

Genome-wide association studies reveal differences in genetic susceptibility between single events versus recurrent events of atrial fibrillation and myocardial infarction: the HUNT study

Martina Hall^{1,2}, Anne Heidi Skogholt², Ida Surakka³, Haavard Dalen^{4,5,6} and Eivind Almaas^{1,2,*}

¹Department of Biotechnology and Food Science, Norwegian University of Science and Technology, Trondheim, Norway

²K. G. Jebsen center for Genetic Epidemiology, Norwegian University of Science and Technology, Trondheim, Norway

³Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA

⁴Clinic of Cardiology, St. Olavs University Hospital, Trondheim, Norway

⁵Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway

⁶Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway

Correspondence*:
Corresponding Author
eivind.almaas@ntnu.no

2 ABSTRACT

3 Genetic research into Atrial Fibrillation (AF) and Myocardial Infarction (MI) has predominantly
4 concentrated on contrasting afflicted individuals with their healthy counterparts. However, this
5 approach lacks granularity, overlooking the subtleties within patient populations. In this study,
6 we explore the distinction between AF and MI patients experiencing only single events to
7 those experiencing recurrent events. Integrating hospital records, questionnaire data, clinical
8 measurements, and genetic data from more than 500,000 HUNT and UKBB participants in our
9 analysis, we compare the two groups for both clinical and genetic characteristics with GWAS
10 meta-analyses, PheWAS-analyses, and gene co-expression networks. We find that the two
11 groups of patients differs in both clinical characteristics and genetic risk. More specifically,
12 recurrent AF patients are significantly younger with a better baseline health, in terms of lower
13 cholesterol and blood pressure, than the single AF patients. The GWAS meta-analysis results
14 indicate that recurrent AF patients seem to have a greater genetic risk of recurrent events and
15 the PheWAS-analysis and gene co-expression network analyses highlight differences in the
16 diseases and genetic functions associated to the set of SNPs and genes for to each group. In
17 contrast, we find that for MI patients experiencing single events are significantly younger and
18 have a better baseline health, yet they exhibit higher genetic risk. The GWAS meta-analysis

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

19 identifies mostly genetic regions uniquely associated with single MI, and the PheWAS-analysis
20 and gene co-expression networks support the genetic differences between these two groups.

21 **Keywords:** GWAS meta-analysis, recurrent events, atrial fibrillation, myocardial infarction, networks, PheWAS, HUNT

1 INTRODUCTION

22 Myocardial infarction (MI) and atrial fibrillation (AF) are two major prevalent cardiovascular diseases. AF
23 is an established risk factor for several other cardiovascular diseases, where the severity and mortality risk
24 increase rapidly with relapse of AF. To treat and prevent new events of AF, medical and interventional
25 therapy are available as treatments in order to normalize rhythm or to stabilize the heart rate. AF ablation
26 has become the leading clinical treatment, but with varying success rate, where 20% to 40% of the patients
27 require new treatment (1). Similarly, MI is a severe heart diagnosis associated with high mortality rates.
28 Upon survival, the heart is most likely weakened, making the patient vulnerable for other diseases. In fact,
29 33% of the patients experiencing MI die within a year (all death causes) (2). On the other hand, some
30 patients experience only one event of MI and have a normal and healthy life afterwards. Most MI patients
31 undergo cardiac catheterization and percutaneous coronary interventions in the acute phase, while medical
32 therapy targeting clotting of blood and lipids as well as life-style interventions, are provided to reduce
33 recurrent events. Still, a significant proportion of individuals suffers from relapse of MI.

34 Many genetic studies have been conducted to identify genetic variants that likely affect the risk of AF
35 (3, 4) and MI (5, 6). While many variants have been identified and replicated in other studies, these variants
36 were identified when comparing all included cases of AF or MI with healthy controls. Some studies have
37 been conducted with the aim of understanding the genetic of patients experiencing recurrent events of
38 AF (1, 7, 8, 9, 10) and MI (11, 12, 13). These are, however, mostly focused on either (A) response after
39 treatment, or (B) the genetic effects on recurrence from known AF/MI variants or genes. Little effort
40 has been made regarding the comparison of genetics between patients experiencing a single event versus
41 patients experiencing recurrent events.

42 To our knowledge, no studies have been conducted with a full GWAS-analysis comparing recurrent
43 to single events of AF and MI. In the presented study, we investigate if there are statistically significant
44 genetic differences between patients experiencing recurrent events (defined as two or more events) of AF
45 or MI compared to patients only experiencing a single event. Note that, we perform a general comparison
46 without separating cases based on type of possible treatment after the first AF or MI event. By this, we
47 perform a general comparison between single and recurrent events of AF and MI, to identify if there are
48 any genetic differences between patients with single events to patients with recurrent events.

2 METHODS

2.1 Cohorts

2.1.1 The HUNT study

51 The Trøndelag Health study (HUNT) is a health related population based longitudinal study, based on
52 four rounds of data collection: HUNT1 (1984-1986), HUNT2 (1995-1997), HUNT3 (2006-2008) and
53 HUNT4 (2017-2019). With a unique database covering clinical measurements, questionnaire data and
54 biological samples from roughly 230,000 of Trøndelag counties' inhabitants from 1984 onward, it is one
55 of the largest health studied ever performed (14). A great benefit of the HUNT study is the link with

56 other health related registries through the Norwegian unique personal identification number. Such health
57 registries include hospital and general practitioner registries, cancer registries, cause of death registries,
58 and the prescription database.

59 In the current study, genotype data for 69,621 participants from HUNT2 and HUNT3 were used, and
60 these were linked to questionnaire data and clinical measurements from HUNT1, HUNT2, and HUNT3,
61 regional hospital records, Nord-Trøndelag Hospital Trust (HNT), and the Norwegian Cause of Death
62 registry (COD). The HNT register contains all ICD9 and ICD10-codes for registered hospital visits of these
63 participants in the time period of August 1987 to April 2017. The COD registry spans the same time period,
64 with registered ICD9 and ICD10-codes for the primary and secondary causes of death.

65 2.1.2 The United Kingdom Biobank

66 The United Kingdom Biobank (UKBB) is a health related population based study consisting of roughly
67 500,000 middle aged UK inhabitants from across the country. Sampling of the participants took place
68 during 2006-2010, where questionnaires, clinical measurements, and biological samples were collected.
69 Similar to the HUNT study, it is also linked to electronic health records, containing information about
70 the participants hospital, general practitioner, and death records with ICD9 and ICD10 diagnose codes
71 (15). Genotyped data are available for more than 480,000 participants, and in the current study, we use
72 these data together with relevant questionnaire data, clinical measurements, and hospital and death records
73 for the genotyped participants (of European ancestry). The hospital records span the time period from
74 December 1992 to September 2021. The death registry spans the same period, starting from 2006.

75 2.2 Genotyping and imputation

76 Genotyping and imputation for the HUNT and UKBB participants have been described elsewhere (16, 17).
77 In short, genotyping was performed using one of three Illumina HumanCoreExome arrays: 12 v.1.0, 12 v.1.1
78 with custom content (UM HUNT Biobank v1.0) according to standard protocols for the HUNT participants,
79 and standard protocols for Affymetrix Applied Biosystems UK BiLEVE Axiom or Applied Biosystems UK
80 Biobank Axiom array for the UKBB participants. Standard quality control was performed for the HUNT
81 genotyping, as well as a UKBB-specific quality control for the UKBB genotyping. Imputation in HUNT
82 was performed using 2,202 whole-genome sequenced samples from HUNT together with the Haplotype
83 Reference Consortium (HRC) reference panel (18, 19), resulting in 25 million genetic markers. For UKBB,
84 the HRC and UK10K+1000 Genomes reference panel was used, resulting in 90 million variants.

85 2.3 Definitions of traits and outcomes

86 Hospital records of HUNT and UKBB participants were used to determine cases of MI and AF, and
87 the number of events for each participant. An event of MI is defined as having a registered diagnosis of
88 ICD10:I21-I24 or ICD9:410. An event of AF is similarly defined with registered diagnosis of ICD10:I48 or
89 ICD9:427.3.

90 In both records, each diagnosis is registered as a main or bi-diagnosis (denoted as first and second
91 diagnosis in UKBB), and we consider them both when determining the number of events for each
92 participant. Considering only the main diagnosis and filtering out bi-diagnosis can cause errors in addition
93 to significantly reducing the amount of data. A bi-diagnosis can be interpreted in several ways: One could
94 have the case where a person is admitted to the hospital with a main diagnosis, and additional diagnoses
95 are detected and registered at this first visit. These will be registered as bi-diagnoses, even though they
96 are new diagnoses and could have been recorded as main diagnoses if they were the only disease and the

97 reason for the hospital visit that day. A different example could be that the doctor could infer from
98 previous records that other diseases (such as AF and MI) are relevant diagnoses for the main diagnosis and
99 record these as bi-diagnoses, even though these are not new events. Due to the different ways of reporting
100 bi-diagnosis, we apply some selective filtering in order to separate single events from recurring events of
101 MI and AF.

102 A first event is defined as the first visit where this diagnosis is listed in the records, either as a main or a
103 bi-diagnose. A second event is defined if there is more than one month time delay between the initial and
104 the following events while one of the following two criteria are fulfilled: (i) the current event is a main
105 diagnosis, (ii) the current event is a bi-diagnosis and the only diagnosis given that day (indicating that this
106 is indeed a new event reported on this day). Following events are similarly registered as the second event,
107 with a minimum one month time delay to the previously defined event. Except for the case of selecting the
108 main diagnoses events exclusively, this is the only event definition that ensures the selection of new events.

109 As we are interested in contrasting patients with recurring events with relapse-free patients, we need to
110 establish that the patients with only a single recorded event is not identified as such by the fact that they died
111 before they experienced a new event. Identifying in the HUNT data that 80% of secondary events happened
112 within a five year span for AF and seven year span for MI, we filter out single event participants who
113 have died either due to the phenotype or within five or seven years after the AF or MI event, respectively
114 (corresponding values were found in UKBB). Also, we take notice of the censoring dates of *06-04-2017*
115 and *12-11-2021* for HUNT and UKBB respectively, and filter out participants registered with only one
116 event of AF or MI, where their last AF or MI event is less than five and seven years prior to these censoring
117 dates.

118 The four trait groups are from here denoted as single AF/MI: participants that experience only one
119 event of AF/MI, and recurrent AF/MI: participants that experience more than one event of AF/MI, while
120 satisfying the conditions specified above.

