

1 **Unbiased characterization of atrial fibrillation phenotypic architecture provides insight to**
2 **genetic liability and clinically relevant outcomes**

3
4 Giovanni Davogustto,¹ Shilin Zhao,² Yajing Li,² Eric Farber-Eger,¹ Brandon D. Lowery,¹ Lauren
5 Lee Shaffer,¹ Jonathan D. Mosley,^{1,3} M. Benjamin Shoemaker,¹ Yaomin Xu,^{2,3} Dan M. Roden,^{1,3}
6 Quinn S. Wells^{1,3}

7
8 Short title: Phenotypic clusters of atrial fibrillation

9
10 ¹ Department of Medicine, Division of Cardiovascular Medicine, Vanderbilt University Medical
11 Center, Nashville TN

12 ² Department of Biostatistics, Vanderbilt University Medical Center, Nashville TN

13 ³ Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville TN

14 Word count: 3,520

15

1 **ABSTRACT:**

2 Background:

3 Atrial Fibrillation (AF) is a common and clinically heterogeneous arrhythmia. Machine
4 learning (ML) algorithms can define data-driven disease subtypes in an unbiased fashion, but
5 whether the AF subgroups defined in this way align with underlying mechanisms, such as high
6 polygenic liability to AF or inflammation, and associate with clinical outcomes is unclear.

7 Methods:

8 We identified individuals with AF in a large biobank linked to electronic health records
9 (EHR) and genome-wide genotyping. The phenotypic architecture in the AF cohort was defined
10 using principal component analysis of 35 expertly curated and uncorrelated clinical features. We
11 applied an unsupervised co-clustering machine learning algorithm to the 35 features to identify
12 distinct phenotypic AF clusters. The clinical inflammatory status of the clusters was defined
13 using measured biomarkers (CRP, ESR, WBC, Neutrophil %, Platelet count, RDW) within 6
14 months of first AF mention in the EHR. Polygenic risk scores (PRS) for AF and cytokine levels
15 were used to assess genetic liability of clusters to AF and inflammation, respectively. Clinical
16 outcomes were collected from EHR up to the last medical contact.

17 Results:

18 The analysis included 23,271 subjects with AF, of which 6,023 had available genome-
19 wide genotyping. The machine learning algorithm identified 3 phenotypic clusters that were
20 distinguished by increasing prevalence of comorbidities, particularly renal dysfunction, and
21 coronary artery disease. Polygenic liability to AF across clusters was highest in the low
22 comorbidity cluster. Clinically measured inflammatory biomarkers were highest in the high
23 comorbid cluster, while there was no difference between groups in genetically predicted levels of

1 inflammatory biomarkers. Subgroup assignment was associated with multiple clinical outcomes
2 including mortality, stroke, bleeding, and use of cardiac implantable electronic devices after AF
3 diagnosis.

4 Conclusion:

5 Patient subgroups identified by unsupervised clustering were distinguished by
6 comorbidity burden and associated with risk of clinically important outcomes. Polygenic liability
7 to AF across clusters was greatest in the low comorbidity subgroup. Clinical inflammation, as
8 reflected by measured biomarkers, was lowest in the subgroup with lowest comorbidities.
9 However, there were no differences in genetically predicted levels of inflammatory biomarkers,
10 suggesting associations between AF and inflammation is driven by acquired comorbidities rather
11 than genetic predisposition.

1 INTRODUCTION

2 Atrial fibrillation (AF) is the most common arrhythmia encountered in clinical practice,¹
3 and is associated with increased mortality,² stroke,³ heart failure (HF),³ myocardial infarction,^{4,5}
4 dementia,⁶ and decreased quality of life.⁷ AF is a complex, dynamic, and heterogeneous condition,
5 and there is considerable variation in the age at onset, arrhythmia burden, presence of risk factors,
6 severity of symptoms, and incidence of complications among affected individuals.⁸⁻¹⁸
7 Heterogeneity is likewise evident in response to AF treatments, including antiarrhythmic drugs
8 and catheter-based ablation.^{19,20} This interindividual variation suggests complex gene-
9 environment interactions and the existence of multiple AF subtypes with differing mechanistic
10 underpinnings.

11 Two sources of risk that may contribute to AF heterogeneity, and by extension, AF
12 subgroups, are genetic susceptibility to AF and inflammation, which itself may be due to acquired
13 or genetic factors. Indeed, polygenic liability conferred by common variants is associated with
14 risk of AF and multiple clinical outcomes.²¹⁻²⁶ Additionally, multiple lines of evidence support a
15 significant role for inflammation in the pathophysiology of AF. For example, elevated levels of
16 inflammatory biomarkers²⁷ and polymorphisms in genes encoding inflammatory cytokines²⁸⁻³¹ are
17 associated with AF risk; and inflammatory infiltrates are observed in atria of AF patients.³²
18 Moreover, many AF risk factors (e.g., HF, coronary artery disease [CAD], hypertension [HTN],
19 obesity) and biologic mechanisms (e.g., fibrosis, thrombogenesis) are themselves associated with
20 increased inflammation,^{25,26,33} and AF itself can promote inflammation and perpetuate atrial
21 remodeling and arrhythmia.^{34,35} However, the contributions of genetic susceptibility to AF and
22 inflammation, both acquired and genetically driven, to AF phenotypic subgroups has not been
23 reported.

1 The emergence of advanced machine learning (ML) methods has facilitated data-driven
2 subtyping of complex, heterogeneous cardiovascular diseases such as HTN,³⁶ HF,³⁷ and
3 pulmonary arterial hypertension.³⁸ ML-derived phenotypic AF subgroups have been identified in
4 post hoc analyses in observational registries.^{39,40} These studies, however, did not explore the
5 mechanistic or genetic basis of subgroups. Therefore, there is no consensus regarding the
6 classification of AF subtypes based on their underlying mechanisms, particularly with respect to
7 patients in clinical care.^{41,42}

8 We hypothesized that contemporary ML methods applied to a large clinical AF population
9 will identify distinct subtypes with different polygenic liability of AF, inflammatory state, and
10 clinical outcomes.

1 MATERIALS AND METHODS

2 The study protocol was approved by the Vanderbilt University Medical Center (VUMC)
3 Institutional Review Board (VUMC IRB# 181403).

4

5 Data source:

6 Clinical data were derived from the Synthetic Derivative (SD), a de-identified version of the
7 Vanderbilt University Medical Center's Electronic Health Record (EHR) intended to support
8 research.⁴³⁻⁴⁵ For the subset of participants with genetic information, genotype data were
9 obtained from BioVU, the Vanderbilt biobank linked to the SD that contains DNA samples and
10 genetic data derived from clinical blood samples that would otherwise be discarded.^{46,47}

11

12 Study population:

13 Subjects with AF were identified using a random-forest classifier applied to routinely collected
14 clinical features. A list of features, including International Classification of Diseases codes (ICD
15 codes), Current Procedural Terminology (CPT) codes, text strings, and medications were used in
16 the classifier (**Table S1**). Every instance of each feature was identified and timestamped across
17 the EHR for each subject. The instance/date pairs were used to generate 3 scoring metrics for
18 each feature: strength (number of distinct days with a mention), persistence (number of days
19 between first and last mention of a feature), and durability (ratio between feature persistence and
20 total record length). The date of the first qualifying feature (e.g., code or text mention) for atrial
21 fibrillation was used as date of AF first recognition for each subject. The classifier was trained
22 and validated locally and demonstrated superior performance compared to conventional
23 phenotyping approaches such as ICD codes. Analyses were limited to subjects receiving ongoing

1 care at our institution, defined as ≥ 3 outpatient visits to primary care or cardiology over 5 years.
2 Clustering was based on feature values at the time of AF first recognition in the EHR. Subjects
3 with missing predictors or outcome data were excluded. When available, genome-wide
4 genotyping data for subjects was obtained from BioVU. Genetic analyses were restricted to
5 subjects of European and African ancestries due to the small number (n=48) of subjects from
6 other ancestries with genotype data. A summary of inclusion and exclusion criteria is presented
7 in **Figure S1**.

8
9 Predictor features:

10 Clustering used 35 clinical features selected as relevant for AF by two cardiologists
11 (GD, QSW), and included demographics, body mass index (BMI), elements of the CHA2DS2-
12 VASc score, sleep apnea, presence of cardiac implantable electronic devices, left ventricular
13 systolic dysfunction, thyroid disease, and valvular heart disease. Features were extracted from
14 the de-identified EHR using a variety of techniques including queries of demographics,
15 anthropometric measures, diagnosis and procedure codes, and natural language processing of
16 clinical notes and procedure reports. A description of features and definitions is presented in
17 Supplemental File 1. Only features prior to or on the day of AF first recognition were utilized
18 for clustering. An exception was made for BMI due to sparsity of weight/height measures,
19 where the closest value to AF first recognition - before or after - was used. As described
20 below, an initial pairwise analysis confirmed that the 35 selected features were not themselves
21 highly correlated.

