 ± 10.0
LA min (ml)	26.0 (19.9–33.6)
LA min, indexed (ml/m ²)	14.1 (10.9–17.9)
LA SV (ml)	39.2 ± 11.1
LA SV, indexed (ml/m ²)	21.1 ± 5.6
LA EF (%)	59.4 ± 8.3
BSA (m ²)	1.9 ± 0.2

Table 1. Baseline characteristics. Continuous variables are presented as mean ± SD, skewed variables (defined as $-1 < \text{skewness} > 1$) as median (IQR min – IQR max) and binary variables as percentages. LA volumes were measured using the biplane method. BMI, denotes body mass index; LA, left atrium; SV, stroke volume; EF, ejection fraction; BSA, body surface area.

information on the cohort is provided in Table 1. A total of 24 genetic variants were removed from MR analyses to reduce risk of weak instrument bias (F-statistic < 10) and 2 genetic variants were excluded during data harmonization. A total of 84 genetic variants were taken forward for further analyses. The total amount of genetic variants varies per outcome due to MR-Steiger filtering. Data supporting the genetic variants selection (F-statistics, data harmonization, Steiger filtering) and single genetic variant-estimates for all outcomes can be found in Supplementary Table 1.

Results of the MR analyses between AF and indexed LA volumes and ejection fraction are shown in Fig. 1 and Supplementary Table 2. Additional information on the MR analyses of the unadjusted LA volumes can be found in Supplementary Table 2. Sensitivity analyses were performed to test whether the assumptions of the MR analyses were fulfilled (Supplementary Table 3). MR-Steiger directionality test indicated that the 84 genetic variants known to be associated with AF explained ~7% of AF variance. The genetic variants explained more of AF variance than indexed LA max volume (1.7%), indexed LA min volume (1.7%), indexed LA stroke volume (1.8%) and LA ejection fraction (2.0%) (Supplementary Table 3).

Using the Rucker framework, we found evidence for unbalanced horizontal pleiotropy in the MR estimates of indexed LA max and indexed LA stroke volume, indicated by significant Q-Q' and MR-Egger intercepts ($P < 0.05$) (Supplementary Table 3). We therefore took forward the MR-Egger model as primary MR-method to assess the genetic association with indexed LA max and indexed LA stroke volume, whereas we adopted the inverse variance weighted random effects (IVW-RE) model for indexed LA min and LA ejection fraction. Using these models, we found evidence for a causal effect of genetic susceptibility to AF on indexed LA max ($\beta = 1.56$, SE = 0.53, $P = 4.0 \times 10^{-3}$), indexed LA min ($\beta = 0.57$, SE = 0.19, $P = 2.0 \times 10^{-3}$) and LA ejection fraction ($\beta = -0.89$, SE = 0.25, $P = 4.1 \times 10^{-4}$) (Fig. 1). Weak-instrument bias was indicated within the MR-Egger estimate of AF on indexed LA max ($I^2_{GX} = 0.94$). We did not find evidence for a causal association between genetic susceptibility to AF and indexed LA stroke volume ($\beta = 0.54$, SE = 0.29, $P = 6.98 \times 10^{-2}$). Scatter- and forest plots of the MR analyses between AF and all LA dimensions are provided in Supplementary Figs. 2–8.

Several sensitivity analyses were performed to test whether valid conclusions on causal inference could be made under different assumptions of possible underlying pleiotropy or instrumental invalidity. We investigated whether the results were consistent under the scenario where a relatively large portion of the genetic instruments is invalid using the weighted median approach. Using this approach, we found additional evidence for a significant causal estimate between genetic susceptibility to AF and indexed LA max ($\beta = 1.36$, SE = 0.47, $P = 3.83 \times 10^{-3}$), indexed LA min ($\beta = 0.89$, SE = 0.30, $P = 2.8 \times 10^{-3}$) and LA EF ($\beta = -1.17$, SE = 0.42, $P = 5.84 \times 10^{-3}$). We then investigated whether the results were consistent under the scenario in which a small proportion of the genetic variants are outliers using the MR-Lasso approach. Using this approach, we find the genetic associations between AF and indexed LA min ($\beta = 0.57$, SE = 0.19, $P = 1.98 \times 10^{-3}$) as well as LA ejection fraction ($\beta = -0.89$, SE = 0.25, $P = 4.09 \times 10^{-4}$) to be robust to this scenario. However, the association between genetic susceptibility to AF and indexed LA max ($\beta = 0.48$, SE = 0.30, $P = 1.13 \times 10^{-1}$) was attenuated (Fig. 1).