121 Baseline and clinical characteristics, as well as information about other relevant diseases identified
122 with the participants, were taken from HUNT and UKBB hospital records, questionnaires, and clinical
123 measurements. Participants were defined to have diabetes and/or hypertension if they have ever been
124 registered with the ICD-codes ICD10:E10-E14 or ICD9:250 for diabetes and/or ICD10:I10-I15 (excluding
125 I11.0), ICD9:401-405 for hypertension. The smoking variable is based on the HUNT question: "Have you
126 ever smoked? (Yes/No)", where the latest possible HUNT participation were used for each patient. The
127 corresponding variable in UKBB reads "Ever smoked (Yes/No)" which was constructed upon sampling.
128 BMI, cholesterol, systolic and diastolic blood pressure were measured in HUNT2 and HUNT3, where the
129 HUNT study closest prior to the disease event were used. Corresponding variables were measured upon
130 sampling in the UKBB. Student's t-test for continuous variables and Fishers' exact test for binary variables
131 were used to test differences of the characteristics (age, diabetes, hypertension, BMI, smoking, cholesterol,
132 systolic, and diastolic blood pressure) between groups of single versus recurrent events. Test statistics with
133 Bonferroni-adjusted p -values ($p < 0.05/8 = 6.25 \cdot 10^{-3}$) were considered as significant findings.

134 **2.4 GWAS meta-analysis**

135 To identify genetic factors associated with single or recurrent events of AF and MI, we conducted three
136 GWAS analyses for each trait in both cohorts separately: (i) patients with single events versus healthy
137 controls, (ii) patients with recurrent events versus healthy controls, and (iii) patients with recurrent events
138 versus patients with single events. As a control, we also conducted a GWAS analysis in each cohort
139 with all cases of each disease against healthy controls. Healthy controls were defined as participants with

140 no registered events of AF and MI, respectively. Variants with minor allele count (MAC) < 3 and an
141 imputation score < 0.3 were excluded from all GWAS results. Participants with non-European recent
142 ancestry were excluded from the analyses in UKBB (all genotyped HUNT participants are of European
143 ancestry). Association analyses were performed with SAIGE, using a generalized linear mixed model
144 adjusting for relatedness and unbalanced case-control ratios (20). Birth year, gender, batch/chip, and the
145 first four principal components were added as covariates in the models. Genomic variants with minor allele
146 frequency (MAF) > 1% in one or both studies were included in the meta-analysis.

147 From the eight GWAS analyses (three for AF and three for MI, and one control for each disease)
148 performed for both the HUNT population and the UKBB population, we performed eight fixed-effect
149 inverse variance weighted (IVW) meta-analyses using METAL (21). In METAL, each variant is assigned a
150 new effect size as a sum of each study's effect size weighted by the corresponding study variance. The
151 p -values in the meta-analysis are calculated based on the Z statistic given by the new effect sizes and
152 standard errors. Variants reaching genome-wide significance (p -values < $5 \cdot 10^{-8}$) from the Z statistic
153 were considered as significant findings. Annotations of the significant SNPs, identification of nearest genes,
154 and search for nearby SNPs associated with relevant traits were performed with the FUMA platform and
155 the GWAS catalog (22, 23). Variants were considered to be in the same genetic region if they were less than
156 500 kb apart, and genetic regions denoted as shared for both the single and recurrent events meta-analysis
157 were either consisting of the same SNPs or SNPs within the same genetic region.

158 2.5 Phenome Wide Association Studies (PheWAS)

159 Phenome Wide Association Studies (PheWAS) were performed on all SNPs from the meta-analyses
160 reaching genome-wide significance. From the comprehensive Pan UKBB resource (24), we collected
161 results from GWAS conducted on 1,326 phenocodes, and we identified the effect of each of our SNPs
162 of interest on each phenocode. All GWAS results from the Pan UKBB are based on UKBB participants,
163 and we selected results for European ancestry exclusively. For each set of SNPs (identified in common or
164 specifically for either single or recurrent AF/MI), phenotypes with a p -value less than $0.05/(1326 \times n_{set})$,
165 where n_{set} is the number of SNPs in the set, were considered as significant associations. For simplicity,
166 only the SNP with the lowest p -value for each phenotype were selected among each set of SNPs.

167 2.6 Gene function and network analyses

168 The sets of nearest genes to the SNPs identified through the GWAS analyses as common or unique to
169 either recurrent or single AF/MI events (in total six sets) were analyzed for tissue specificity (differentially
170 expressed gene sets in each tissue). We employed both FUMA(22) and gene ontology enrichment using the
171 Fishers Exact over-representation test in PANTHER (protein annotation through evolutionary relationship)
172 (25). Here, biological processes with false discovery rate (FDR) adjusted for multiple testing < 0.05 were
173 considered as functionally enriched for the gene set. To further investigate processes connected to these
174 genes, we performed gene co-expression network analysis (26, 27, 28), where the hypothesis is that highly
175 correlated genes have a regulatory relationship or similar response in a condition (29). Using the identified
176 gene sets as target genes in an egocentric gene co-expression network analysis, we generate a network
177 from the shared neighborhoods among the closest neighbour-genes of each target gene (in the gene) set
178 and inspect the gene functions in the network.

179 Creating these egocentric networks involves several steps. First, using gene expression data from
180 GTEx v.8(30) (<https://www.gtexportal.org>) gene co-expression networks for seven tissue sub-types from
181 heart, muscle skeletal, artery, and kidney (GTEx_Analysis_v8_eQTL_expression_matrices.tar: *Heart Atrial*

182 *Appendage, Heart Left Ventricle, Muscle Skeletal, Artery Aorta, Artery Coronary, Artery Tibial* and
183 *Kidney Cortex*) were created. Since co-expression patterns may vary in different tissues (29), a separate
184 network was created for each tissue. Following the WGCNA-approach (31), the link weight (strength of
185 co-expression) between each pair of genes (i and j) were defined by the weighted topological overlap
186 (wTO):

$$wTO_{ij} = \frac{A_{ij} + \sum_{k \neq i,j} A_{ik}A_{kj}}{\min(\sum_u A_{iu}, \sum_u A_{ju}) + 1 - A_{ij}}, \quad (1)$$

187 where $A_{ij} = |cor(i, j)|^6$ is the absolute Pearson correlation of the gene expressions raised to a power 6
188 to emphasize the strongest correlations. The resulting gene co-expression network is then an all-to-all
189 network where pairs of genes with high wTO -link weights represent strong connections between the genes
190 and their topological neighbourhood. Only the 15% strongest links from each tissue were included in the
191 following analysis (still leaving about 30 million links), to avoid inclusion of genes based on weak (and
192 likely spurious) connections.

193 Next, for each of the seven tissues, egocentric networks for each target gene were extracted from the
194 co-expression networks. The egocentric networks were filtered to include only the 25 genes with the
195 strongest wTO -link weights with each target gene. By merging and further reducing the complexity of
196 the networks, the 25 strongest linked genes to each target gene across all tissues were selected in the final
197 network. Here, we weighted the link strengths using $wTO_{ij}^{weighted} = \sqrt{\sum_{w=1}^W wTO_{ij,w}^2}$, where W is the
198 number of tissues in which the linked gene is among the 25 strongest linked gene to the target gene, and
199 $wTO_{ij,w}$, is the corresponding wTO -link weight in tissue w .

200 The final six sets of egocentric networks for target genes identified as common or as unique for the single
201 or recurrent AF/MI events, were analyzed with the *igraph* R-package (32, 33). Shared neighbouring genes
202 were defined as genes linked to two or more of the target genes. The set of shared neighbourhood genes for
203 each network was plotted separately with Cytoscape (v. 3.8.1) (34) and gene ontology enrichment of these
204 gene sets were obtained through the PANTHER over-representation test (25).

3 RESULTS

205 3.1 Characteristics of trait groups

206 Among the genotyped participants with European ancestry included in this study, there are 7, 127 and
207 29, 330 hospital patients registered with AF in HUNT and UKBB, respectively. With the filtering described
208 in the methods section, we identify 1, 425 HUNT and 9, 561 UKBB patients with single AF events, and
209 2, 267 HUNT and 7, 267 UKBB patients with recurrent AF events. Correspondingly, 5, 805 HUNT and
210 14, 592 UKBB participants are registered with MI events, and 1, 651 HUNT and 6, 584 UKBB patients are
211 identified with single MI events, and 1, 615 HUNT and 1, 615 UKBB patients are identified with recurrent
212 MI events.

213 Baseline and clinical characteristics of these patients are presented in Tab. 1. Comparing the two AF
214 groups in HUNT, there seems to be a tendency that the group experiencing single AF events are older with
215 worse health conditions, having a higher percentage of diabetes and hypertension and being significantly
216 older (adjusted p -value $2 \cdot 10^{-10}$) with significantly higher levels of cholesterol and systolic blood pressure
217 (adjusted p -values $6.1 \cdot 10^{-6}$ and $5 \cdot 10^{-4}$). Comparing the AF groups in UKBB, we see the same tendency
218 with significantly higher age (adjusted p -value $< 10^{-16}$) and somewhat higher levels of BMI and systolic
219 blood pressure for the single AF event group compared to the recurrent AF group. Comparing the MI groups

	AF HUNT		AF UKBB	
	Single <i>n</i> = 1,435	Recurrent <i>n</i> = 2,267	Single <i>n</i> = 9,561	Recurrent <i>n</i> = 7,267
Male (%)	58	62	65	66
Birth year	1933(12)	1936(12)	1946(6)	1947(6)
Age first event	71(12)*	68(12)	65(7)*	64(8)
Diabetes (%)	21	18	19	19
Hypertension (%)	59	57	67	70*
BMI	28.2(4.4)	28.3(4.6)	29.0(5.3)	28.9(5.2)
Smoking (%)	62	64	66	66
Cholesterol	5.73(1.33)*	5.72(1.21)	5.6(1.2)	5.6(1.2)
Systolic BP	144(24)*	141(22)	142(19)	141(19)
Diastolic BP	79.4(14.2)	79.0(13.3)	82.6(10.8)	82.5(10.7)

	MI HUNT		MI UKBB	
	Single event <i>n</i> = 1,651	Recurrent events <i>n</i> = 1,615	Single event <i>n</i> = 6,584	Recurrent events <i>n</i> = 1,615
Male (%)	70.7	67.2	78	78
Birth year	1935(12)	1934(13)	1947(6)	1947(6)
Age first event	64(11)	68(12)*	59(8)	61(9)*
Diabetes (%)	19	25*	21	35*
Hypertension (%)	52	58*	74	88*
BMI	27.7(4.0)	28.0(4.1)	28.8(4.5)	29.3(5.0)*
Smoking (%)	76*	71	72	74
Cholesterol	5.67(1.45)	5.92(1.44)*	5.50(1.27)	5.45(1.28)
Systolic BP	140(22)	144(23)*	140(20)	141(20)
Diastolic BP	78.7(12.9)	80.0(13.4)	81.1(11.0)	80.5(11.0)

Table 1. Characteristics of sample groups of single and recurrent events of AF and MI in the HUNT and UKBB population. Continuous variables are presented as mean and (sd), and dichotomous variables are reported in percentages. Starred values indicate a significant difference between the single and recurrent group, where the starred value is marking the larger value.