22

23

1 Outcomes:

2 Clinical outcomes included mortality, bleeding-related hospitalization, stroke,
3 pacemaker implantation, implantable cardioverter defibrillator (ICD) implantation,
4 cardioversion, and AF ablation. Inflammatory burden in clusters was assessed using clinical
5 inflammatory biomarker levels ascertained within 6 months of first AF mention. Polygenic
6 risk scores, derived from published GWAS, were used to assess genetic liability to AF and
7 imputed inflammatory cytokine levels (further details below). Outcomes were required to first
8 appear in the medical record at least one day after the first mention of AF. A full description of
9 outcomes and definitions is presented in Supplemental File 2.

10
11 Genetic data, AF polygenic risk score calculation, and imputed cytokine calculation

12 Genotyping data for BioVU subjects were acquired using the Illumina Multi-Ethnic
13 Genotyping Array (MEGA) platform. QC was performed using PLINK v1.90 β 6.17, and
14 included duplicate removal, sex-check, and removal of one of each pair of related individuals
15 (π -hat>0.2). Data were then pre-processed using the HRC-1000G-check tool v4.2.11 and
16 imputed using the Michigan Imputation server and the 10/2014 release of the 1,000 Genomes
17 cosmopolitan reference haplotypes.

18 Genetic predisposition to AF within clusters was assessed using a previously published
19 polygenic risk score (PRS) that comprises 1,168 SNPs
20 (<https://www.pgscatalog.org/score/PGS000035/>).⁴⁸ In our dataset, 8 SNPs did not pass QC and
21 hence the PRS value was computed by summing the product of allele weighting and allele
22 dosage across 1,160 available SNPs. The raw scores were then normalized to the sample mean

1 and standard deviation. The SNPs and associated weights included in the AF PRS are presented
2 in Supplemental File 3.

3 Genetic predisposition for inflammation between clusters was assessed using genetically
4 predicted cytokine levels, as previously described by Mosley et al.⁴⁹ Briefly, using published
5 GWAS summary statistics of measured cytokine levels, we calculated a PRS value (“imputed
6 cytokine level”) of BioVU subjects for each cytokine of interest with at least one SNP meeting
7 genome-wide significance (GWS) in published summary statistics.⁴⁹ Again, PRS were computed
8 by summing the product of allele weighting and allele dosage across the SNPs passing GWS.

9 Imputed cytokine levels of 41 cytokines that met inclusion criteria were created. To
10 validate that imputed cytokine levels were capturing a clinically relevant predisposition to
11 disease, we conducted phenome-wide association studies (PheWAS) for each cytokine PRS in
12 Vanderbilt’s de-identified EHR. Of 41 imputed cytokines, 13 had at least one significant
13 association (FDR <0.1) with clinical phenotype and were used in downstream analyses. A
14 summary of imputed cytokines selection is presented in Figure S2, and the results of validation
15 PheWAS of all the included cytokines are available for download in the supplemental material.

16

17 Clustering:

18 Due to requirements of the clustering algorithm, continuous variables such as age and
19 BMI were converted into categories. Age was categorized into <50, 50-70, and > 70 years at first
20 AF mention in EHR. BMI was categorized into <18 Kg/m², 18-24.99 Kg/m², 25-29.99 Kg/m²,
21 30-39.99 Kg/m², and >=40 Kg/m². Transformed data were analyzed using the BlockCluster
22 package, employing the coclusterBinary function specifically designed for binary data, and
23 therefore all feature elements were coded in the form of dummy variables. The clustering

1 methodology requires the number of clusters to be prespecified, which in this case was set to 3
2 clusters based on the distribution of features. The clustering algorithm involves estimating a
3 latent mixture block model, and simultaneously conducts block clustering in both subject and
4 features sets. The number of clusters were prespecified to 3 based on the number of features
5 used.

6 Other statistical considerations and analyses

7 Statistical analyses were performed with the R statistical program (v4.3.1), and primarily
8 using the “Blockcluster” (4.5.3), “Hmisc” (v5.1.1), and “rms” (v6.7.1) packages. Unless
9 otherwise specified, continuous variables are presented as median (interquartile range), and
10 categorical variables as counts (percentage). Differences in measured inflammatory biomarkers
11 among clusters were assessed using one way ANOVA. The associations of cluster assignment
12 with normalized AF PRS and imputed cytokine levels from genetic data were determined using
13 an ordinal regression model. The models tested the association between PRS as continuous
14 variables and cluster assignment, and included age, sex, and 4 principal components of ancestry
15 as covariates. For all these sets of tests, false discovery rates (FDRs) were calculated from the
16 nominal *p-values* using the R function `p.adjust` using the Benjamini & Hochberg option. FDRs
17 less than 0.1 were considered statistically significant. A Cox proportional hazards regression
18 model was used to test the association of cluster assignment and time-to-event mortality,
19 adjusting for age of first AF mention and CHA₂DS₂-VASc score. The functions *adjustedsurv* and
20 *adjusted_surv_quantile* of the package `adjustedCurves` (v0.10.1) were used to create the adjusted
21 survival curves and adjusted median survival estimates. For other outcomes, unadjusted cox
22 proportional hazards regression models were used to test the association of cluster assignment
23 and time-to-event outcomes.

1 RESULTS

2

3 Cohort characteristics:

4 We identified 23,271 subjects with AF meeting inclusion criteria. The cohort median age
5 was 68 years (IQR: 59-76), 42.6% female, and 90.4% self-reported White. The most prevalent
6 comorbidity was hypertension (62.3%), followed by renal dysfunction (36.8%), coronary artery
7 disease (CAD) (35.8%), and fluid-electrolyte disorders (25.9%). Heart failure was also common
8 (18.4%), and 12.1% of patients had history of left ventricular systolic dysfunction (Table 1).

9 The median follow-up after AF first mention was 5.4 (IQR:2.8-8.6) years. Mortality
10 during follow-up was 14%, with a median time to death of 4.5 years (IQR 1.6-7.8) after first AF
11 mention. Approximately half of subjects (55.2%) received antiarrhythmic drug therapy, usually
12 initiated shortly after diagnosis (median 14.6 days, IQR 0.0-401.5). Cardioversion was
13 performed in approximately one fifth of the subjects (19.9%) and AF ablation in 8.1%. Device
14 placement was common, with pacemaker implantation in 13.7% of subjects and 8.5%
15 undergoing ICD implantation. Stroke occurred in 12.3% of subjects while bleeding events were
16 rare (Table 2).

17

18 Feature exploration:

19 To gain additional insight into the phenotypic architecture of our cohort prior to
20 clustering, we conducted a series of exploratory analyses. We first assessed pairwise correlations
21 between features to identify patterns among variables, detect pairs of highly correlated features
22 that could affect clustering, and inform *a priori* estimates of the expected number of groups
23 derived by clustering. The selected features were not highly correlated (all $r < 0.5$), and all were

1 retained for clustering analyses. Based on initial visual evaluation, we estimated that clustering
2 of would result in approximately 3 distinct phenotypic subgroups based on correlation clustering
3 (Figure 1A). We then conducted data dimensionality reduction using principal component
4 analysis (PCA) and evaluation of variable contributions to components using a loading plot
5 (Figure 1B). This revealed that comorbidity burden, particularly left ventricular systolic
6 dysfunction, fluid-electrolyte disorders, coronary artery disease, and renal dysfunction, were the
7 main contributors to principal component (PC) 1, while age group was the main contributor for
8 PC2. Analysis of PC associations with mortality revealed that PC1 values were associated with
9 increased risk of death (PC1 interquartile OR 1.99, 95% CI: 1.91-2.07).

10

11 Unbiased clustering of features at AF diagnosis:

12 The co-clustering algorithm *blockcluster* groups subjects and features simultaneously to
13 identify blocks of subjects with similar characteristics. Based on results above, 3 clusters of
14 subjects were prespecified, but also identified 3 distinct groups of predictors determining cluster
15 assignment. Clusters were distinguished by increasing burden of comorbidities (Figure 2A),
16 particularly renal dysfunction, fluid-electrolyte disorders, and CAD, as shown in Figure 2B.
17 Based on comorbidity burden, we defined groups as “Low”, “Mid”, and “High” comorbidity
18 clusters. Figure S3 shows that other comorbidities demonstrated similar patterns of distribution
19 across clusters, but the magnitude was less pronounced than renal dysfunction, coronary artery
20 disease, and fluid-electrolyte disorders (Figure S3).