We examined which genetic variant(s) drove the attenuation of the association between genetic susceptibility to AF and LA size and function by performing leave-one-out analyses. Results of the leave-one-out analyses using an IVW and MR-Egger approach are provided in Supplementary Table 4 and can be visually inspected in Supplementary Figs. 9–15. We observed that the MR-Egger estimate of AF on indexed LA max was attenuated after exclusion of rs67249485 ($\beta = 1.41$, SE = 0.82, $P = 9.05 \times 10^{-2}$), a genetic variant located on the long arm

of chromosome 4 in the proximity of the *PITX2* gene. However, the Wald estimate of rs67249485 did show a significant association for indexed LA max ($\beta = 1.38$, $SE = 0.58$, $P = 1.65 \times 10^{-2}$). The results are shown in Fig. 1. The leave-one-out analyses also showed an attenuation of IWR-RE estimates after exclusion of rs67249485 for indexed LA min ($\beta = 0.36$, $SE = 0.22$, $P = 1.00 \times 10^{-1}$), and LA EF ($\beta = -0.73$, $SE = 0.30$, $P = 1.52 \times 10^{-2}$). The Wald statistics for the association between rs67249485 and indexed LA min ($\beta = 1.13$, $SE = 0.36$, $P = 1.44 \times 10^{-3}$) as well as LA ejection fraction ($\beta = -1.29$, $SE = 0.47$, $P = 5.94 \times 10^{-3}$) were significant (Fig. 1).

We performed several quality controls to gain insights in the statistical validity of rs67249485 driving the association between genetic susceptibility to AF and LA dimensions and functions. Histograms of LA dimension distributions per AF increasing T allele showed absence of outliers which could drive current MR estimates (Supplementary Fig. 16). The genetic variant rs67249485 explained more variance for AF (MR-Steiger $R^2 = 1.58\%$) than for any LA size or function, which ranged up to a maximum explained variance of 0.23% for LA min. This indicates that the Wald estimates assessed the true causal direction (Supplementary Table 1).

Lastly, we performed multivariable MR analyses to assess whether the described genetic associations between AF and LA size and function are independent of blood pressure as it can affect both AF¹² and LA size and function^{13,14}. In brief, all multivariable Mendelian randomization analyses were similar to the univariable results. For example, the main MR-Egger analyses of AF on index LA max ($\beta = 1.56$, $SE = 0.53$, $P = 4.0 \times 10^{-3}$) had similar effect estimates as in the multivariable MR in which we corrected for systolic blood pressure ($\beta = 1.68$, $SE = 0.53$, $P = 1.6 \times 10^{-3}$). Please see Supplementary Table 2 for the full results and Supplementary Table 3 for the sensitivity analyses.

The MR analyses for the non-indexed LA volumes are provided in Supplementary Tables 1–4. The results were consistent to the results on the indexed LA volumes. The MR analyses for LA min (indexed and non-indexed) were repeated using genetic variant-outcome effect estimates obtained from their log-transformed equivalents to account for right skewness. Results were comparable to the primary analyses (Supplementary Table 2).

Discussion

Our study provides evidence to support the hypothesis that genetically susceptibility to AF increases indexed LA max, LA min and decreases LA ejection fraction. We pinpoint that rs67249485, near the *PITX2* gene, is the driver of the association with indexed LA max and LA min and contributes strongly to the association with LA ejection fraction. However, we did not find evidence for a causal association between AF and LA stroke volume.

Our primary analyses indicate that genetic susceptibility to AF causally increases indexed LA max and LA min. A causal association between AF and LA stroke volume was not established. One potential explanation for this discrepancy is that AF increases indexed LA max and indexed LA min in a similar degree, nullifying the effect on LA stroke volume. Another potential explanation is that a larger passive conduit function of the LA could compensate for a decreased pump function at larger maximal LA volume through the Frank-Starling law^{15,16}. This would result in similar LA stroke volume and lower LA ejection fraction^{15,16}. In fact, we do find that genetic susceptibility to AF is associated with decreased LA ejection fraction.

The described associations between AF and indexed LA max, indexed LA min and LA EF were attenuated after exclusion of rs67249485, located in an intergenic region near the *PITX2* gene¹⁷. Our results suggest rs67249485 to be the main driver of the genetic association between AF and indexed LA max and LA min as the main analyses were nullified after exclusion of rs67249485, while the Wald estimates of rs67249485 was significant. We still find a causal estimate between genetic susceptibility to AF and LA EF after exclusion of this variant, which may suggest that other genetic variants may also contribute to the genetic association between AF and LA EF. The validity of rs67249485 as important driver in the association between AF and LA size and function is statistically supported by several sensitivity analyses which indicate that the large effect of this genetic variants is very unlikely caused by measurement error, uneven population distribution or incorrect direction of causality. The biological role of *PITX2* in AF development has been extensively studied and many potential mechanisms have been suggested, including deviations in LA myocyte automaticity, impaired response to oxidative stress, inflammation and a role in the embryonic development of the heart^{18–21}. The *PITX2* gene does not only increase the risk of AF development, but has been suggested as a determinant in the success of pulmonary vein ablation in preventing AF recurrence as well²². Our results provide evidence for another possible biological consequence of *PITX2*, as we show that LA volumes increase and LA ejection fraction decreases through the AF increasing T allele of rs67249485. However, further experimental validation is needed to investigate details of the mechanisms underlying the association of rs67249485, *PITX2*, AF and LA size and function.