220 in HUNT, we see the reverse trends, where patients experiencing recurrent MI events are significantly older
 221 (adjusted p -value $< 10^{-16}$) with significantly higher percentages of diabetes and hypertension (adjusted
 222 p -values $8.8 \cdot 10^{-5}$ and $7.6 \cdot 10^{-3}$) and significantly higher levels of cholesterol and systolic blood pressure
 223 (adjusted p -values $5.4 \cdot 10^{-6}$ and $1.4 \cdot 10^{-5}$). There is also a tendency of higher BMI and diastolic blood
 224 pressure for this group. The same trends are found in the UKBB MI groups, with significantly higher
 225 age, BMI and percentage of diabetes and hypertension (adjusted p -values $6 \cdot 10^{-14}$, $3.9 \cdot 10^{-5}$, $< 10^{-16}$
 226 and $< 10^{-16}$) for the recurrent MI events group compared to the single MI events group. There is also a
 227 tendency towards higher levels of systolic blood pressure for the UKBB single MI events group.

228 In total, these are indications of patients experiencing only single AF events being older at their first event
 229 and with worse health condition and lifestyle factors compared to the recurrent AF cases. We therefore
 230 hypothesise that single AF events could be more likely driven by age and other lifestyle factors, while there
 231 are genetic factors driving recurrent AF events. For MI, the characteristics point in the opposite direction,
 232 where patients experiencing recurrent MI events are older and generally of worse health conditions and
 233 lifestyle factors compared to patients with only one MI event and surviving it. For MI, we therefore
 234 formulate two hypothesis: (i) that recurrent MI events are driven by the age at first event and worse health
 235 conditions, while single MI events are driven by genetic factors, or (ii) that both single and recurrent MI

236 are driven by the same genetic factors, but recurring MI are due to higher age and other lifestyle factors
237 affecting the risk of obtaining new MI events.

238 **3.2 Genetic differences**

239 In the following, we explore our hypothesis (defined above) for AF and MI by investigating genetic
240 differences between the groups identified through the GWAS meta-analyses.

241 **3.2.1 Genetic differences in AF**

242 To test our hypothesis that patients experiencing recurrent AF events are more genetically susceptible
243 than patients experiencing single AF, we perform three GWAS-meta analyses (see Methods). The GWAS
244 meta-analysis comparing single to recurrent AF events found no regions with significantly different effects.
245 Some SNPs were identified genome-wide significant in the HUNT population, but these were rare variants
246 ($MAF \leq 0.2\%$) and removed through filtering prior to the META-analysis. Comparing the GWAS meta-
247 analyses of each group against healthy controls (Tab. 2 and Fig. 1), we find that 18 regions are specifically
248 associated with recurrent AF events, 2 are specifically associated with single AF events and 16 are identified
249 in both GWAS investigations. Many regions comprise multiple SNPs that exhibit significant effects in only
250 one of the study groups. Five regions identified in the recurrent AF GWAS study consist of only one SNP,
251 yet these are identified with similarly strong effects in both the HUNT and the UKBB studies, indicating a
252 genuine association. Regional plots of the single SNP hits uniquely associated with recurrent AF are shown
253 in Additional File 1, Fig.S15-S19. The presented results show that over half of the identified regions are
254 specifically linked to single or recurrent AF, supporting the hypothesis that patients who have experienced
255 recurrent AF events are genetically more susceptible than those who have only experienced one event and
256 survived it.

257 All regions had previously been associated with AF, and all regions except one (the region at chromosome
258 7 in the KCNH2 gene) were identified in the full AF GWAS meta-analysis by comparing all AF cases
259 against healthy controls. This indicates that all the regions identified as unique for single or recurrent AF
260 (excluding the KCNH2 gene region) have an effect when compared to healthy controls, with the true effect
261 being mainly or solely for patients experiencing single or recurrent AF. The five SNPs in the KCNH2 gene,
262 however, are not detected in our full GWAS meta-analysis and therefore only show an effect for patients
263 experiencing recurrent AF.

264 To our knowledge, only seven genes have been found to be associated with AF recurrence: SOX5, CAV1,
265 EPHX2, ITGA9, SLC8A1, TBX5 and PITX2 (10, 8, 35, 1). Our findings show that regions proximate
266 to the SOX5, CAV1, TBX5 and PITX2 genes are identified in both the single and the recurrent GWAS,
267 yielding comparable effects. Thus, no differences in the impact of these regions between the two groups
268 are evident. No regions were identified near the EPHX2, SCL8A1 and ITGA9 genes. Variants near the
269 NAV2 and SCN10A genes have previously been tested for their effect in recurrent AF events without any
270 significant findings (36, 35). In this study, we discovered 26 and 16 SNPs located within and nearby the
271 NAV2 and SCN10A genes, respectively, that are exclusively associated with recurrent AF, suggesting that
272 these SNPs have a distinct effect on recurrent AF patients compared to single AF cases.

273 Several of the genes listed in Tab.2 code for functions related to AF. Two of the genes listed as "Common"
274 (KCNN3 and HCN4) and three genes identified uniquely for recurrent AF (SCN10A, KCNH2 and KCNJ5)
275 are related to electrophysiological activity, coding for potassium and sodium channels. Other genes listed as
276 "Common" in Tab. 2 code for functions directly linked to heart activity and AF (TTN, TBX5, SYNE2 and
277 RPL3L) or indirectly linked through comorbidities (ATXN1, CAV1, SH3PXD2A and ZFH3). Two of the

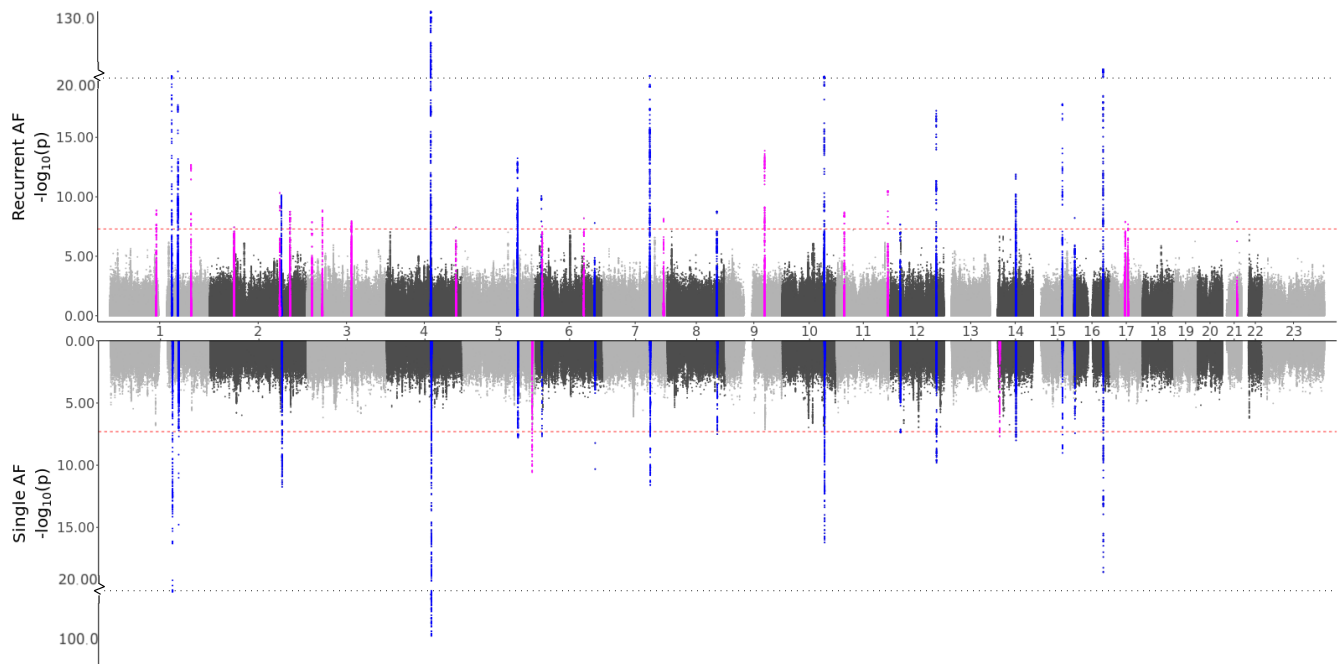


Figure 1. GWAS meta analysis results for Atrial Fibrillation (AF). Top: Comparing recurrent AF patients to AF-free controls. Bottom: Comparing single AF patients to AF-free controls. Blue colored spikes represent regions of SNPs found to be statistically significant in both GWAS studies (common), while magenta spikes represent statistically significant regions of SNPs specifically associated with the given AF group.

278 recurrent AF genes also code for functions directly or indirectly linked to AF (CASQ2 and GOSR2), and
279 some indicate a possible indirect link related to comorbidities, e.g. hypertension or malignancy (PPFIA4,
280 USP34, WIPF1, SPATS2L, CAND2 and AOPEP). The two genes uniquely identified for single AF have
281 been shown important for myocardial diseases and cardiac abnormalities, coding for functions found
282 central in malformation in heart (NKX2-5) and myosin (MYH7).

283 3.2.2 Genetic differences in MI

284 Based on the characteristics of the two MI groups, we formulate two hypotheses: (i) Single MI events
285 may be driven by genetic factors and may be more susceptible to medical treatment, while recurrent MI
286 events may be driven by higher age at the first event and other lifestyle factors, and (ii) Both single and
287 recurrent MI are driven by the same genetic factors, but recurring MI events occur due to higher age and
288 lifestyle factors. Testing for direct genetic differences between the two MI groups, the GWAS meta-analysis
289 (comparing single to recurrent events) did not detect any regions with significant effects. When testing for
290 genetic effects in each groups when compared to MI-free controls, the GWAS meta-analyses shown in
291 Fig. 2 and Tab. 3 identified four regions that are in common for both groups, 24 regions that are specifically
292 identified for the single event groups, and two regions that are unique for the recurrent events group. Hence,
293 some genetic factors are common for both groups, but most identified genetic effects are unique to patients
294 experiencing only one event of MI and surviving it. This results is in support of our first hypothesis.