21

22

23 Association of clusters with clinical outcomes:

1 Unadjusted survival differed significantly among clusters (Median survival 23.2 years
2 [95% CI 22-greater than 25] for Low cluster, 16.7 years [95% CI 15.1-18.5] for Mid cluster, and
3 9.5 years [95% CI 8.3-10.9] for High cluster), with a dose-response effect observed between
4 increasing cluster comorbidity burden and worse survival (Figure 3A). A similar relationship
5 was seen with mortality when the 3 key variables determining cluster assignment (renal
6 dysfunction, CAD, fluid-electrolyte disorders) were considered (Figure 3B, Median survival >25
7 years when all these variables were absent, 20.2 years [95% CI 19.1-22.3] with one variable
8 present, and 13.6 years [95% CI 12.5-15.1] with two or three variables present). Furthermore,
9 when adjusting for age and CHA2DS2-VASc score at the time of first AF mention, cluster
10 assignment remained significantly associated with survival (adjusted median survival 22.02
11 years [95% CI 21.34-23.17] for Low cluster, 13.98 years [95% CI 13.51-14.53] for Mid cluster,
12 and 7.96 years [95% CI 7.39-8.64] for High cluster, Figure 3C).

13 Like survival, there was a significant difference in unadjusted event-rates for stroke,
14 bleeding, pacemaker placement, ICD placement, ablation, and cardioversion. Again, clusters
15 with greater burden of comorbidities had more events observed for these outcomes (Figure 4,
16 figure S4). In general, patients in the “High” cluster were more likely to have episodes of stroke
17 or bleeding and ICD implantation, while patients in the “Low” cluster were more likely to have
18 ablations and cardioversions. Patients in the “Mid” cluster were most likely to have pacemaker
19 implantation. Antiarrhythmic drug therapy was comparable across clusters (Figure S4).

20

21 Association of clusters with AF polygenic risk, measured inflammatory biomarkers, and imputed
22 cytokine levels:

1 There was a significant difference among clusters with respect to genetic liability to AF
2 assessed by normalized AF PRS after adjustment for age, sex, and first 4 principal components
3 of ancestry, with the highest genetic susceptibility to AF seen in the “Low” comorbidity cluster
4 and lowest genetic susceptibility to AF in the “High” comorbidity clusters (Figure 5).

5 We compared clusters with respect to clinically ascertained levels of C-reactive protein,
6 erythrocyte sedimentation rate, and white blood cell count measured within 6 months of AF first
7 mention. There was a significant association between AF cluster and measured inflammatory
8 biomarkers (Table 3): all biomarkers ($p < 0.001$ for each) had lowest values in the “Low”
9 comorbidity cluster and highest values in the “High” comorbidity cluster, with intermediate
10 values in “Mid” cluster.

11 The PheWAS of genetically predicted circulating cytokines identified 13 markers (IFN-
12 GR1, IL-10RB, IL-17F, IL-17RD, IL-1RL2, IL-1RN, IL-23R, IL-34, IL-37, IL6R, IL-6ST,
13 TNFRSF-11B, TNFRSF-6B) to take forward in this analysis. After accounting for multiple
14 comparison testing, there were no significant differences among clusters with respect to
15 genetically imputed levels for these inflammatory cytokines and cytokine-associated circulating
16 proteins. The normalized imputed cytokine levels values overall and by cluster are presented in
17 Supplemental File 4.

18

1 **DISCUSSION**

2 We conducted an analysis of 23,271 subjects with atrial fibrillation from a large tertiary
3 medical center that identified 3 phenotypic disease clusters associated with mortality, stroke, and
4 bleeding. Cluster membership was largely driven by comorbid diseases, in particular the
5 presence of renal dysfunction, CAD, and fluid-electrolyte disturbances. Greater comorbidity
6 burden across groups was associated with higher levels of clinically measured inflammatory
7 biomarkers and inversely associated with a polygenic predisposition for AF. There was no
8 detectable difference among groups with respect to genetic predisposition to inflammation as
9 ascertained by imputed cytokine levels.

10 Many of our findings are in agreement with prior studies reporting data-driven subgroups
11 of AF, and thereby expands and strengthens the existing evidence base.^{39,40} For example, two
12 earlier studies (n= 9,749 and 2,458 subjects, respectively), both using observational cohorts and
13 hierarchical clustering, identified 3-4 clusters, including clusters comprising subjects with young
14 age or low comorbidity burden and a cluster with CAD as driving feature. Additionally, both
15 studies reported associations between cluster assignment and major adverse cardiovascular
16 events. Like these studies, we observed a similar number of clusters, with CAD and other
17 comorbidities (renal dysfunction and fluid-electrolyte disturbances) as key drivers of cluster
18 assignment and an association of cluster assignment with mortality, stroke, bleeding,
19 cardioversion, ablation, and cardiac implantable electronic device (CIED) implant. The present
20 study adds to the evidence base of data-driven analyses of AF phenotypic architecture and
21 supports the importance of comorbidities as determinates of interindividual variability in AF
22 outcomes.

1 Our work also extends previous work in important ways. One notable distinction from
2 prior studies is the size of the study population, which comprised over 23,000 subjects from a
3 large medical center with dense phenotype data and more than 5 years of follow-up after AF first
4 mention in the EHR. This population is more than twice the size of earlier reports and was
5 sufficiently large to enable more detailed analyses of AF phenotypic architecture including
6 evaluation of a broader range of comorbidities, inflammatory markers, and polygenic risk. For
7 example, we identified renal dysfunction as an important determinate of cluster assignment, a
8 finding that has not been previously reported. This observation warrants replication in other
9 studies but suggests a previously underappreciated contribution of renal dysfunction to AF
10 phenotypic heterogeneity, especially in clinical populations.

11 The current report also provides new insights into potential mechanisms underlying AF
12 subgroups, including clinically acquired and genetically determined inflammation as well as
13 genetic susceptibility to AF itself. We observed higher levels of clinically measured
14 inflammatory biomarkers in clusters with more comorbidities. However, there were no
15 differences with respect to genetically predicted levels of inflammatory cytokines. This suggests
16 that observed epidemiologic associations between inflammation and AF are driven by acquired
17 comorbidities and/or environmental factors rather than a genetic predisposition to increased
18 inflammation. These observations could explain, at least in part, the lack of universal efficacy of
19 anti-inflammatory agents as AF therapies⁵⁰⁻⁵³, and further support the importance of
20 cardiovascular risk factor optimization in the management of AF.

21 We also observed greater polygenic risk for AF in groups with lower comorbidity burden,
22 with genetic risk highest in the lowest burden of CAD and other comorbidities. Clinically, this
23 cluster has many similarities with a group of patients previously referred to as having “lone AF”.

1 Collectively, in the context of prior literature, our findings suggest that observed phenotypic
2 variability in AF, in part, is distributed along a gradient with differing contributions of genetic
3 risk and acquired comorbidity. This framework for understanding AF heterogeneity,
4 incorporating both environmental and genetic factors, aligns with clinical experience and has
5 been previously articulated in a theoretical context.⁵⁴ However, to our knowledge, our analysis is
6 the first to empirically demonstrate the complementary roles of comorbidity and genetic liability
7 as major drivers of phenotypic variability using hypothesis-free methods.

8 Limitations

9 Even though we took steps to improve quality and completeness of EHR-derived data,
10 including restricting the analysis to subjects receiving longitudinal care in our health system,
11 misclassification and under-ascertainment of clinical phenotypes and outcomes, notably
12 mortality which may be under-estimated in the EHR, are possible due to misdiagnosis or
13 underdiagnosis. Additionally, measured biomarkers were not prospectively collected and
14 therefore are susceptible to ascertainment bias. Only a portion of patients included had genetic
15 data available which limited the power of the imputed cytokine analyses. It is also possible that
16 imputed cytokines did not capture sufficient genetic variability of cytokine levels; we attempted
17 to minimize this issue by only using imputed cytokines validated by phenome-wide scanning in
18 the analyses.

19 20 Conclusions

21 Patient subgroups identified by unsupervised clustering were associated with differing
22 risk of mortality, stroke, bleeding, and ICD implant. These groups were distinguished by
23 comorbidity burden and polygenic liability to AF, which were inversely correlated. Clinical

1 inflammation, as reflected by higher measured biomarker levels, was associated with increasing
2 comorbidity burden, but there were no differences in genetically predicted levels of
3 inflammatory biomarkers, consistent with the idea that inflammatory burden is driven by
4 acquired comorbidities.

6 **Acknowledgements:**

7 Clinical data were obtained using Vanderbilt's Synthetic Derivative. The Synthetic
8 Derivative resource is supported by CTSA award No. UL1TR000445 from the National Center
9 for Advancing Translational Sciences. Genomic data were obtained using Vanderbilt's de-
10 identified biobank BioVU. Vanderbilt University Medical Center's BioVU projects are
11 supported by numerous sources: institutional funding, private agencies, and federal grants. These
12 include NIH funded Shared Instrumentation Grant S10OD017985, S10RR025141, and
13 S10OD025092; CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. The contents
14 of this publication are solely the responsibility of the authors and do not necessarily represent
15 official views of the National Center for Advancing Translational Sciences or the National
16 Institutes of Health.