One cardiovascular risk factor that could potentially affect our results is hypertension, as blood pressure is known to affect both AF and LA size and function^{12–14}. We therefore performed additional multivariable MR analyses and find that the described associations between AF and LA size and function are independent of systolic blood pressure, diastolic blood pressure and pulse pressure²³.

Our study has several strengths. The strengths include the use state-of-art genetic and CMR data. The MR design is less susceptible to confounding and strongly contributes to previous work in the field²⁴. We excluded individuals with known prevalent AF and the MR was designed to study the effect of increased AF risk on LA dimensions before onset of the disease. Extensive sensitivity analyses were performed to further reduce the risk of pleiotropy and reversed causation and support our hypothesis.

Some limitations should be noted as well. First, the genetic variants used as proxy for AF explained approximately 7% of AF variance, which is a proportion of total genetic variance of 62% that has been suggested in a previous twin study²⁵. We note that we did not include all previously established genetic variants associated with AF as the UK Biobank was used as discovery cohort in the most recent GWAS of AF¹⁷. We therefore took forward the largest set of genetic variants using effect sizes obtained without the UK Biobank to limit overlap of the exposure and outcome cohorts. In addition, a part of the heritability of AF and LA size is still unknown

Figure 1. Summary MR estimates of the causal association between AF and LA size and function. The figure displays the MR estimates on the association between AF and body surface area indexed left atrial maximal volume (LA max), minimal volume (LA min), stroke volume and ejection fraction. Inverse-variance-weighted (random effects) model, MR-Egger, MR pleiotropy residual sum and outlier (MR-PRESSO), weighted median, weighted mode-based estimator and MR-Mix are shown. Outlier-corrected MR-PRESSO estimates are not included, since no genetic variants were removed in the MR-PRESSO analyses. On the X-axis, the beta coefficient and its upper and lower bound standard error are shown. The main analyses, i.e. inverse-variance-weighted random effects under the scenario of balanced horizontal pleiotropy or MR-Egger estimate under the scenario of unbalanced horizontal pleiotropy, are underlined per outcome. We considered a stringent two-sided Bonferroni corrected $P < 0.05/7$ outcomes statistically significant for the main analyses. Significant results for the main analysis are annotated with a single asterisk (*). A P -value threshold of $P < 0.05$ was adopted for the sensitivity MR analyses. Significant sensitivity MR analyses are annotated with a double asterisk (**). SE denotes standard error. The plot was made using the forestplot package (version 1.10.1, <https://CRAN.R-project.org/package=forestplot>) in R (version 3.6.3)⁵⁹.