295 Some distinct regions, including the SNPs in the NBEAL1 and ATXN2 genes for single MI events, and
296 SNPs in the MIA3 gene for recurrent events, exhibit substantial effects for multiple SNPs in the region,
297 with comparable effects in both the HUNT and UKBB populations. Several regions represent suggestive
298 findings comprising only single SNPs and are only identified in the HUNT population (regional plots

Hall et al.

AFull	AF Study	Known	rsID	Chr:pos_Ref/Alt	Gene	Function	Effect	StdErr	Pvalue	Dir	Nsnps
Yes	Common	Yes	rs6691463	1:154814538_C/G	KCNN3	intronic	-0.16	0.015	3.5×10^{-24}	--	166
Yes	Common	Yes	rs72700114	1:170193825_G/C	NTMT2	intergenic	0.36	0.031	9.1×10^{-32}	++	239
Yes	Common	Yes	rs1429094	2:179515774_A/G	TTN	intergenic	-0.14	0.02	1.8×10^{-12}	--	64
Yes	Common	Yes	rs6843082	4:111718067_G/A	PITX2	intergenic	-0.5	0.021	2.4×10^{-132}	--	553
Yes	Common	Yes	rs678897	5:137441065_G/C	NME5	intergenic	0.13	0.017	5.7×10^{-14}	++	229
Yes	Common	Yes	rs112899072	6:16417852_C/G	ATXN1	intronic	0.15	0.023	8.6×10^{-11}	++	10
Yes	Common	Yes	rs117984853	6:149399100_G/T	UST	downstream	0.16	0.025	4.9×10^{-11}	++	2
Yes	Common	Yes	rs11773845	7:116191301_C/A	CAV1	intergenic	0.16	0.016	4×10^{-24}	++	218
Yes	Common	Yes	rs17337621	8:124542519_G/C	FBXO32	intronic	0.19	0.032	1.7×10^{-9}	++	5
Yes	Common	Yes	rs1389189	10:105486077_A/G	SH3PXD2A	intronic	-0.14	0.017	6.1×10^{-17}	--	122
Yes	Common	Yes	rs137913153	12:24776752_A/G	SOX5	intergenic	0.11	0.02	4.2×10^{-8}	++	8
Yes	Common	Yes	rs883079	12:114793240_C/T	TBX5	UTR3	0.15	0.017	5.8×10^{-18}	++	67
Yes	Common	Yes	rs1152591	14:64680848_A/G	SYNE2	intronic	0.11	0.015	1.3×10^{-12}	++	67
Yes	Common	Yes	rs7172038	15:73667255_T/G	HCN4	intergenic	-0.18	0.021	1.6×10^{-18}	--	44
Yes	Common	Yes	rs140185678	16:2003016_G/A	RPL3L	exonic	0.25	0.043	6.1×10^{-9}	++	1
Yes	Common	Yes	rs4499262	16:73059159_C/A	ZFH3	intronic	0.26	0.021	2.5×10^{-35}	++	108
Yes	Recurrent	Yes	rs9428227	1:116309200_C/T	CASQ2	intronic	0.093	0.015	1.3×10^{-9}	++	24
Yes	Recurrent	Yes	rs2270543	1:203030685_T/C	PPFIA4	intronic	0.11	0.015	2×10^{-13}	++	26
Yes	Recurrent	Yes	rs12463885	2:61457996_A/C	USP4	intronic	0.086	0.016	3.6×10^{-8}	++	1
Yes	Recurrent	Yes	rs56181519	2:175555714_C/T	WIPF1	intergenic	-0.12	0.018	4.8×10^{-11}	--	4
Yes	Recurrent	Yes	rs7605146	2:201183888_G/A	SPATS2L	intronic	0.094	0.016	1.8×10^{-9}	++	40
Yes	Recurrent	Yes	rs4642101	3:12842223_T/G	CAND2	intronic	-0.091	0.016	1.3×10^{-8}	--	2
Yes	Recurrent	Yes	rs6790396	3:38771925_C/G	SCN10A	intronic	-0.095	0.016	1.4×10^{-9}	--	16
Yes	Recurrent	Yes	rs73228569	3:111614052_T/C	PHLDB2	intronic	0.12	0.021	1.5×10^{-8}	++	34
Yes	Recurrent	Yes	rs12646050	4:174634261_A/G	RANP6	intergenic	0.14	0.026	3.7×10^{-8}	++	1
Yes	Recurrent	Yes	rs34969716	6:18210109_G/A	KDM1B	intronic	0.095	0.017	3.1×10^{-8}	++	1
Yes	Recurrent	Yes	rs2684249	6:122392511_T/C	HSF2	intergenic	0.09	0.016	6.4×10^{-9}	++	1
No	Recurrent	Yes	rs758890	7:150655643_G/A	KCNH2	intronic	-0.094	0.016	7×10^{-9}	--	5
Yes	Recurrent	Yes	rs10993463	9:97807233_C/T	AOPEP	intronic	0.12	0.016	1.3×10^{-14}	++	88
Yes	Recurrent	Yes	rs2568119	11:20004957_G/A	NAV2	intronic	0.11	0.018	2×10^{-9}	++	26
Yes	Recurrent	Yes	rs3765618	11:128769876_C/G	KCNJ5	UTR3	-0.18	0.027	3.2×10^{-11}	--	12
Yes	Recurrent	Yes	rs12944882	17:37983492_T/C	IKZF3	intronic	0.088	0.015	1.2×10^{-8}	++	2
Yes	Recurrent	Yes	rs17608766	17:45013271_T/C	GOSR2	UTR3	-0.12	0.022	2.1×10^{-8}	--	2
Yes	Recurrent	Yes	rs2834618	21:36119111_T/G	CLIC6	intergenic	0.15	0.026	1.2×10^{-8}	++	1
Yes	Single	Yes	rs1223535129	5:172664353_CG/C	NKX2-5	intergenic	0.11	0.017	2.7×10^{-11}	?+	31
Yes	Single	Yes	rs28631169	14:23888183_C/T	MYH7	intronic	0.1	0.018	2.1×10^{-8}	++	2

Table 2. Atrial Fibrillation (AF) variants found to be significant in the GWAS meta analysis. The lead variant (lowest *p*-value) for each independent region is listed. The "AFfull" column describes if the variant (or any of the significant variants in this region) is also identified in the full AF GWAS meta analysis. The "AF Study" column shows in which of the studies the variant/region was found to be significant ("common" meaning significant in both the single and the recurrent GWAS meta analysis). The "Known" column reports if this region is previously known for AF. The subsequent columns are "rsID," position, nearest gene, and the function of the lead SNP as well as effect size, standard error, and *p*-value from the meta-analysis. "Dir" corresponds to the direction of the effect in the HUNT and UKBB GWAS meta-analysis respectively (an entry of ? meaning not included in the meta-analysis - see Additional File 4 for allele frequencies), and "Nsnps" shows the number of significant SNPs in the region.

299 of the single SNP hits uniquely associated with single or recurrent MI are shown in Additional File 1,
 300 Fig.S20-S39). However, as shown in Additional File 4, these variants are not HUNT-specific since they are
 301 reported with relative high frequencies in the general European population. Hence, although they are rare in
 302 the UKBB population and thereby not included in the META-analysis, including a different European study
 303 population could validate or dispute the effect identified here. Also, many of these regions are well-known
 304 for MI, further indicating that these findings might be valid.

305 Comparing the GWAS meta-analysis of all MI cases to MI-free controls, we find that all four regions that
 306 were identified in common for single and recurrent MI (regions in or close to the genes HPCAL1, LPA,
 307 CDKN2A and CXADR) were also found in the full MI GWAS meta-analysis. The two regions that were
 308 specifically associated with recurrent MI events (regions close to the MIA3 and NOVA1 genes) were also
 309 identified in the full MI GWAS, but nine of the regions specifically associated with single MI were not

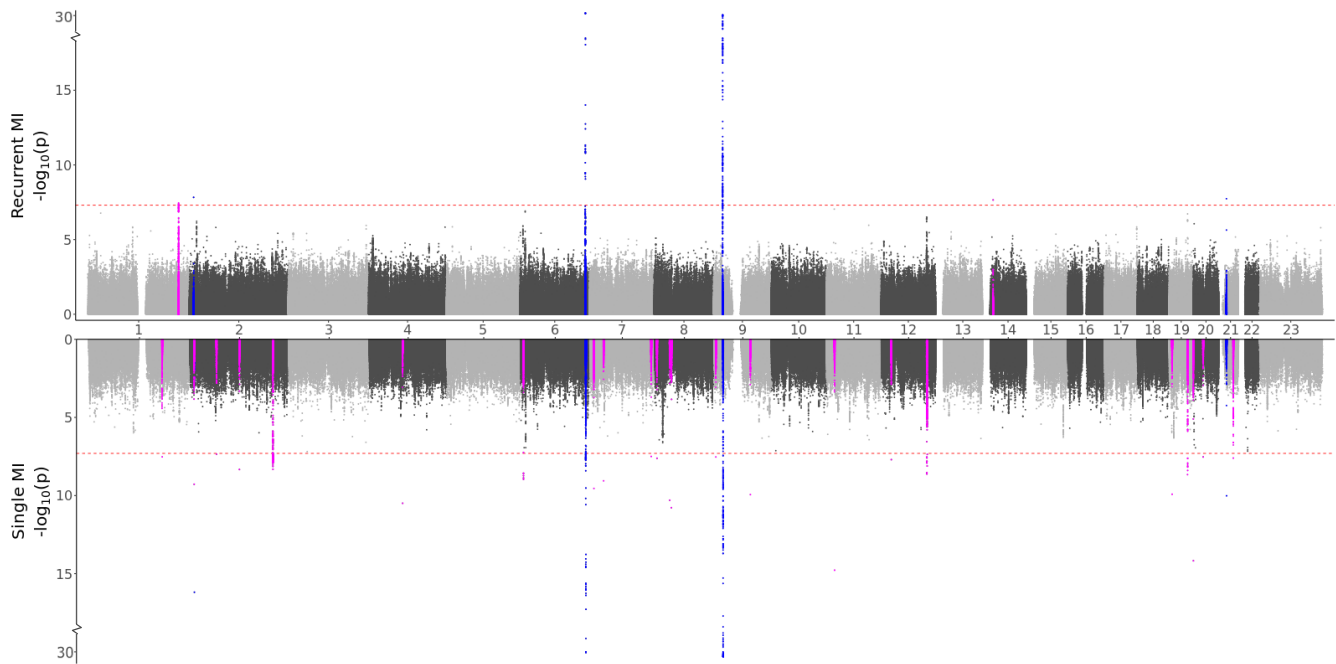


Figure 2. GWAS meta-analysis results for Myocardial Infarction (MI). Top: Comparing recurrent MI patients to MI-free controls. Bottom: Comparing single MI patients to MI-free controls. Blue colored spikes represent regions of SNPs found to be statistically significant in both GWAS studies (common), while magenta spikes represents statistically significant regions of SNPs that are specifically associated with the given MI group.