18 **Sources of Funding:**

19 This project was supported by Vanderbilt's participation in the American Heart
20 Association's Atrial Fibrillation Strategically Focused Research Network (18SFRN34110369
21 and 18SFRN34230089). This project also received support from NIH grants R01 HL149826
22 (DMR), R01 HL142856 (JDM).

23

1 **Disclosures:**

2 None.

3 **Data and Materials Availability:**

4 Data used in this study are available upon request to the corresponding author and after
5 approval from institutional review.

Literature Cited

- 1
2
- 3 1. Colilla S, Crow A, Petkun W, Singer DE, Simon T, Liu X. Estimates of current and
4 future incidence and prevalence of atrial fibrillation in the U.S. adult population. *The*
5 *American journal of cardiology*. 2013;112:1142-1147. doi:
6 10.1016/j.amjcard.2013.05.063
- 7 2. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, Castella M, Diener
8 HC, Heidbuchel H, Hendriks J, et al. 2016 ESC Guidelines for the management of atrial
9 fibrillation developed in collaboration with EACTS. *European heart journal*.
10 2016;37:2893-2962. doi: 10.1093/eurheartj/ehw210
- 11 3. Stewart S, Hart CL, Hole DJ, McMurray JJ. A population-based study of the long-term
12 risks associated with atrial fibrillation: 20-year follow-up of the Renfrew/Paisley study.
13 *The American journal of medicine*. 2002;113:359-364. doi: 10.1016/s0002-
14 9343(02)01236-6
- 15 4. Soliman EZ, Safford MM, Muntner P, Khodneva Y, Dawood FZ, Zakai NA, Thacker EL,
16 Judd S, Howard VJ, Howard G, et al. Atrial fibrillation and the risk of myocardial
17 infarction. *JAMA internal medicine*. 2014;174:107-114. doi:
18 10.1001/jamainternmed.2013.11912
- 19 5. Miyasaka Y, Barnes ME, Gersh BJ, Cha SS, Bailey KR, Seward JB, Iwasaka T, Tsang
20 TS. Coronary ischemic events after first atrial fibrillation: risk and survival. *The*
21 *American journal of medicine*. 2007;120:357-363. doi: 10.1016/j.amjmed.2006.06.042
- 22 6. Kwok CS, Loke YK, Hale R, Potter JF, Myint PK. Atrial fibrillation and incidence of
23 dementia: a systematic review and meta-analysis. *Neurology*. 2011;76:914-922. doi:
24 10.1212/WNL.0b013e31820f2e38
- 25 7. Calkins H, Kuck KH, Cappato R, Brugada J, Camm AJ, Chen SA, Crijns HJ, Damiano
26 RJ, Jr., Davies DW, DiMarco J, et al. 2012 HRS/EHRA/ECAS expert consensus
27 statement on catheter and surgical ablation of atrial fibrillation: recommendations for
28 patient selection, procedural techniques, patient management and follow-up, definitions,
29 endpoints, and research trial design: a report of the Heart Rhythm Society (HRS) Task
30 Force on Catheter and Surgical Ablation of Atrial Fibrillation. Developed in partnership
31 with the European Heart Rhythm Association (EHRA), a registered branch of the
32 European Society of Cardiology (ESC) and the European Cardiac Arrhythmia Society
33 (ECAS); and in collaboration with the American College of Cardiology (ACC),
34 American Heart Association (AHA), the Asia Pacific Heart Rhythm Society (APHRS),
35 and the Society of Thoracic Surgeons (STS). Endorsed by the governing bodies of the
36 American College of Cardiology Foundation, the American Heart Association, the
37 European Cardiac Arrhythmia Society, the European Heart Rhythm Association, the
38 Society of Thoracic Surgeons, the Asia Pacific Heart Rhythm Society, and the Heart
39 Rhythm Society. *Heart rhythm*. 2012;9:632-696.e621. doi: 10.1016/j.hrthm.2011.12.016
- 40 8. Krahn AD, Manfreda J, Tate RB, Mathewson FA, Cuddy TE. The natural history of atrial
41 fibrillation: incidence, risk factors, and prognosis in the Manitoba Follow-Up Study. *The*
42 *American journal of medicine*. 1995;98:476-484. doi: 10.1016/s0002-9343(99)80348-9
- 43 9. Wang TJ, Parise H, Levy D, D'Agostino RB, Sr., Wolf PA, Vasan RS, Benjamin EJ.
44 Obesity and the risk of new-onset atrial fibrillation. *Jama*. 2004;292:2471-2477. doi:
45 10.1001/jama.292.20.2471

- 1 10. Kannel WB, Abbott RD, Savage DD, McNamara PM. Epidemiologic features of chronic
2 atrial fibrillation: the Framingham study. *The New England journal of medicine*.
3 1982;306:1018-1022. doi: 10.1056/nejm198204293061703
- 4 11. Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA. Independent
5 risk factors for atrial fibrillation in a population-based cohort. The Framingham Heart
6 Study. *Jama*. 1994;271:840-844.
- 7 12. Dewland TA, Olgin JE, Vittinghoff E, Marcus GM. Incident atrial fibrillation among
8 Asians, Hispanics, blacks, and whites. *Circulation*. 2013;128:2470-2477. doi:
9 10.1161/circulationaha.113.002449
- 10 13. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, Singer DE. Prevalence
11 of diagnosed atrial fibrillation in adults: national implications for rhythm management
12 and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation
13 (ATRIA) Study. *Jama*. 2001;285:2370-2375. doi: 10.1001/jama.285.18.2370
- 14 14. Alonso A, Lopez FL, Matsushita K, Loehr LR, Agarwal SK, Chen LY, Soliman EZ,
15 Astor BC, Coresh J. Chronic kidney disease is associated with the incidence of atrial
16 fibrillation: the Atherosclerosis Risk in Communities (ARIC) study. *Circulation*.
17 2011;123:2946-2953. doi: 10.1161/circulationaha.111.020982
- 18 15. Andersen K, Rasmussen F, Neovius M, Tynelius P, Sundstrom J. Body size and risk of
19 atrial fibrillation: a cohort study of 1.1 million young men. *Journal of internal medicine*.
20 2018;283:346-355. doi: 10.1111/joim.12717
- 21 16. Rosenberg MA, Kaplan RC, Siscovick DS, Psaty BM, Heckbert SR, Newton-Cheh C,
22 Mukamal KJ. Genetic variants related to height and risk of atrial fibrillation: the
23 cardiovascular health study. *American journal of epidemiology*. 2014;180:215-222. doi:
24 10.1093/aje/kwu126
- 25 17. Lubitz SA, Yin X, Rienstra M, Schnabel RB, Walkey AJ, Magnani JW, Rahman F,
26 McManus DD, Tadros TM, Levy D, et al. Long-term outcomes of secondary atrial
27 fibrillation in the community: the Framingham Heart Study. *Circulation*. 2015;131:1648-
28 1655. doi: 10.1161/circulationaha.114.014058
- 29 18. Moss TJ, Calland JF, Enfield KB, Gomez-Manjarres DC, Ruminski C, DiMarco JP, Lake
30 DE, Moorman JR. New-Onset Atrial Fibrillation in the Critically Ill. *Critical care*
31 *medicine*. 2017;45:790-797. doi: 10.1097/ccm.0000000000002325
- 32 19. Shoemaker MB, Bollmann A, Lubitz SA, Ueberham L, Saini H, Montgomery J, Edwards
33 T, Yoneda Z, Sinner MF, Arya A, et al. Common genetic variants and response to atrial
34 fibrillation ablation. *Circ Arrhythm Electrophysiol*. 2015;8:296-302. doi:
35 10.1161/circep.114.001909
- 36 20. Andrade JG, Deyell MW, Verma A, Macle L, Champagne J, Leong-Sit P, Novak P,
37 Badra-Verdu M, Sapp J, Khairy P, et al. Association of Atrial Fibrillation Episode
38 Duration With Arrhythmia Recurrence Following Ablation: A Secondary Analysis of a
39 Randomized Clinical Trial. *JAMA Network Open*. 2020;3:e208748-e208748. doi:
40 10.1001/jamanetworkopen.2020.8748
- 41 21. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander
42 ES, Lubitz SA, Ellinor PT, et al. Genome-wide polygenic scores for common diseases
43 identify individuals with risk equivalent to monogenic mutations. *Nat Genet*.
44 2018;50:1219-1224. doi: 10.1038/s41588-018-0183-z