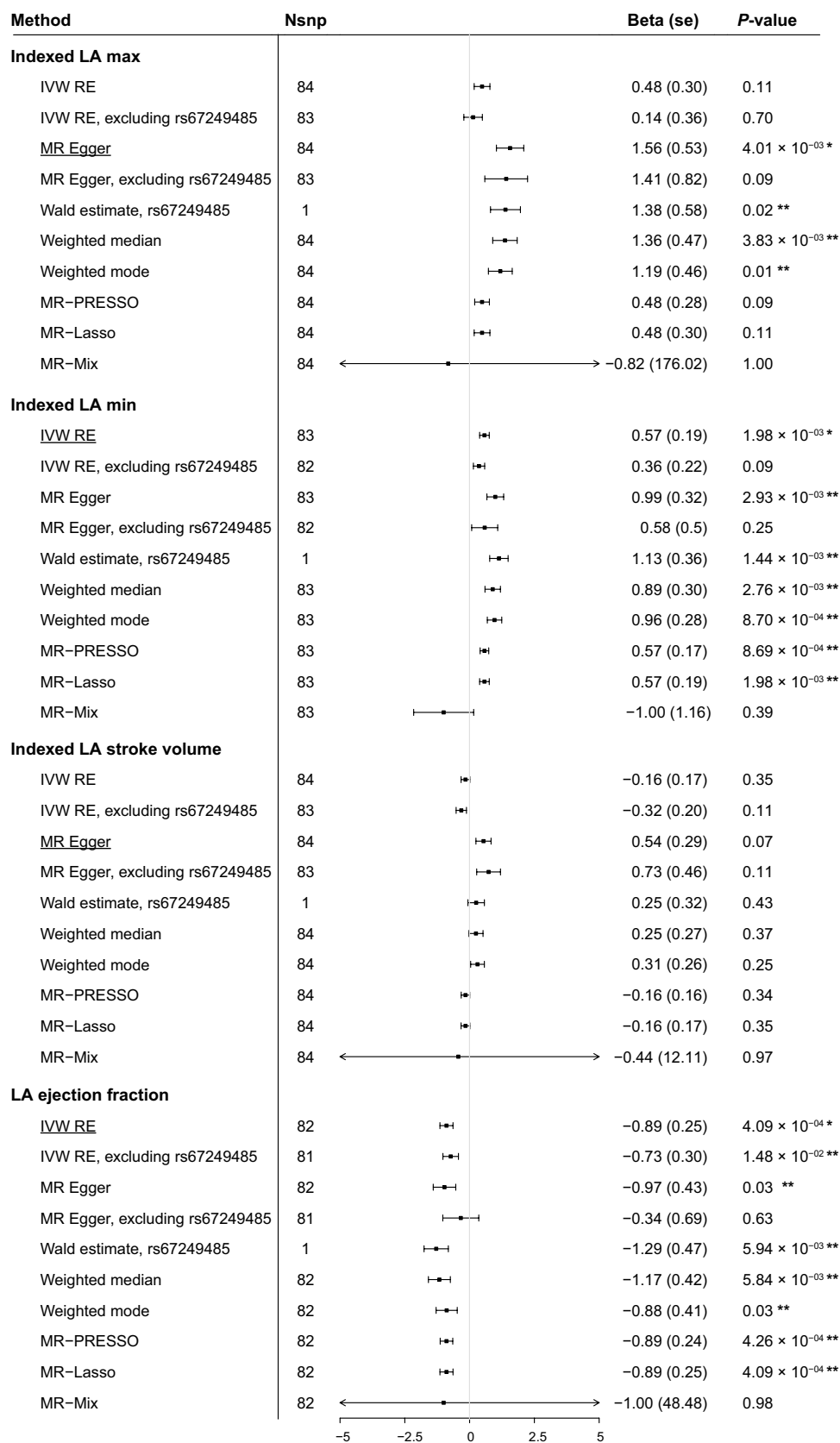
and there remains a gap between SNP-based and classic heritability estimates²⁶. Several reasons for the missing heritability have been hypothesized, including the focus of GWAS on common genetic variants and the inclusion of individuals that are mainly from European descent²⁶. In addition, GWAS assumes an additive model which overlook epistatic effects and possible interactions between genetics and the environment²⁶. Further research to the genetics of AF by studying whole exome sequencing data^{27,28}, expanding the reference genome with other ancestries²⁹, gene–gene^{30,31} and gene–lifestyle interaction^{32,33} could increase our insights in AF and consequently the certainty of the described genetic association between AF and LA size and function. We did not have data on LA volume at the onset of atrial contraction and were therefore unable to differentiate the effect of AF on the LA conduit and pump function separately. Pleiotropy cannot be ruled out completely despite rigorous sensitivity analyses. We were unable to perform a bidirectional MR to further entangle the cause and consequence in the association between AF and LA size and function as the current cohort is too small to identify robustly associated genetic variants. Lastly, the AF associated variants were obtained from a multi-ethnic GWAS meta-analysis, while the outcome cohort included individuals that were mainly from European descent. Population stratification could introduce confounding in the MR analyses through hidden population structure if the ancestry is correlated with both the phenotypes and genotypes³⁴. However, we believe this to be unlikely given the stringent adjustments for genetic ancestry in the GWAS of AF and in the regression analyses on atrial size and function³⁵.

In conclusion, we provide evidence that a higher genetic susceptibility to AF increases indexed LA max and LA min, while it decreases LA EF. We pinpoint that the genetic variant rs67249485, near the *PITX2* gene, drives the association between AF and indexed LA max and LA min and contributes strongly to the genetic association between AF and LA EF. The association between AF and LA EF was robust to multiple sensitivity analyses and indicate that genetic susceptibility to AF causally decreases LA EF.