310 detected in the full MI GWAS (regions in or close to the OVAAL, TAF1B, GLI2, BBS9, MCPH1, GLIS3,
311 HECTD4, INSR and SYNDIG1). Hence, certain regions identified in the full MI GWAS are exclusively
312 linked to either single or recurrent MI, and some regions are only observed when patients with single MI
313 events are filtered out, emphasizing the need for dividing the MI groups.

314 Ten of the 24 regions specifically associated with single MI (the regions in or close to OVAAL, BMP3,
315 RIOK1, AC096553.5, MCPH1, KCNU1, GLIS3, TUT7, TRIB3 and SYNDIG1) are novel findings for
316 MI, and thus, have not previously been associated with CVD-related traits. While these regions consist
317 of only a single SNP (with the exception of the five SNPs close to the RIOK1 gene) and are only found
318 in the HUNT population (except for the SNP close to the KCNU1 gene), some of them are coding for
319 similar functions to the identified previously known MI genes. Three of these genes are commonly related
320 to malignancy (OVAAL, RIOK1, TUT7), similar to NBEAL1, in which we identify a known MI region
321 consisting of 254 SNPs uniquely associated with single MI. Other genes are coding for proteins involved in
322 calcium handling (BMP3, SYNDIG1) or associated with diabetes mellitus (GLIS3, TRIB3), which points
323 toward a more aggressive development of atherosclerosis. Similarly, both ATXN2 and HECTD4, for which
324 we identify known MI regions uniquely associated with single MI, are associated with diabetes mellitus.
325 The two regions exclusively linked to recurrent MI events are both found with negative effects in both
326 the HUNT and the UKBB population. The region close to the MIA3 gene has been associated with MI
327 previously, while the single SNP close to the NOVA1 gene (also with possible relations to malignancy) is a
328 novel finding. Taken together, these results highlight some interesting genes with a possible relation to MI.
329 Further studies are needed to investigate if similar effects are found in other European and non-European
330 populations and if these SNPs/genes link to MI and specifically link to either single or recurrent MI events.

Hall et al.

Mifull	MI Study	Known	rsID	lead	Gene	Function	Effect	StdErr	Pvalue	Dir	Nsnps
Yes	Common	No*	rs7600627	2:10543580_T/G	HPCAL1	intronic	0.41	0.049	6.5×10^{-17}	+	1
Yes	Common	Yes	rs55730499	6:161005610_C/T	LPA	intronic	0.3	0.031	1.4×10^{-21}	++	59
Yes	Common	Yes	rs1333049	9:22125503_G/C	CDKN2A	intergenic	0.18	0.016	3.8×10^{-27}	++	184
Yes	Common	No	rs2846882	21:18930412_T/A	CXADR	intronic	-0.29	0.044	9.6×10^{-11}	-?	1
No	Single	No	rs72720974	1:180559567_A/G	OVAAL	intergenic	0.23	0.042	3×10^{-8}	+	1
No	Single	No*	rs1620983	2:10028920_T/G	TAF1B	intronic	0.27	0.043	5.3×10^{-10}	+	1
Yes	Single	No*	rs2422293	2:65033227_A/G	SERTAD2	intergenic	0.26	0.047	4.4×10^{-8}	+	1
No	Single	No*	rs34337514	2:121458109_A/T	GLI2	intergenic	0.22	0.038	4.7×10^{-9}	+	1
Yes	Single	Yes	rs147932234	2:203951846_T/C	NBEAL1	intronic	-0.15	0.025	4.8×10^{-9}	--	254
Yes	Single	No	rs2868125	4:81955408_T/G	BMP3	intronic	0.27	0.041	3.2×10^{-11}	+	1
Yes	Single	No	rs576614613	6:7442904_T/C	RIOK1	intergenic	-0.26	0.043	1.1×10^{-9}	-?	5
Yes	Single	No	rs34945893	7:9691418_A/T	AC096553.5	intergenic	0.26	0.041	2.9×10^{-10}	+	1
No	Single	No*	rs1544558	7:33508624_T/G	BBS9	intronic	0.26	0.042	8.8×10^{-10}	+	1
Yes	Single	Yes	rs3918226	7:150690176_C/T	NOS3	intronic	0.17	0.03	3.2×10^{-8}	++	1
No	Single	No	rs7016007	8:5781499_G/T	MCPH1	intergenic	-0.28	0.05	2.4×10^{-8}	-?	1
Yes	Single	No	rs56097134	8:37067371_G/A	KCNU1	intergenic	-0.24	0.037	5×10^{-11}	--	1
Yes	Single	No*	rs1472790078	8:41141915_A/G	SFRP1	intronic	0.28	0.042	1.7×10^{-11}	+	1
No	Single	No	rs10814885	9:4209824_C/G	GLIS3	intronic	0.24	0.044	2.9×10^{-8}	+	1
Yes	Single	No	rs111991434	9:89155924_A/G	TUT7	intergenic	0.28	0.043	1.2×10^{-10}	+	1
Yes	Single	No*	rs58771640	11:19767929_T/A	NAV2	intronic	-0.34	0.043	1.7×10^{-15}	-?	1
Yes	Single	No*	rs4625573	12:24378427_T/C	SOX5	intronic	0.27	0.048	2×10^{-8}	+	1
Yes	Single	Yes	rs1876263690	12:111907431_A/AC	ATXN2	intronic	0.098	0.016	2.4×10^{-9}	++	10
No	Single	Yes	rs77215829	12:112618346_A/C	HECTD4	intronic	0.14	0.024	7.8×10^{-9}	++	2
No	Single	No*	rs2352955	19:7152404_A/G	INSR	intronic	0.23	0.035	1.2×10^{-10}	++	1
Yes	Single	Yes	rs55766194	19:45013423_A/G	TOMM40	intronic	0.29	0.048	2.2×10^{-9}	+	9
Yes	Single	No	rs6076475	20:360933_A/C	TRIB3	upstream	0.41	0.052	6.7×10^{-15}	+	1
No	Single	No	rs1569677	20:24404965_A/G	SYNDIG1	intergenic	0.22	0.04	3×10^{-8}	+	1
Yes	Single	Yes	rs28451064	21:35593827_G/A	MRPS6	intergenic	0.14	0.024	2.5×10^{-8}	++	1
Yes	Recurrent	Yes	rs4618978	1:222779187_C/G	MIA3	intergenic	-0.15	0.028	3.6×10^{-8}	--	5
Yes	Recurrent	No	rs59875208	14:26557764_C/A	NOVA1	intergenic	-0.23	0.042	2.2×10^{-8}	--	1

Table 3. Myocardial Infarction (MI) variants that are significant in the GWAS meta analysis. The lead variant (lowest *p*-value) for each independent region is listed. The "Mifull" column describes if the given variant (or any of the significant variants in this region) is also identified in the full MI GWAS meta analysis. The "MI Study" column shows in which of our studies the variant/region was found to be significant (an entry of "common" means that it was present in both the single and the recurrent GWAS meta analysis). The "Known" column shows if this region is previously related to MI ("Yes"), a relevant CVD-trait ("No*"), or neither ("No"). The next columns are "rsID," nearest gene, the function of the lead SNP as well as effect size, standard error, and *p*-value from the meta-analysis. "Dir" corresponds to the direction of the effect in the HUNT and UKBB GWAS meta-analysis respectively (a value "?" indicates that it was not included in the meta-analysis - see Additional File 4 for allele frequencies), and "Nsnps" shows the number of significant SNPs in the region.

331 3.3 Identification of additional phenotypes affected by SNPs through PheWas

332 To delve deeper into the genetic distinctions observed between single and recurrent AF and MI, we
 333 conducted a PheWAS (Phenome-Wide Association Study) analysis. This enabled us to pinpoint other
 334 phenotypes that are associated with the same set of SNPs that were designated as either common or unique
 335 for single and recurrent AF and MI.

336 Our PheWAS investigation of the SNPs identified as common for both single and recurrent AF revealed
 337 a total of 1,903 SNPs which were linked with 236 phenocodes (shown in Fig. 3 and Additional File
 338 2). Not surprisingly, the strongest associations were found for *Atrial fibrillation and flutter* and *Cardiac*
 339 *dysrhythmia* (*p*-value 10^{-400} and 10^{-220}). Furthermore, we discerned robust associations to phenocodes
 340 related to *Appendiceal conditions* and *Coagulation defects*. Notably, the circulatory system category
 341 emerged as the predominant category, encompassing 54 phenocodes. This includes, but is not limited
 342 to, conditions such as *Phlebitis and thrombophlebitis*, *Sinoatrial node dysfunction (Bradycardia)*, *Heart*
 343 *failure*, and *Hypertension*.