- 1 22. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM, Almgren P,
2 Alonso A, Anderson CD, Aragam KG, et al. Multi-ethnic genome-wide association study
3 for atrial fibrillation. *Nat Genet.* 2018;50:1225-1233. doi: 10.1038/s41588-018-0133-9
- 4 23. Choi SH, Weng L-C, Roselli C, Lin H, Haggerty CM, Shoemaker MB, Barnard J, Arking
5 DE, Chasman DI, Albert CM, et al. Association Between Titin Loss-of-Function Variants
6 and Early-Onset Atrial Fibrillation. *JAMA.* 2018;320:2354-2364. doi:
7 10.1001/jama.2018.18179
- 8 24. Yoneda ZT, Anderson KC, Ye F, Quintana JA, O'Neill MJ, Sims RA, Sun L, Glazer AM,
9 Davogusto G, El-Harasis M, et al. Mortality Among Patients With Early-Onset Atrial
10 Fibrillation and Rare Variants in Cardiomyopathy and Arrhythmia Genes. *JAMA*
11 *cardiology.* 2022;7:733-741. doi: 10.1001/jamacardio.2022.0810
- 12 25. Guo Y, Lip GY, Apostolakis S. Inflammation in atrial fibrillation. *Journal of the*
13 *American College of Cardiology.* 2012;60:2263-2270. doi: 10.1016/j.jacc.2012.04.063
- 14 26. Harada M, Van Wagoner DR, Nattel S. Role of inflammation in atrial fibrillation
15 pathophysiology and management. *Circ J.* 2015;79:495-502. doi: 10.1253/circj.CJ-15-
16 0138
- 17 27. Jacob KA, Nathoe HM, Dieleman JM, van Osch D, Kluin J, van Dijk D. Inflammation in
18 new-onset atrial fibrillation after cardiac surgery: a systematic review. *European journal*
19 *of clinical investigation.* 2014;44:402-428. doi: 10.1111/eci.12237
- 20 28. Gungor B, Ekmekci A, Arman A, Ozcan KS, Ucer E, Alper AT, Calik N, Yilmaz H,
21 Tezel T, Coker A, et al. Assessment of interleukin-1 gene cluster polymorphisms in lone
22 atrial fibrillation: new insight into the role of inflammation in atrial fibrillation. *Pacing*
23 *and clinical electrophysiology : PACE.* 2013;36:1220-1227. doi: 10.1111/pace.12182
- 24 29. Marcus GM, Whooley MA, Glidden DV, Pawlikowska L, Zaroff JG, Olgin JE.
25 Interleukin-6 and atrial fibrillation in patients with coronary artery disease: data from the
26 Heart and Soul Study. *American heart journal.* 2008;155:303-309. doi:
27 10.1016/j.ahj.2007.09.006
- 28 30. Kato K, Oguri M, Hibino T, Yajima K, Matsuo H, Segawa T, Watanabe S, Yoshida H,
29 Satoh K, Nozawa Y, et al. Genetic factors for lone atrial fibrillation. *International journal*
30 *of molecular medicine.* 2007;19:933-939.
- 31 31. Wang YH, Fu L, Wang B, Li SF, Sun Z, Luan Y. Genetic variants of interleukin-18 are
32 associated with reduced risk of atrial fibrillation in a population from Northeast China.
33 *Gene.* 2017;626:269-274. doi: 10.1016/j.gene.2017.05.034
- 34 32. Frustaci A, Chimenti C, Bellocci F, Morgante E, Russo MA, Maseri A. Histological
35 substrate of atrial biopsies in patients with lone atrial fibrillation. *Circulation.*
36 1997;96:1180-1184. doi: 10.1161/01.cir.96.4.1180
- 37 33. Gutierrez A, Van Wagoner DR. Oxidant and Inflammatory Mechanisms and Targeted
38 Therapy in Atrial Fibrillation: An Update. *J Cardiovasc Pharmacol.* 2015;66:523-529.
39 doi: 10.1097/fjc.0000000000000313
- 40 34. Marcus GM, Smith LM, Ordovas K, Scheinman MM, Kim AM, Badhwar N, Lee RJ,
41 Tseng ZH, Lee BK, Olgin JE. Intracardiac and extracardiac markers of inflammation
42 during atrial fibrillation. *Heart Rhythm.* 2010;7:149-154. doi:
43 10.1016/j.hrthm.2009.10.004
- 44 35. Marcus GM, Smith LM, Glidden DV, Wilson E, McCabe JM, Whiteman D, Tseng ZH,
45 Badhwar N, Lee BK, Lee RJ, et al. Markers of inflammation before and after curative

- 1 ablation of atrial flutter. *Heart Rhythm*. 2008;5:215-221. doi:
2 10.1016/j.hrthm.2007.10.007
- 3 36. Guo Q, Lu X, Gao Y, Zhang J, Yan B, Su D, Song A, Zhao X, Wang G. Cluster analysis:
4 a new approach for identification of underlying risk factors for coronary artery disease in
5 essential hypertensive patients. *Scientific reports*. 2017;7:43965. doi: 10.1038/srep43965
- 6 37. Ahmad T, Pencina MJ, Schulte PJ, O'Brien E, Whellan DJ, Pina IL, Kitzman DW, Lee
7 KL, O'Connor CM, Felker GM. Clinical implications of chronic heart failure phenotypes
8 defined by cluster analysis. *Journal of the American College of Cardiology*.
9 2014;64:1765-1774. doi: 10.1016/j.jacc.2014.07.979
- 10 38. Parikh KS, Rao Y, Ahmad T, Shen K, Felker GM, Rajagopal S. Novel approach to
11 classifying patients with pulmonary arterial hypertension using cluster analysis.
12 *Pulmonary circulation*. 2017;7:486-493. doi: 10.1177/2045893217705891
- 13 39. Inohara T, Shrader P, Pieper K, Blanco RG, Thomas L, Singer DE, Freeman JV, Allen
14 LA, Fonarow GC, Gersh B, et al. Association of of Atrial Fibrillation Clinical
15 Phenotypes With Treatment Patterns and Outcomes: A Multicenter Registry Study.
16 *JAMA cardiology*. 2018;3:54-63. doi: 10.1001/jamacardio.2017.4665
- 17 40. Inohara T, Piccini JP, Mahaffey KW, Kimura T, Katsumata Y, Tanimoto K, Inagawa K,
18 Ikemura N, Ueda I, Fukuda K, et al. A Cluster Analysis of the Japanese Multicenter
19 Outpatient Registry of Patients With Atrial Fibrillation. *Am J Cardiol*. 2019;124:871-878.
20 doi: 10.1016/j.amjcard.2019.05.071
- 21 41. Fuster V, Ryden LE, Asinger RW, Cannom DS, Crijns HJ, Frye RL, Halperin JL, Kay
22 GN, Klein WW, Levy S, et al. ACC/AHA/ESC Guidelines for the Management of
23 Patients With Atrial Fibrillation: Executive Summary A Report of the American College
24 of Cardiology/American Heart Association Task Force on Practice Guidelines and the
25 European Society of Cardiology Committee for Practice Guidelines and Policy
26 Conferences (Committee to Develop Guidelines for the Management of Patients With
27 Atrial Fibrillation) Developed in Collaboration With the North American Society of
28 Pacing and Electrophysiology. *Circulation*. 2001;104:2118-2150.
- 29 42. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification
30 for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-
31 based approach: the euro heart survey on atrial fibrillation. *Chest*. 2010;137:263-272. doi:
32 10.1378/chest.09-1584
- 33 43. Roden DM, Denny JC. Integrating electronic health record genotype and phenotype
34 datasets to transform patient care. *Clin Pharmacol Ther*. 2016;99:298-305. doi:
35 10.1002/cpt.321
- 36 44. Wei WQ, Denny JC. Extracting research-quality phenotypes from electronic health
37 records to support precision medicine. *Genome Med*. 2015;7:41. doi: 10.1186/s13073-
38 015-0166-y
- 39 45. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to
40 Advance Precision Medicine. *Annu Rev Genomics Hum Genet*. 2016;17:353-373. doi:
41 10.1146/annurev-genom-090314-024956
- 42 46. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsler JR, Masys DR.
43 Development of a large-scale de-identified DNA biobank to enable personalized
44 medicine. *Clin Pharmacol Ther*. 2008;84:362-369. doi: 10.1038/clpt.2008.89