Methods

Study population. The UK Biobank is a large, population-based cohort that included 503,325 individuals via general practitioners of the UK National Health Service (NHS) between 2006 and 2010. Informed consent was obtained from all included individuals and the North West Multi-centre Research Ethics Committee approved of the study and the North West Multi-centre Research Ethics Committee approved of the study³⁶. The UK Biobank study has been carried out in accordance with relevant guidelines and regulations and has approval from all relevant institutional review boards, including the North West Multi-centre Research Ethics Committee for the UK, the National Information Governance Board for Health and Social Care for England and Wales, and the Community Health Index Advisory Group for Scotland³⁶. Hospital episode statistics were available up to 31-03-2017 for English participants, 29-02-2016 for Welsh participants and 31-10-2016 for Scottish participants. Individuals with contoured CMR data, as previously performed by Petersen et al., were included in the current study³⁷. Individuals were excluded in case of missing information on body surface area or any covariates (please see below), failure of genetic quality control (including heterozygosity, high missingness and a discrepancy between reported and inferred gender), familial relatedness, or a medical history of mitral valve disease, heart failure, valvular surgery, pulmonary hypertension or prevalent AF at the time of CMR. Definitions of prevalent incident and incident disease are presented in Supplementary Table 1 and a flowchart depicting the study sample selection is shown in Supplementary Fig. 1.

Left atrial size and function. CMR protocol and image analyses of left atrial dimensions have been described previously¹⁰. In brief, all CMR examinations in UK Biobank were performed on a clinical wide bore 1.5 T scanner (MAGNETOM Aera, Syngo Platform VD13A, Siemens Healthcare, Erlangen, Germany) in Cheddle, United Kingdom. The LA dimensions were manually analyzed by two core laboratories based in London and Oxford and the returned volumes were used in the current study³⁷. In each CMR examination, endocardial LA contours were manually traced at end-systole (maximal LA area) and end-diastole (minimal LA area) in the HLA (4-chamber) view and VLA (2-chamber) view. The biplane method was applied to calculate maximal and minimal areas. Maximal LA volume (LA max volume) is defined as the end of left ventricular systole. Minimal LA volume (LA min volume) is defined as the end of left ventricular diastole. LA stroke volume and LA ejection fraction were calculated as followed: LA stroke volume = $(LA\ max - LA\ min)$ and LA ejection fraction = $100 \times (LA\ max - LA\ min) / (LA\ max)$.



LA volumes (LA max and LA min and LA stroke volume) were indexed to body surface area (BSA) to account for body size as well as gender differences¹². We took forward these seven outcomes to evaluate the association between AF associated genetic variants and LA size and function. As sensitivity analyses, we log-transformed LA min (indexed and non-indexed) to account for right skewness.

Genotype and imputed data. The Wellcome Trust Centre for Human Genetics performed genotyping and quality control before imputation in the individuals of UK Biobank, and imputed to HRC v1.1 panel. The quality control of samples and variants, and imputation was previously described in detail³⁸.

Genetic variants: atrial fibrillation. In this study, 111 genetic variants associated with AF (P -value $< 5 \times 10^{-8}$) from the prior GWAS of Nielsen et al. were selected as genetic instruments in current analyses³⁹. The effect sizes of the genetic variants associated with AF within the independent cohorts of the Broad AF Study, BBJ, EGCUT, PHB, SiGN and the Vanderbilt AF Registry published by Roselli et al. were used (number of cases = 32,957, number of controls = 83,546)¹⁷. We opted for this approach to obtain one of the largest sets of robust AF genetic instruments, while also being able to use effect sizes that were independent of the UK Biobank to limit overlap of the exposure and outcome cohorts. One genetic variant (rs17005647) was a priori removed as we were unable to precisely calculate the beta with the provided odds ratio of 1.0.

Genetic variants: left atrial size and function. Effect estimates of the AF associated genetic variants on LA size and function were obtained from all individuals included in the current study. Effect sizes were obtained by performing linear regression analyses on LA size and function, which were corrected for age during the imaging visit, sex, 30 principal components and genotyping array.

Genetic variants: blood pressure traits. Effect estimates of the AF associated genetic variants on systolic blood pressure, diastolic blood pressure and pulse pressure were obtained from a cohort of 408,212 unrelated individuals from the UK Biobank that were not included in the estimates of LA size and function. Systolic and diastolic blood pressure values were obtained during the baseline visit through two automated and/or two manual blood pressure measurements and the average of all measurements was used. The automated measurements were corrected according to previously described methodology⁴⁰. Pulse pressure was calculated by subtracting diastolic from systolic blood pressure. Blood pressure altering medication use was taken into account by adding respectively 15, 10 mmHg and 5 mmHg to the blood pressure trait⁴¹. Effect sizes were obtained by performing linear regression analyses, which were corrected for age during the baseline visit, sex, 30 principal components and genotyping array.