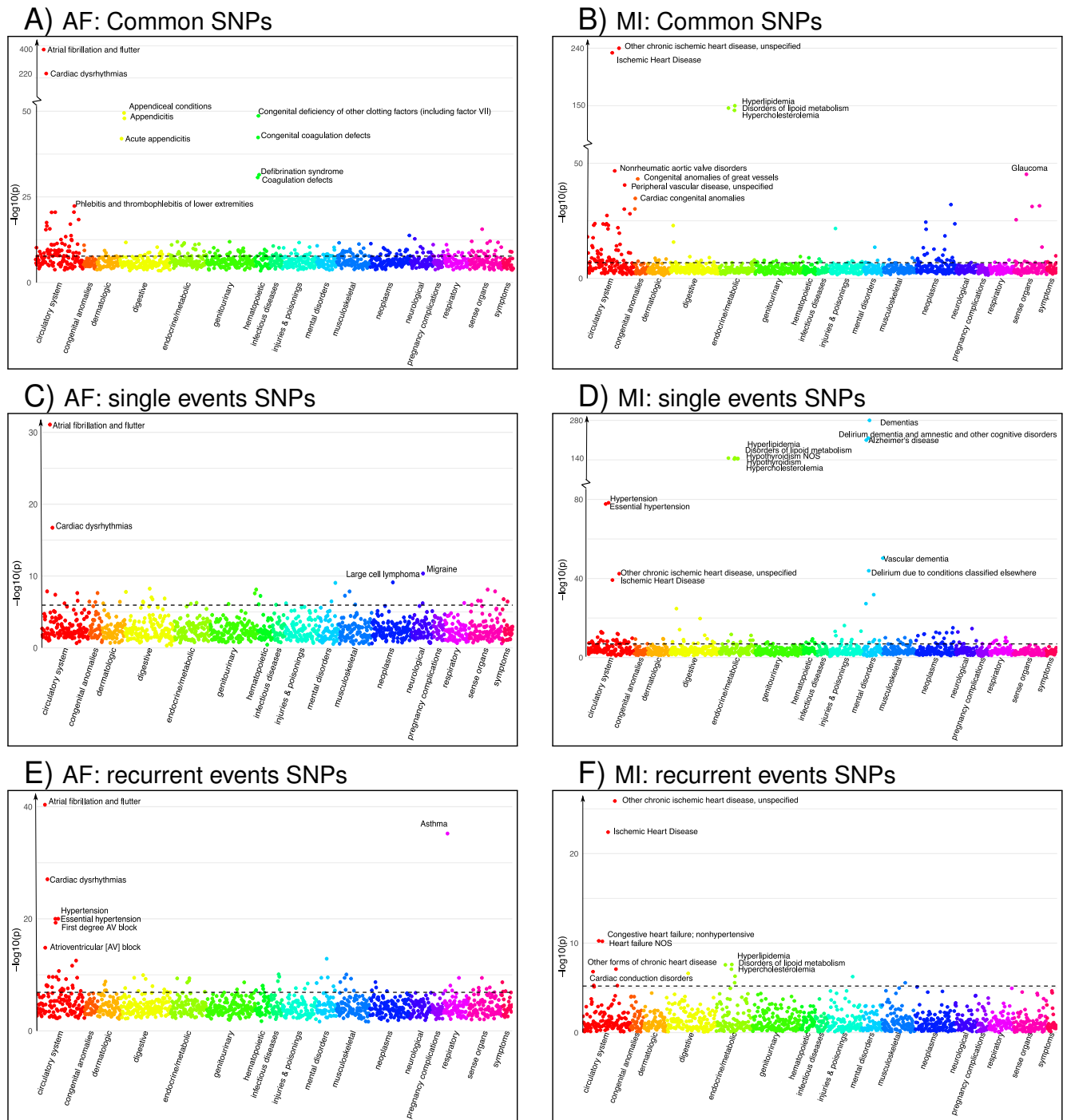


Figure 3. Phenocodes associated with each set of SNPs found for both single and recurrent AF/MI or uniquely for one of them. The *x*-axis shows each of the 1326 phenocodes sorted by phenocode category, and the *y*-axis shows the lowest *p*-value for the association between the phenocode and the SNPs in the set. The dotted line shows the threshold for significant associates, which vary according to the number of SNPs in each set. A) set of 1,903 SNPs found in common for both single and recurrent AF. B) set of 245 SNPs found in common for both single and recurrent MI. C) set of 33 SNPs found uniquely for single AF events. D) set of 299 SNPs found uniquely for single MI events. E) set of 286 SNPs found uniquely for recurrent AF events. F) set of 6 SNPs found uniquely for recurrent MI events.

344 Intriguingly, the two identified regions specifically associated with single AF consist of 33 SNPs
345 that exhibit significant association with 44 phenocodes (see Additional File 2). Not surprisingly, the
346 strongest associations for these SNPs pertain to the phenocodes *Atrial fibrillation and flutter* and *Cardiac*
347 *dysrhythmias*, the remaining 42 phenocodes span a diverse array of phenocode categories. These include
348 not only *Migraine* and *Large cell lymphoma*, but also conditions such as *Arrhythmia (cardiac) NOS*,
349 *Paroxysmal supraventricular tachycardia*, and *Cerebral atherosclerosis*.

350 In contrast, the 18 regions specifically associated with recurrent AF events consisting of 286 SNPs
351 show significant association with 91 phenocodes (see Additional File 2), and a majority of the strong
352 associations pertain to phenocodes of the circulatory system category. Again, the phenocode with the
353 most potent associations are *Atrial fibrillation and flutter*, and *Cardiac dysrhythmias*. Additionally, these
354 SNPs also display strong significant associations with *Asthma* and 27 phenocodes from the circulatory
355 system, including conditions such as *Hypertension*, *Atrioventricular block*, *Cardiomyopathy*, *Heart failure*,
356 *Ischemic heart disease*, *Cardiac arrest*, and *Palpitations*. Collectively, these results underscore genetic
357 susceptibility disparities between patients experiencing single versus recurrent AF events. Specifically,
358 SNPs specifically tied to recurrent AF are linked to a broad range of phenocodes related to the heart and
359 circulatory system compared to SNPs specifically associated to single AF events.

360 Regarding MI, we identified four regions associated with both single and recurrent MI, comprising 245
361 SNPs that exhibit significant associations to 144 phenocodes (see Additional File 2). The most prominent
362 associations are observed with *Ischemic heart disease* and *Hyperlipidemia* disorders. Additionally,
363 numerous diseases within the circulatory system category, such as *Nonrheumatic aortic valve disorders*,
364 *Peripheral vascular disease*, *Stricture of artery*, *Hypertension*, and *Heart valve disorders*, are also strongly
365 associated.

366 The 24 regions specifically identified for single MI events consist of 299 SNPs that are associated with
367 128 phenocodes (see Additional File 2). These include *Ischemic heart disease*, *Hypertension*, and diseases
368 of *Hyperlipidemia*. Additionally, there are strong associations with neurodegenerative disorders, such
369 as *Dementia*, *Alzheimer's*, and *Delirium*. These SNPs are furthermore linked with 33 phenocodes from
370 the circulatory system category, highlighting conditions such as *Cerebral ischemia*, *Cardiac conduction*
371 *disorders*, *Heart failure*, *Aortic valve disease*, and *Pulmonary heart disease*.

372 Notably, the two regions consisting of 6 SNPs specifically identified for recurrent MI were associated
373 with a mere 16 phenocodes (see Additional File 2). While these included *Ischemic heart disease*, *Heart*
374 *failure*, *Cardiac conduction disorders*, and diseases of *Hyperlipidemia*, they lacked the other 27 circulatory
375 system disorders identified for the single MI SNPs. Once again, these findings emphasise the genetic
376 differences between patients experiencing single and recurrent MI. SNPs specifically associated with single
377 MI events appear to be associated with a broader and more diverse range of cardiovascular disorders
378 compared to those solely linked to recurrent MI.

379 **3.4 Gene sets and co-expression network neighbourhood**

380 In our final analysis, we leverage multiple sets of gene expression data from the GTEx consortium
381 (30) measured in tissue sub-types taken from heart, muscle, skeletal, artery, and kidney to generate gene
382 co-expression networks (see Methods for details). Here, our expectation is that highly correlated genes
383 have a regulatory relationship or similar response in a condition (29). Thus, this approach should uncover
384 genes that display an expression profile that most closely links to the set of target genes found through our
385 GWAS analyses, and we investigate their functions.

386 3.4.1 AF associated genes in co-expression networks

387 Differential gene expression analysis of the 18 genes identified in recurrent AF (listed as Recurrent
 388 in Tab. 2) reveals a significant up-regulation of these genes in atrial appendage tissues from the heart.
 389 Furthermore, elevated expression levels are discerned in left ventricular heart, artery tibial, and muscle
 390 skeletal tissues (see Additional File 1, Fig. S9). Gene ontology analysis indicates that this set of genes
 391 is significantly enriched for cell-cell signaling involved in cardiac conduction, (Fold Enrichment (FE)
 392 > 100 , $FDR = 1.17 \times 10^{-2}$), cardiac muscle cell action potential ($FE = 70.03$, $FDR = 3.15 \times 10^{-2}$), and
 393 regulation of heart rate ($FE = 43.99$, $FDR = 1.62 \times 10^{-2}$).