- 1 47. Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects
2 protections applied in an opt-out, de-identified biobank. *Clinical and translational*
3 *science*. 2010;3:42-48. doi: 10.1111/j.1752-8062.2010.00175.x
- 4 48. Weng LC, Preis SR, Hulme OL, Larson MG, Choi SH, Wang B, Trinquart L, McManus
5 DD, Staerk L, Lin H, et al. Genetic Predisposition, Clinical Risk Factor Burden, and
6 Lifetime Risk of Atrial Fibrillation. *Circulation*. 2018;137:1027-1038. doi:
7 10.1161/circulationaha.117.031431
- 8 49. Mosley JD, Benson MD, Smith JG, Melander O, Ngo D, Shaffer CM, Ferguson JF,
9 Herzig MS, McCarty CA, Chute CG, et al. Probing the Virtual Proteome to Identify
10 Novel Disease Biomarkers. *Circulation*. 2018;138:2469-2481. doi:
11 10.1161/circulationaha.118.036063
- 12 50. Imazio M, Brucato A, Ferrazzi P, Pullara A, Adler Y, Barosi A, Caforio AL, Cemin R,
13 Chirillo F, Comoglio C, et al. Colchicine for prevention of postpericardiotomy syndrome
14 and postoperative atrial fibrillation: the COPPS-2 randomized clinical trial. *Jama*.
15 2014;312:1016-1023. doi: 10.1001/jama.2014.11026
- 16 51. Imazio M, Brucato A, Ferrazzi P, Rovere ME, Gandino A, Cemin R, Ferrua S, Belli R,
17 Maestroni S, Simon C, et al. Colchicine reduces postoperative atrial fibrillation: results of
18 the Colchicine for the Prevention of the Postpericardiotomy Syndrome (COPPS) atrial
19 fibrillation substudy. *Circulation*. 2011;124:2290-2295. doi:
20 10.1161/circulationaha.111.026153
- 21 52. Nigam A, Talajic M, Roy D, Nattel S, Lambert J, Nozza A, Jones P, Ramprasath VR,
22 O'Hara G, Kopecky S, et al. Fish oil for the reduction of atrial fibrillation recurrence,
23 inflammation, and oxidative stress. *Journal of the American College of Cardiology*.
24 2014;64:1441-1448. doi: 10.1016/j.jacc.2014.07.956
- 25 53. Darghosian L, Free M, Li J, Gebretsadik T, Bian A, Shintani A, McBride BF, Solus J,
26 Milne G, Crossley GH, et al. Effect of omega-three polyunsaturated fatty acids on
27 inflammation, oxidative stress, and recurrence of atrial fibrillation. *Am J Cardiol*.
28 2015;115:196-201. doi: 10.1016/j.amjcard.2014.10.022
- 29 54. Roselli C, Rienstra M, Ellinor PT. Genetics of Atrial Fibrillation in 2020. *Circulation*
30 *Research*. 2020;127:21-33. doi: doi:10.1161/CIRCRESAHA.120.316575

31

32

33

34

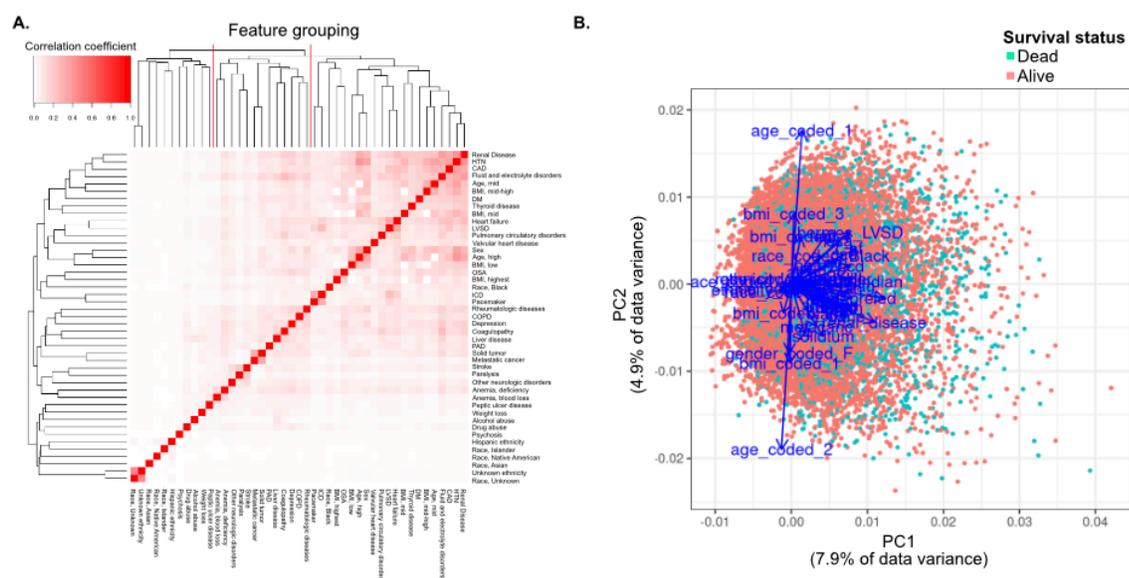
35

36

37

38

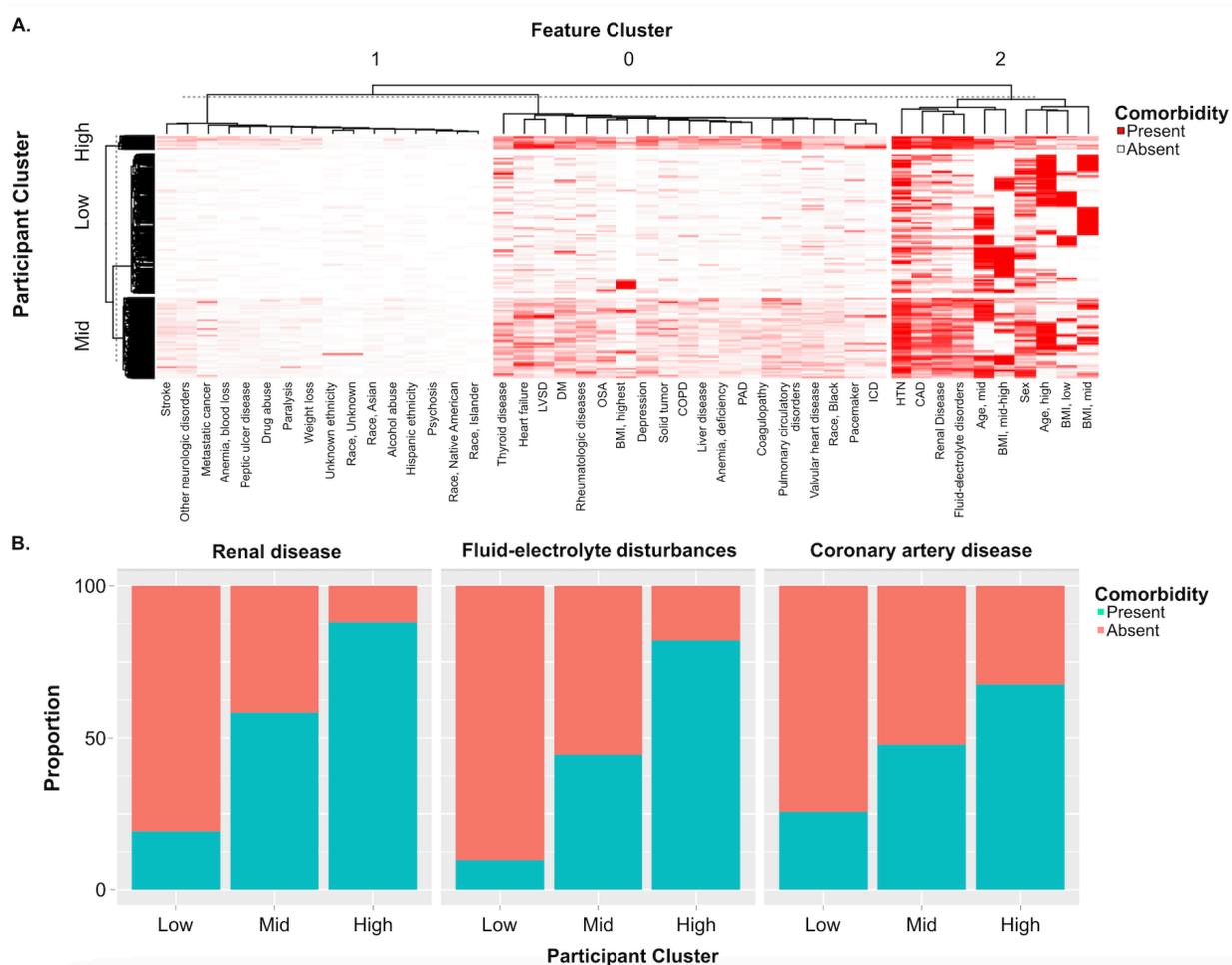
1
2 **Fig. 1: Overall exploration, and correlation of clinical variables used as predictors in**
3 **clustering algorithm**



4
5 A) Observed correlation between clinical predictors used for clustering. There were no highly
6 correlated features ($r < 0.5$). Correlation clustering of features suggest approximately 3 groups
7 (separated by vertical red lines in top horizontal axis) B) Principal component analysis of
8 predictors reveal PC1 is mainly determined by comorbidity burden, while PC2 is mostly
9 determined by age at time of AF diagnosis and BMI. There is an association between positive
10 PC1 and death during follow up ($p < 0.001$). Abbreviations: AF: Atrial fibrillation, CAD:
11 coronary artery disease, BMI: Body mass index, COPD: Chronic obstructive pulmonary disease,
12 PAD: Peripheral arterial disease, LVSD: Left ventricular systolic dysfunction, ICD: Implantable
13 cardioverter defibrillator, OSA: Obstructive sleep apnea, DM: Diabetes mellitus, HTN:
14 Hypertension, PC: Principal component.