Mendelian randomization analysis. The genetic variants were tested for weak instrument bias (F-statistic) and reversed causation (MR-Steiger). F-statistics were calculated per genetic variant using the following formula: $F = R^2(n-2)/(1-R^2)$. Here, n is the sample size of the exposure and R^2 is the amount of variance of the exposure explained by the genetic variant⁴². An F-statistic < 10 was considered to indicate weak-instrument bias and these genetic variants were removed from further analyses. Reversed causation was assessed through MR-Steiger filtering and genetic variants with a significantly higher ($P < 0.05$) R^2 for the outcome than for the exposure were removed⁴³. The R^2 for AF (on the liability scale)⁴⁴ and linear outcomes⁴⁵ were calculated based on the summary statistics provided in Supplementary Table 1 using previously established formulae.

MR estimates were generated using inverse-variance weighted random effects meta-analysis. The Rucker framework was applied to assess heterogeneity and thus potential pleiotropy within the MR effect estimates⁴⁶. Balanced horizontal pleiotropy was assessed by calculating Cochran's Q ($P < 0.05$) and I^2 index ($> 25\%$) as indicators of heterogeneity within the IVW model⁴⁷. Potential unbalanced pleiotropy was assessed by performing MR-Egger regression as the MR-Egger allows for a non-zero intercept⁴⁸. The Rucker framework then assesses the difference between heterogeneity within the IVW effect estimate (Cochran's Q) and heterogeneity within the MR-Egger regression (Rucker's Q), called Q-Q'. A significant Q-Q' ($P < 0.05$), in combination with a significant non-zero intercept of the MR-Egger regression ($P < 0.05$), was considered to indicate unbalanced horizontal pleiotropy. Under this scenario, we report the MR-Egger effect estimates as it provides a causal estimate if the general InSIDE (Instrument Strength Independent of Direct Effect) assumption holds⁴⁸. Weak instrument bias within the MR-Egger regression was assessed by I^2_{GX} . An $I^2_{GX} > 95\%$ was considered low risk of weak instrument bias within the MR-Egger estimates⁴⁹. The main analysis consisted of either the IVW-RE (under the scenario of balanced horizontal pleiotropy) or the MR-Egger estimate (under the scenario of unbalanced horizontal pleiotropy).

Additional sensitivity analyses included the Mendelian randomization-Pleiotropy Residual Sum and Outlier (MR-PRESSO)⁵⁰, MR-Lasso⁵¹, leave-one-out analyses^{52,53}, weighted median⁵⁴, weighted mode⁵⁵ and MR-Mix⁵⁶, multivariable MR-IVW²³, multivariable MR-Egger⁵⁷ and multivariable MR-PRESSO⁵⁰. These all have their own strength and weaknesses and jointly provide information on the possibility of a true causal relationship. Outlier robust methods include MR-PRESSO (excludes outliers), leave-one-out analyses (excludes genetic variants one by one and re-performs IVW and MR-Egger analyses) and MR-Lasso (downweights outliers). Weighted median (majority valid), weighted mode and MR-MIX (plurality valid) generally have the potential to estimate true causal effects when larger proportions of genetic variants violate MR assumptions (generally at the cost of power). The multivariable MR-IVW²³, multivariable MR-Egger⁵⁷ and multivariable MR-PRESSO⁵⁰ analyses were performed to correct for the potential influence of systolic blood pressure, diastolic blood pressure and pulse pressure in the causal association between AF and LA size and function¹²⁻¹⁴. Effect estimates for blood pressure traits were obtained in an independent cohort from the UK Biobank (See: Genetic variants: blood pressure traits). Weak instrument bias within the multivariable MR setting was considered unlikely if Q_{x1} and Q_{x2} were larger than the

critical value at the χ^2 , calculated by subtracting one degree of freedom from the amount of SNPs at a P value of 0.05²³. Potential pleiotropy within the multivariable MR setting was assessed using the Q_a , which was considered to indicate potential pleiotropy when larger than the critical value on the χ^2 distribution as calculated by the amount of SNPs minus two degrees of freedom at a P value of 0.05²³. A multivariable MR-Egger intercept with a P value < 0.05 was considered a robust proof of unbalanced horizontal pleiotropy and the MR-Egger regression to provide a robust causal estimate⁵⁷.

Causal effect estimates are reported in β values, since LA volumes and fractions are continuous variables. The main analyses were considered significant at a Bonferroni corrected $\alpha = 0.05/7$ outcomes. For the sensitivity analyses, we adapted $\alpha = 0.05$ to ascertain statistical significance when replicating the findings of the main analysis. Continuous variables are displayed as mean \pm standard deviation when normally distributed and as median and interquartile ranges when skewed. Categorical variables are displayed as percentages. Regression analyses to obtain genetic variant-outcome associations were performed using statistical software STATA 15 (StataCorp LP)⁵⁸. MR analyses were performed using R (version 3.6.3)⁵⁹, the TwoSampleMR package 0.5.3⁶⁰, MR-PRESSO (version 1.0)⁵⁰, MR-Lasso⁵¹, MR-mix (version 0.1.0)⁵⁶, MendelianRandomization (version 0.5.0)⁶¹ and MVMR (version 0.3)²³.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 19 November 2020; Accepted: 30 March 2021