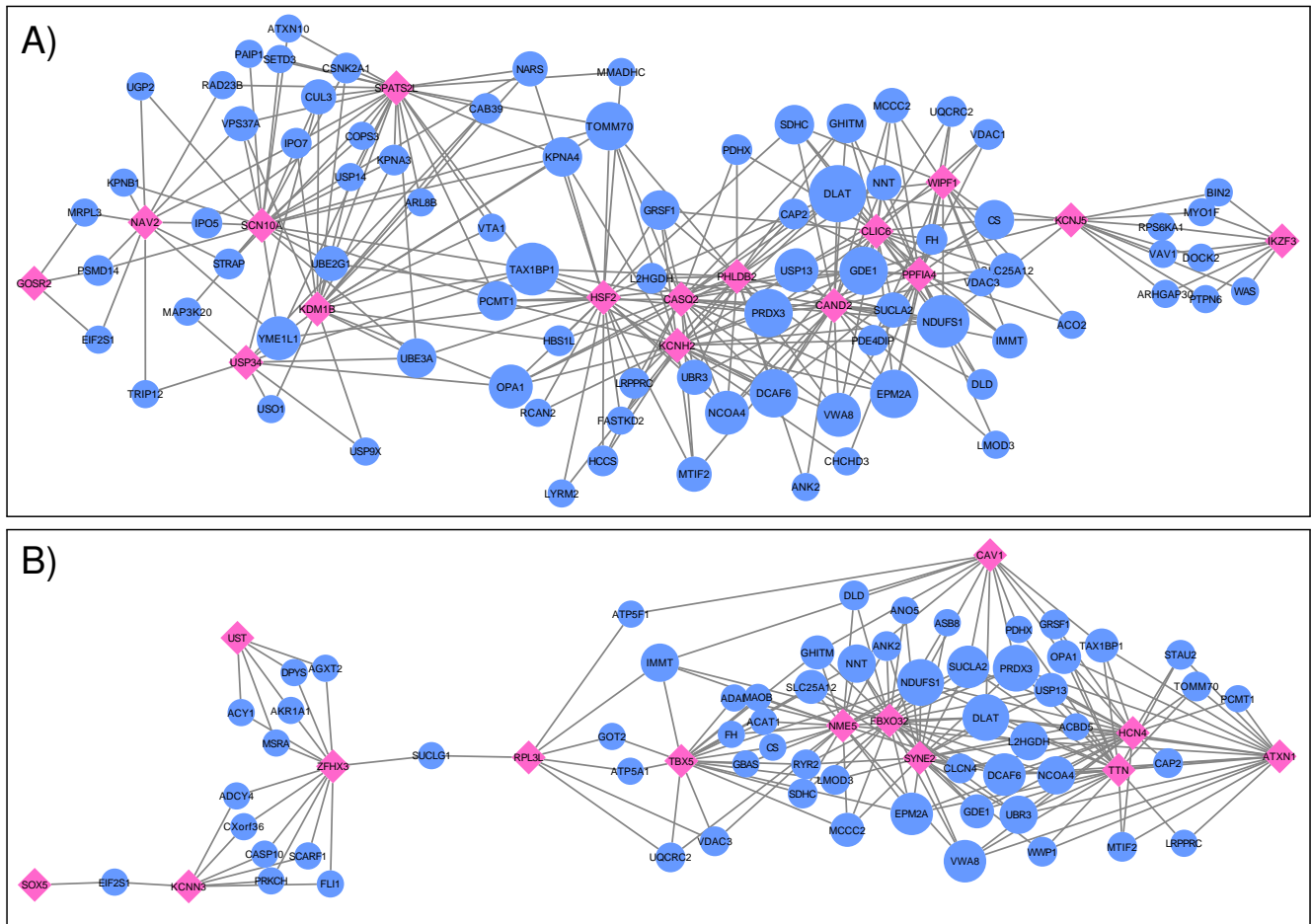


Figure 4. Networks showing the strongest shared neighbourhood of co-expressed genes for the GWAS (target) genes associated with A) recurrent AF uniquely and B) both single and recurrent AF. Pink diamond nodes represent the target genes and blue circular nodes represent the neighbouring genes. The sizes of the blue nodes are scaled according to their number of nearest neighbours in the network.

394 Following the co-expression analysis approach detailed in the Methods section, we find that 16 of the 18
 395 recurrent AF genes show strong co-expression with other genes in heart, artery, kidney, and muscle skeletal
 396 tissues. Selecting the top-25 genes with the strongest connection to each of the 16 target genes, Fig. 4 A)
 397 shows that all of the 16 target genes are connected through 82 shared neighbouring genes (see Additional
 398 File 3), i.e. the 82 shared genes are among the top-25 strongest connections for two or more of the
 399 target genes. These 82 neighbouring genes are significantly enriched for a variety of biological processes,
 400 including acetyl-CoA biosynthetic process from pyruvate ($FE > 100$, $FDR = 4.41 \times 10^{-3}$), tricarboxylic

401 acid cycle (FE= 54.70, FDR = 1.38×10^{-6}), NLS-bearing protein import into nucleus (FE = 50.99,
402 FDR = 1.93×10^{-3}), inner mitochondrial membrane organization (FE = 32.73, FDR = 9.49×10^{-4}),
403 respiratory electron transport chain (FE = 10.72, FDR = 4.32×10^{-2}), regulation of proteasomal protein
404 catabolic process (FE = 7.49, FDR = 4.94×10^{-2}), proteasome-mediated ubiquitin-dependent protein
405 catabolic process (FE = 5.58, FDR= 3.79×10^{-2}), and regulation of cellular catabolic process (FE = 4.01,
406 FDR= 9.85×10^{-3}).

407 Focusing on the two genes specific to single AF events (listed as *Single* in Tab. 2), our analysis reveals
408 that these genes shows significant up-regulation in left ventricle tissues of the heart and also high expression
409 levels for atrial appendage tissues of the heart (see Additional File 1, Fig. S10). Gene ontology analysis
410 confirms that these genes are closely linked to adult heart development (FE> 100, FDR = 7.77×10^{-3}),
411 ventricular cardiac muscle tissue morphogenesis (FE> 100, FDR = 4.17×10^{-2}), myofibril assembly
412 (FE> 100, FDR = 2.98×10^{-2}), cardiac muscle contraction (FE> 100, FDR = 2.47×10^{-2}), and
413 regulation of striated muscle contraction (FE> 100, FDR = 2.87×10^{-2}). Thus, although both target
414 genes exhibit the specified enriched functions, an egocentric network analysis reveals that they do not
415 share mutual genes with strong co-expression across the heart, artery, kidney, and skeletal muscle tissues.
416 Therefore, while they may have functional overlap, their co-expressing gene partners diverge for each
417 target gene.

418 In our comparative analysis of gene sets uniquely associated with either single or recurrent AF, we also
419 evaluated genes that were consistent across both AF categories. Among the 16 genes identified for both
420 AF groups (labelled as *Common* in Tab. 2), there was a significant up-regulation in tissues of the heart's
421 left ventricle, atrial appendage, and skeletal muscles (see Additional File 1, Fig. S11). Notably, these
422 genes did not show any enrichment in specific gene ontology categories. Upon conducting an egocentric
423 network analysis, we found that 14 out of these 16 genes displayed strong co-expression patterns with
424 genes differentially expressed in the heart, arteries, kidneys, and skeletal muscles. Moreover, 13 of these
425 genes were interconnected via 61 shared neighbor genes (see Fig.4 B) and Additional File 3). The 61 genes
426 found in common for the 13 target genes show significant enrichment for fumarate metabolic process (FE
427 > 100, FDR = 1.83×10^{-2}), regulation of atrial cardiac muscle cell action potential (FE > 100, FDR
428 = 2.51×10^{-2}), regulation of mitochondrial RNA catabolic process (FE> 100, FDR = 3.36×10^{-2}),
429 mitochondrial acetyl-CoA biosynthetic process from pyruvate (FE > 100, FDR = 3.32×10^{-2}), succinyl-
430 CoA catabolic process (FE > 100, FDR = 4.10×10^{-2}), tricarboxylic acid cycle (FE = 73.80, FDR
431 = 5.34×10^{-8}), branched-chain amino acid catabolic process (FE = 46.69, FDR = 1.11×10^{-2}), inner
432 mitochondrial membrane organization (FE= 35.33, FDR = 2.16×10^{-3}), aerobic electron transport chain
433 (FE = 15.20, FDR = 2.98×10^{-2}), mitochondrial ATP synthesis coupled electron transport (FE = 14.37,
434 FDR = 3.50×10^{-2}), and alpha-amino acid catabolic process (FE = 14.21, FDR = 3.53×10^{-2}).

435 Thus, the gene sets identified in common and uniquely for single and recurrent AF events are up-regulated
436 in atrial appendage tissues of the heart, left ventricular tissues of the heart, and muscle skeletal tissues,
437 where the gene set identified for recurrent AF and single AF events show significant up-regulation in atrial
438 appendage and left ventricular tissues of the heart, respectively, while the gene set identified for both single
439 and recurrent AF shows significant up-regulation of all three tissues. While both gene sets specifically
440 identified for single and recurrent AF events show enrichment for functions related to heart and cardiac
441 muscle, the recurrent AF gene set shows interesting enrichment of functions related to regulation of heart
442 rate and cardiac muscle cell action potential. Collectively, these findings underscore the distinct genetic
443 underpinnings between patients experiencing isolated versus recurrent AF events, with the recurrent AF
444 genes revealing more intricate processes.

445 3.4.2 Network MI

446 The 24 nearest genes to the regions specifically identified for single MI events (listed as Single in Tab. 3)
447 show no significant enrichment for any gene ontology terms. Differential gene expression analysis shows
448 no significant up or down regulation of these genes in any tissue, but significant expression levels in tibial
449 nerve and high expression levels in tibial artery and aorta artery tissues (see Additional File 1, Fig S12).

450 When inspecting gene co-expression in heart, artery, kidney, and muscle skeletal tissues, we find that
451 21 of these genes show high co-expression with other genes in these tissues. Again creating egocentric
452 networks for each of these 21 target genes, we find that all 21 target genes cluster in a giant component,
453 where 110 genes are connected to two or more of them. The network depicted in Fig. 5 A) illustrates genes
454 that are interconnected with multiple single MI target genes. Notably, there's a dense cluster in the upper
455 left portion of the figure, dominated by GLI2, GLIS3, TRIB3, and SFRP1, all of which are interconnected
456 and share a majority of their neighborhood genes. The known associations of both GLI2 and SFRP1 with
457 CVD-related traits, coupled with the detection of TRIB3 and SFRP1 in the comprehensive MI GWAS,
458 suggest potential shared functional roles of these four genes. Furthermore, they also exhibit significant
459 interconnections with BBS9, HECTD4, ATXN2, and SYNDIG1. It's worth noting that HECTD4 and
460 ATXN2 have recognized associations with MI.

461 Gene ontology analysis of the 110 genes shared between two or more target genes (see Additional File
462 3) show significant enrichment for positive regulation of several functional categories, more specifically
463 establishment of protein localization to telomere ($FE = 54.18, FDR = 1.95 \times 10^{-2}$), positive regulation
464 of protein localization to Cajal body ($FE = 49.26, FDR = 2.28 \times 10^{-2}$), NLS-bearing protein import
465 into nucleus ($FE = 38.02, FDR = 3.85 \times 10^{-3}$), positive regulation of telomerase RNA localization
466 to Cajal body ($FE = 36.12, FDR = 4.17 \times 10^{-2}$), cristae formation ($FE = 36.12, 4.00 \times 10^{-2}$),
467 tricarboxylic acid cycle ($FE = 29.13, FDR = 1.59 \times 10^{-3}$), regulation of proteasomal protein catabolic
468 process ($FE = 8.38, FDR = 1.63 \times 10^{-3}$), regulation of ubiquitin-dependent protein catabolic process
469 ($FE = 7.35, FDR = 2.27 \times 10^{-2}$), proteasome-mediated ubiquitin-dependent protein catabolic process
470 ($FE = 4.68, FDR = 4.76 \times 10^{-2}$), purine-containing compound metabolic process ($FE = 4.21, FDR =$
471 4.69×10^{-2}), nucleotide metabolic process ($FE = 3.84, FDR = 4.69 \times 10^{-2}$), and cellular component
472 biogenesis ($FE = 2.06, FDR = 3.57 \times 10^{-2}$).

473 Upon analyzing the two genes uniquely associated with recurrent MI (listed as Recurrent in Tab. 3), we
474 observe neither significant enrichment in gene ontology terms nor any distinctive expression patterns across
475 tissue types (see Additional File 1, Fig. S13). In constructing egocentric networks for these genes, we
476 identify seven overlapping genes (see Additional File 3) among the top-25 most correlated for each gene,
477 as depicted in Fig. 5 B). While four of these seven neighboring genes (TOMM70, FASTKD2, MMADHC
478 and OPA1) are related to mitochondrial function and energy production, gene ontology analysis yielded no
479 significant enrichment.