15

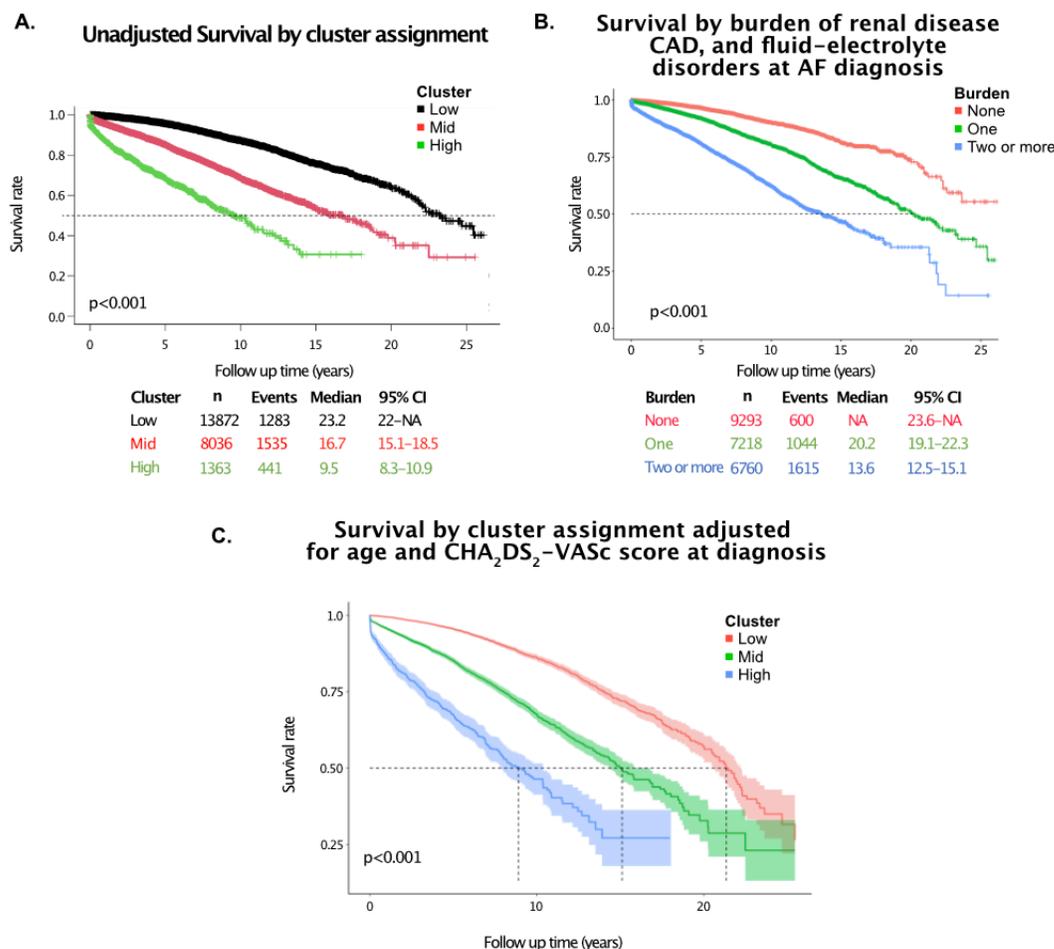
1 **Fig. 2: Clustering assignment and feature prevalence at AF diagnosis time by cluster**



2
 3
 4 A) Heatmap of cluster assignment by participant and feature. Each row represents a participant,
 5 and each column a feature. White denotes absence of feature at AF first mention time while red
 6 presence of feature. Renal dysfunction, fluid-electrolyte disorders, and CAD are the most
 7 striking features which determine participant cluster assignment. B) Comorbidity burden for
 8 renal dysfunction, fluid-electrolyte disorders, and CAD distribution across clusters.
 9 Abbreviations: AF: Atrial fibrillation, VHD: Valvular heart disease, CAD: coronary artery
 10 disease, BMI: Body mass index, COPD: Chronic obstructive pulmonary disease, PAD:
 11 Peripheral arterial disease, LVSD: Left ventricular systolic dysfunction, ICD: Implantable

- 1 cardioverter defibrillator, OSA: Obstructive sleep apnea, DM: Diabetes mellitus, HTN:
- 2 Hypertension, Rheum: Rheumatologic diseases
- 3

1 **Fig. 3: Survival by cluster assignment and by burden of renal dysfunction, coronary artery**
 2 **disease, and fluid-electrolyte disorders**



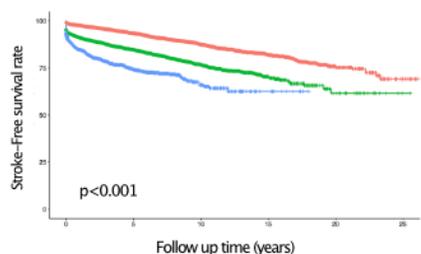
16

17 A) Kaplan-Meier curves demonstrating observed survival by cluster assignment. B) Kaplan-
 18 Meier curves demonstrating observed survival by burden of renal dysfunction, coronary artery
 19 disease (CAD), and fluid-electrolyte disorders. C) Predicted survival by cluster assignment
 20 adjusted for age at CHA₂DS₂-VASc score at time of first AF mention. The solid line represents
 21 point-estimate and the shaded area the 95% point-wise confidence interval. Abbreviations: AF:
 22 Atrial fibrillation, CAD: coronary artery disease.

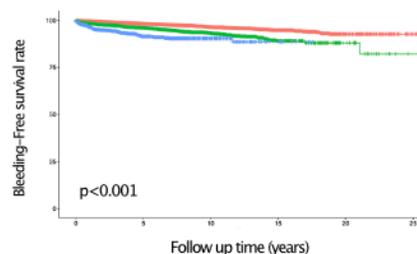
23

1 **Fig. 4: Unadjusted clinical outcomes during follow up by cluster assignment**

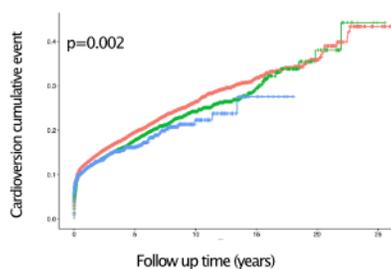
A. Unadjusted Stroke-free survival by cluster assignment



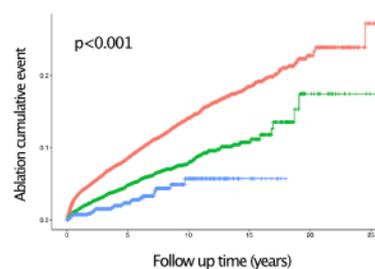
B. Unadjusted Major Bleeding-free survival by cluster assignment



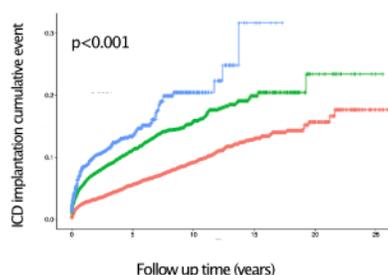
C. Cumulative Cardioversion event rate by cluster assignment



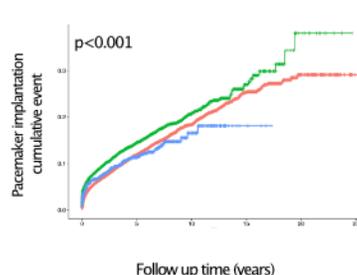
D. Cumulative Ablation event rate by cluster assignment