Published online: 19 April 2021

References

1. Chugh, S. S. *et al.* Worldwide epidemiology of atrial fibrillation: A global burden of disease 2010 study. *Circulation* **129**, 837–847 (2014).
2. Kirchhof, P. *et al.* 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur. Heart J.* **37**, 2893–2962 (2016).
3. Zhuang, J. *et al.* Association between left atrial size and atrial fibrillation recurrence after single circumferential pulmonary vein isolation: A systematic review and meta-analysis of observational studies. *Eurospace*. **14**, 638–645 (2012).
4. Benjamin, E. J., D'Agostino, R. B., Belanger, A. J., Wolf, P. A. & Levy, D. Left atrial size and the risk of stroke and death: The Framingham Heart Study. *Circulation* **92**, 835–841 (1995).
5. Flaker, G. C. *et al.* Clinical and echocardiographic features of intermittent atrial fibrillation that predict recurrent atrial fibrillation. *Am. J. Cardiol.* **76**, 355–358 (1995).
6. Nattel, S., Burstein, B. & Dobrev, D. Atrial remodeling and atrial fibrillation: Mechanisms and implications. *Circ. Arrhythm. Electrophysiol.* **1**, 62–73 (2008).
7. Casacang-Verzosa, G., Gersh, B. J. & Tsang, T. S. M. Structural and functional remodeling of the left atrium. Clinical and therapeutic implications for atrial fibrillation. *J. Am. Coll. Cardiol.* **51**, 1–11 (2008).
8. Allesie, M. A. Atrial electrophysiologic remodeling: Another vicious circle? In *Journal of Cardiovascular Electrophysiology* vol. 9 1378–1393 (Futura Publishing Company Inc., 1998).
9. Wijffels, M. C. E. F., Kirchhof, C. J. H. J., Dorland, R. & Allesie, M. A. Atrial fibrillation begets atrial fibrillation: A study in awake chronically instrumented goats. *Circulation* **92**, 1954–1968 (1995).
10. Lang, R. M. *et al.* Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J. Am. Soc. Echocardiogr.* **28**, 1–39.e14 (2015).
11. Davey Smith, G. & Hemani, G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–98 (2014).
12. Dzeshka, M. S., Shantsila, A., Shantsila, E. & Lip, G. Y. H. Atrial fibrillation and hypertension. *Hypertension* **70**, 854–861 (2017).
13. Eshoo, S., Ross, D. L. & Thomas, L. Impact of mild hypertension on left atrial size and function. *Circ. Cardiovasc. Imaging* **2**, 93–99 (2009).
14. Vaziri, S. M., Larson, M. G., Lauer, M. S., Benjamin, E. J. & Levy, D. Influence of blood pressure on left atrial size: The Framingham Heart Study. *Hypertension* **25**, 1155–1160 (1995).
15. Anwar, A. M., Geleijnse, M. L., Soliman, O. I. I., Nemes, A. & Ten Cate, F. J. Left atrial Frank–Starling law assessed by real-time, three-dimensional echocardiographic left atrial volume changes. *Heart* **93**, 1393–1397 (2007).
16. Tenekecioglu, E. *et al.* Disturbed left atrial function is associated with paroxysmal atrial fibrillation in hypertension. *Arq. Bras. Cardiol.* **102**, 253–261 (2014).
17. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* **50**, 1225–1233 (2018).
18. Mommersteeg, M. T. M. *et al.* Pitx2c and Nkx2-5 are required for the formation and identity of the pulmonary myocardium. *Circ. Res.* **101**, 902–909 (2007).
19. Kirchhof, P. *et al.* PITX2c is expressed in the adult left atrium, and reducing Pitx2c expression promotes atrial fibrillation inducibility and complex changes in gene expression. *Circ. Cardiovasc. Genet.* **4**, 123–133 (2011).
20. Wang, J. *et al.* Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9753–9758 (2010).
21. Tao, G. *et al.* Pitx2 promotes heart repair by activating the antioxidant response after cardiac injury. *Nature* **534**, 119–123 (2016).
22. Shoemaker, M. B. *et al.* Genetic susceptibility for atrial fibrillation in patients undergoing atrial fibrillation ablation. *Circ. Arrhythm. Electrophysiol.* **13**, e007676 (2020).
23. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int. J. Epidemiol.* **48**, 713–727 (2019).
24. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease?. *Int. J. Epidemiol.* **32**, 1–22 (2003).
25. Christophersen, I. E. *et al.* Familial aggregation of atrial fibrillation: A study in danish twins. *Circ. Arrhythm. Electrophysiol.* **2**, 378–383 (2009).
26. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
27. Petersen, B. S., Fredrich, B., Hoepfner, M. P., Ellinghaus, D. & Franke, A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet.* **18**, Article 14 (2017).

28. Wainschein, P. *et al.* Recovery of trait heritability from whole genome sequence data. *bioRxiv* <https://doi.org/10.1101/588020> (2019).
29. Sherman, R. M. *et al.* Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* **51**, 30–35 (2019).
30. Hu, J. K., Wang, X. & Wang, P. Testing gene–gene interactions in genome wide association studies. *Genet. Epidemiol.* **38**, 123–134 (2014).
31. Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10**, 392–404 (2009).
32. Abdullah Said, M., Verweij, N. & Van Der Harst, P. Associations of combined genetic and lifestyle risks with incident cardiovascular disease and diabetes in the UK biobank study. *JAMA Cardiol.* **3**, 693–702 (2018).
33. Said, M. A. *et al.* Contributions of interactions between lifestyle and genetics on coronary artery disease risk. *Curr. Cardiol. Rep.* **21**, Article 89 (2019).
34. Vanderweele, T. J., Tchetgen Tchetgen, E. J., Cornelis, M. & Kraft, P. Methodological challenges in Mendelian randomization. *Epidemiology* **25**, 427–435 (2014).
35. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
36. *UK Biobank Ethics and Governance Framework*. <https://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf> (2012). Accessed 7 Mar 2019.
37. Petersen, S. E. *et al.* Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. *J. Cardiovasc. Magn. Reson.* **19**, 18 (2017).
38. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
39. Nielsen, J. B. *et al.* Genome-wide study of atrial fibrillation identifies seven risk loci and highlights biological pathways and regulatory elements involved in cardiac development. *Am. J. Hum. Genet.* **102**, 103–115 (2018).
40. Stang, A. *et al.* Algorithms for converting random-zero to automated oscillometric blood pressure values, and vice versa. *Am. J. Epidemiol.* **164**, 85–94 (2006).
41. Tobin, M. D., Sheehan, N. A., Scurrah, K. J. & Burton, P. R. Adjusting for treatment effects in studies of quantitative traits: Anti-hypertensive therapy and systolic blood pressure. *Stat. Med.* **24**, 2911–2935 (2005).
42. Palmer, T. M. *et al.* Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat. Methods Med. Res.* **21**, 223–242 (2012).
43. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
44. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**(3), 214–224 (2012).
45. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
46. Bowden, J. *et al.* A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat. Med.* **36**, 1783–1802 (2017).
47. Del Greco, M. F., Minelli, C., Sheehan, N. A. & Thompson, J. R. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat. Med.* **34**, 2926–2940 (2015).
48. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
49. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: The role of the I² statistic. *Int. J. Epidemiol.* **45**, dyw220 (2016).
50. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
51. Rees, J. M. B., Wood, A. M., Dudbridge, F. & Burgess, S. Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. *PLoS ONE* **14**, e0222362 (2019).
52. Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* **36**, 111–133 (1974).
53. Corbin, L. J. *et al.* BMI as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using mendelian randomization. *Diabetes* **65**, 3002–3007 (2016).
54. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
55. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
56. Qi, G. & Chatterjee, N. Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nat. Commun.* **10** Article 1941 (2019).
57. Rees, J. M. B., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat. Med.* **36**, 4705–4718 (2017).
58. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC. (2017).
59. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (2020).
60. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7** Article e34408 (2018).
61. Yavorska, O. O. & Burgess, S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).

Acknowledgements

We would like to thank the Centre for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high-performance computing cluster. We thank Ruben N. Eppinga, MD PhD, Tom Hendriks, MD, M. Abdullah Said, MD, M. Yldau van der Ende, MD PhD, Yanick Hagemeyer, MSc, Luis Juarez Orozco, MD PhD, Jan-Walter Benjamins, BEng, and Ming W. Yeung, MSc (Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands), for their contributions to the extraction and processing of data in the UK Biobank.

Author contributions

All authors contributed to the study conception and design. Formal analyses were performed by Y.J.V. Y.J.V., J.E.S., M.R., P.H. were involved in interpreting the data. Writing of the first draft was performed by J.E.S. and Y.J.V. Supervision, funding, manuscript reviewing and editing were performed by M.R., P.H. All authors read and approved the final manuscript. All authors take full responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

Funding

This work was supported by funding of the Netherlands Heart Foundation [Grant NHS2010B233]; and the Netherlands Heart Foundation [Grant NHS2010B280]; and the Netherlands Organization for Health Research and Development [Grant 95103007]; and the Dutch Kidney Foundation [Grant E0.13].

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-87859-8>.

Correspondence and requests for materials should be addressed to P.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021