E. Cumulative ICD implant event rate by cluster assignment



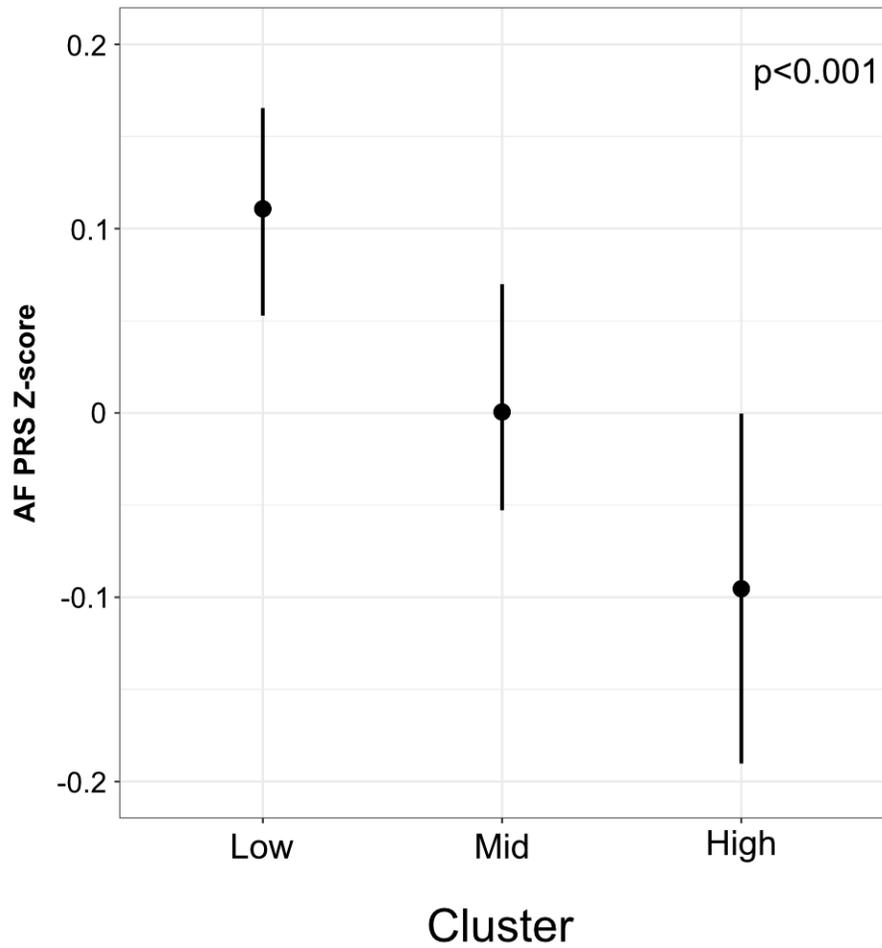
F. Cumulative Pacemaker implant event rate by cluster assignment



- 2
- 3 A) Kaplan-Meier curves demonstrating observed stroke-free survival by cluster assignment. B)
- 4 Kaplan-Meier curves demonstrating observed major bleeding-free survival by cluster
- 5 assignment. C) Observed cumulative cardioversion event rate by cluster assignment. D)
- 6 Observed cumulative atrial fibrillation ablation event rate by cluster assignment. E) Observed
- 7 cumulative ICD implantation event rate by cluster assignment. C) Observed cumulative

- 1 pacemaker implantation event rate by cluster assignment. Abbreviations: AF: Atrial fibrillation,
- 2 ICD: Implantable cardioverter-defibrillator.

1 **Fig. 5: Normalized atrial fibrillation polygenic risk score by cluster adjusted for age, sex,**
2 **and first 4 principal components of ancestry**



18 Normalized (Z-score) of atrial fibrillation polygenic risk score (AF PRS) by cluster group
19 adjusted for first 4 principal components of ancestry, age at first AF mention, and sex. Data are
20 presented as median predicted values after adjustment (dot) and pointwise 95% confidence
21 interval (bars). Abbreviations: AF: Atrial fibrillation, PRS: Polygenic risk score.

22

23

1 **Table 1: Baseline characteristics of the included participants at time of first AF mention in**

2 **EHR**

3
4

Characteristic	All participants (n=23,271)
Age, years	68 (59-76)
Female Sex	9,908 (42.6)
Self-reported Race	
White	21,032 (90.4)
Black	1,782 (7.7)
Asian	177 (0.8)
Native American	36 (0.2)
Islander	31 (0.1)
Unknown	213 (0.9)
Ethnicity	
Non-Hispanic	22,873 (98.3)
Hispanic	156 (0.7)
Unknown	242 (1.0)
BMI, Kg/m ²	29 (25.2-33.8)
CHA ₂ DS ₂ -VASc score	3 (2-4)
Diabetes mellitus	3,621 (15.6)
Hypertension	14,500 (62.3)
Thyroid medication use	4,264 (18.3)
Obstructive sleep apnea	2,562 (11.0)
Rheumatologic diseases	2,971 (12.8)
Chronic Obstructive Pulmonary Disease	1,811 (7.8)
Prior stroke	1,093 (4.7)
Congestive heart failure	4,281 (18.4)
Left ventricular systolic dysfunction	2,805 (12.1)
Coronary artery disease	8,332 (35.8)
Valvular heart disease	2,200 (9.5)
Prior pacemaker	1,748 (7.5)
Prior ICD	1,411 (6.1)
Renal dysfunction	8,559 (36.8)
Peripheral arterial disease	1,272 (5.5)
Pulmonary circulatory disorders	2,312 (9.9)
Paralysis	418 (1.8)
Other neurological disorders	1,024 (4.4)
Liver disease	1,519 (6.5)

Solid tumor	2,052 (8.8)
Coagulopathy	2,433 (10.5)
Weight loss	351 (1.5)
Fluid electrolyte disorders	6,024 (25.9)
Blood loss anemia	494 (2.1)
Deficiency anemia	1,392 (6.0)
Alcohol abuse	169 (0.7)
Drug abuse	453 (1.9)
Psychoses	76 (0.3)
Depression	2,585 (11.1)

1
2 Age, BMI, and CHA₂DS₂-VASc score are presented as median (IQR). All other variables are
3 categorical or dichotomous presented as n (%). Abbreviations: ICD: Implantable Cardioverter
4 Defibrillator, EHR: Electronic Health Record.

5
6
7

1 **Table 2: Clinically relevant incident outcomes of the included participants after first AF**
 2 **mention in EHR**

3
4

Outcome	All participants (n=23,271)
Follow up, years	5.4 (2.8-8.6)
Dead, n (%)	3,259 (14.0)
Time to death, years	4.5 (1.6-7.8)
Stroke after AF diagnosis, n (%)	2,861 (12.3)
Time to first stroke, years	1.6 (0.1-5.1)
Bleeding related hospitalization, n (%)	784 (3.4)
Bleeding source, n (%)	
Intracranial	294 (1.3)
Gastrointestinal	276 (1.2)
Genitourinary	69 (0.3)
Other	145 (0.6)
Time to initial bleeding related hospitalization, years	2.6 (0.6-5.4)
Traumatic bleeding hospitalization, n (%)	333 (1.4)
Traumatic bleeding source, n (%)	
Gastrointestinal	295 (1.3)
Intracranial	20 (0.1)
Genitourinary	2 (0)
Other	16 (0.1)
Time to initial traumatic bleeding hospitalization, years	2.3 (0.6-5.7)
AF ablation, n (%)	1,884 (8.1)
Time to initial AF ablation, years	2.4 (0.6-5.5)
Cardioversion, n (%)	4630 (19.9)
Time to initial cardioversion, years	0.3 (0.0-3.0)
Antiarrhythmic drug therapy, n (%)	12,847 (55.2)
Time to initial antiarrhythmic, years	0.04 (0.0-1.1)
Pacemaker after AF diagnosis, n (%)	3,184 (13.7)
Time to pacemaker implant, years	1.48 (0.2-4.2)
ICD after AF diagnosis, n (%)	1,972 (8.5)
Time to ICD implant, years	1.3 (0.2-4.4)

5
6
7
8

Abbreviations: ICD: Implantable Cardioverter Defibrillator, EHR: Electronic Health Record.

1 **Table 3: Inflammatory biomarkers levels by cluster**

2
3

Biomarker	Overall	Cluster			p-value
		Low	Mid	High	
Erythrocyte sedimentation rate	28 (12-53) n=2,206	19 (8.3-42) n=855	32 (14.4-56) n=1,052	42 (21-69.5) n=299	<0.001
C-reactive protein (mg/L)	21 (4.7-85.6) n=2,174	11.5 (2.8-57.4) n=639	26.2 (5.5-93.72) n=1,164	34.2 (9.5-102) n=371	<0.001
Platelet count	212 (168-261.5) n=15,771	213 (173-261) n=7,851	212 (165-262.3) n=6,619	206 (153-261) n=1,301	<0.001
White blood cell count	7.9 (6.3-9.8) n=16,438	7.6 (6.1-9.5) n=8,339	8.2 (6.5-10.2) n=6,790	8.2 (6.5-10.2) n=1,309	<0.001
Neutrophil percentage	66.5 (58.3-74.5) n=11,533	64.7 (56.8-72.7) n=5,146	67.7 (59.5-74.5) n=5,210	69.4 (61.5-76.7) n=1,177	<0.001
Red cell distribution width	14.2 (13.3-15.4) n=16,335	13.8 (13.1-14.7) n=8,268	14.6 (13.6-15.9) n=6,759	15.5 (14.3-17) n=1,308	<0.001

4
5

Data are presented in median (IQR), and all comparisons made with one-way ANOVA test.