480 Investigating the genes that are in common for both MI groups (listed as Common in Tab. 3), we find
481 no significant differential co-expression in any of the tissues (see Additional File 1, Fig. S14). No gene
482 ontology terms were found to be significantly enriched for these four genes. The egocentric networks
483 (Fig. 5 C)) revealed that three of them are linked through 18 shared neighborhood genes (see Additional
484 File 3). The 18 genes connecting the three target genes show significant enrichment for mitochondrial
485 acetyl-CoA biosynthetic process from pyruvate ($FE > 100, FDR = 5.92 \times 10^{-3}$), tricarboxylic acid
486 cycle ($FE > 100, FDR = 5.37 \times 10^{-7}$), negative regulation of release of cytochrome c from mitochondria
487 ($FE > 100, FDR = 4.88 \times 10^{-2}$), cell redox homeostasis ($FE = 81.70, FDR = 3.38 \times 10^{-3}$), aerobic

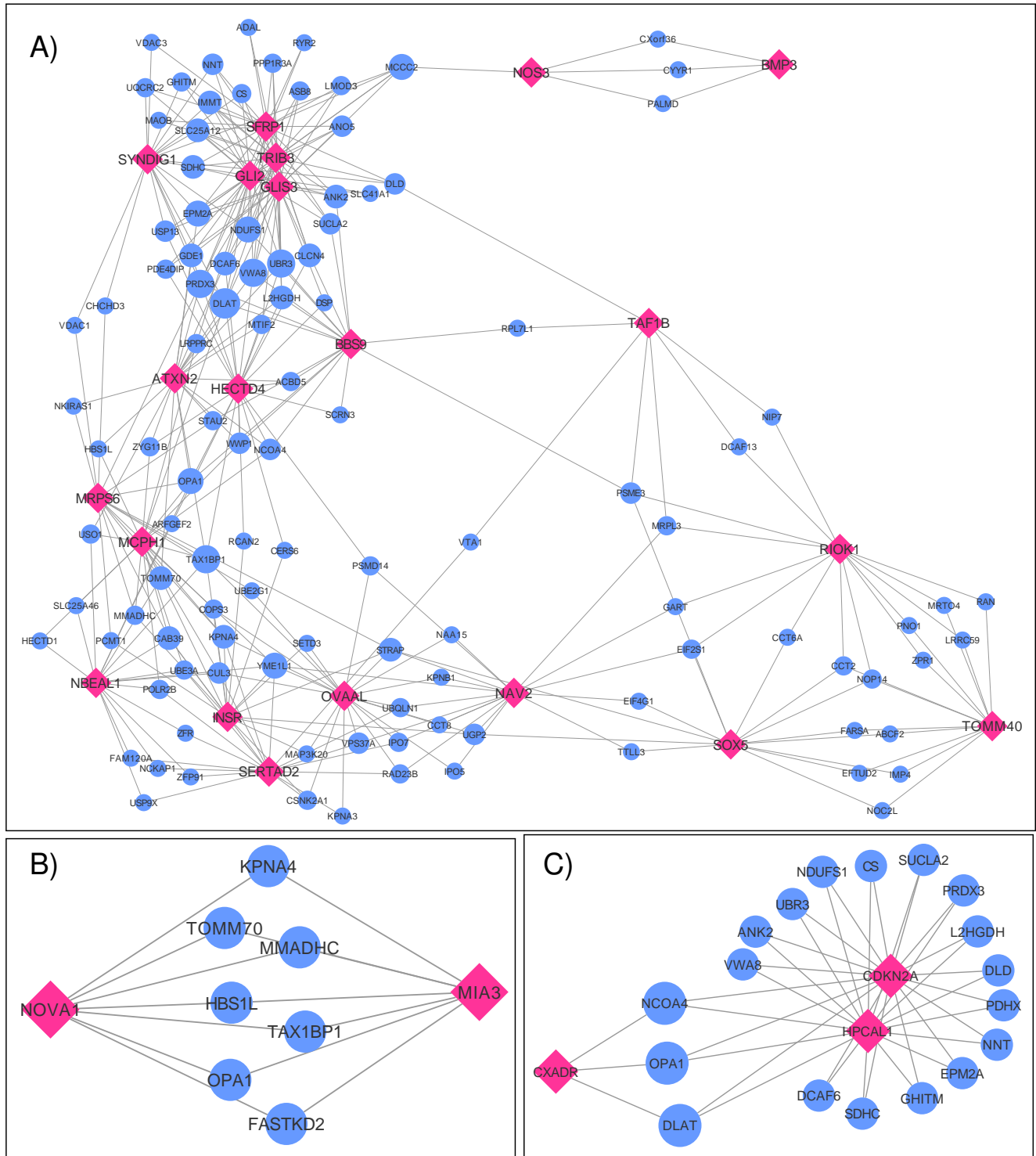


Figure 5. Networks showing the strongest shared neighbourhood of co-expressed genes for the GWAS (target) genes associated with A) single MI uniquely, B) recurrent MI uniquely and C) both single and recurrent MI. Pink diamond nodes represent the target genes and blue circular nodes represents the neighbouring genes. The sizes of the blue nodes are scaled according to their number of nearest neighbours in the network.

488 electron transport chain (FE = 39.90, FDR = 2.07×10^{-2}), mitochondrial ATP synthesis coupled electron

489 transport (FE = 37.71, FDR = 2.38×10^{-2}), reactive oxygen species metabolic process (FE = 32.68,
490 FDR = 3.25×10^{-2}), and mitochondrion organization (FE = 12.85, FDR = 1.15×10^{-2}).

491 In summary, MI gene sets do not exhibit notable tissue specificity or functional enrichment. Yet, the
492 gene sets linked to both common and single MI incidents share multiple genes within their co-expression
493 network neighborhoods. These shared genes exhibit functional enrichment for several pertinent processes.
494 Conversely, the shared gene neighborhood specifically related to recurrent MI does not present any
495 functional enrichment. However, some of the genes are related to mitochondrial function and energy
496 production, similar to several of the enriched functions for the shared neighbouring genes for common MI
497 genes. Notably, the neighbouring genes of the single MI genes shows significant enrichment for several
498 function not identified for the recurrent of common neighbouring gene sets. Only one of the biological
499 processes identified in the neighborhood of single MI genes is also seen in the neighborhood of common
500 genes. This suggests unique functions associated with genes specific to single MI incidents, shedding light
501 on potential reasons why certain patients experience only one MI event.

4 DISCUSSION

502 The study demonstrates distinct clinical characteristics and genetic predispositions between patients who
503 experience a single atrial fibrillation/myocardial infarction (AF/MI) event and those with recurrent events.
504 Notably, the single AF and single MI patient groups represent relapse-free patients, where we have filtered
505 the data to ensure they have only experienced one event, with a buffer of at least 5 years for AF and 7 years
506 for MI since the episode and before the study either ended or the patient died.

507 Single AF incidents appear more influenced by lifestyle factors and age, with only two unique genetic
508 regions identified. In contrast, patients with recurrent AF are typically younger at their first event and exhibit
509 a stronger genetic basis, as evidenced by 18 unique genetic regions linked to recurrent AF. Among these,
510 two regions are near the NAV2 and SCN10A genes, previously hypothesized, but unconfirmed (36, 35)
511 ($N = 42$, 585 East Asian population, $N = 660$ German population), to affect recurrent AF. The region with
512 the lowest p -values and also the largest amount of significant SNPs was found near the PITX2 gene for both
513 the single and the recurrent AF groups. Though PITX2 is the most known gene associated with recurrent
514 AF (10, 7, 1) ($N = 295$ Turkish population, $N = 195$ Caucasian population, $N = 991$ German and
515 American population), our results give no indication of such an effect. This observation extends to six other
516 genes (SOX5, CAV1, EPHX2, ITGA9, SLC8A1, and TBX5) (35, 10, 8) ($N = 660$ German population,
517 $N = 295$ Turkish population, $N = 42$, 585 East Asian population, $N = 486$ Caucasian population), where
518 variants near the SOX5, CAV1 and TBX5 genes are identified in both the single and recurrent AF group,
519 and no variants near the EPHX2, ITGA9 and SLC8A1 genes are identified in any of the groups. PheWAS
520 analysis of these unique single and recurrent AF SNPs further highlight distinct susceptibilities: SNPs
521 associated with single AF correlate with 44 phenocodes across various categories, whereas recurrent AF
522 SNPs are linked to 91 phenocodes, predominantly in the circulatory system category. Network analysis of
523 gene sets revealed that 16 of the 18 genes associated with recurrent AF are connected through 82 shared,
524 highly co-expressed neighboring genes. These recurrent AF genes, along with their neighboring genes, are
525 involved in complex processes related to heart rate regulation and cardiac muscle cell action potential. In
526 contrast, the two genes associated with single AF events are linked to heart and cardiac muscle processes
527 but do not share highly co-expressed genes.

528 We also find distinct clinical and genetic differences between patients with single and recurrent MI.
529 Unlike AF, recurrent MI is more associated with older age at first MI event, lifestyle factors, and age-related

530 issues. This is despite the fact that the single MI group is adjusted for early death due to MI and/or related
531 comorbidities, and this should not affect the results. The genetic predisposition seems stronger in single
532 MI cases, with a total of 24 uniquely associated regions. In contrast, recurrent MI is only associated with
533 two regions that did not share the association with single MI cases. Of the 24 genetic regions uniquely
534 identified for single MI, 14 are previously known for MI or other CVD-related traits. The remaining 10 are
535 novel for MI and have previously not been reported for other CVD-related traits. While most of these novel
536 regions consist of single SNPs primarily identified in the HUNT population, their nearest genes code for
537 functions similar to known MI regions. Looking into the allele frequencies in each population (Additional
538 File 4), we see that these variants are rather common in HUNT and in the general European population
539 (based on reporting from gnomAD (37)), while they are rare variants in the UKBB population and thereby
540 not included in the META-analysis. Further studies are needed to investigate if these variants show similar
541 effects in other European and non-European populations.

542 PheWAS analysis reinforces the genetic distinction between single and recurrent MI groups. The single
543 MI group's SNPs are linked to 128 phenotypes, predominantly in the circulatory system category, with
544 additional associations in the endocrine/metabolic category and notable links to neurodegenerative disorders.
545 In contrast, the SNPs related to recurrent MI correlate with 16 phenocodes, involving both circulatory and
546 endocrine/metabolic categories, but the associations are not as pronounced as those in the single MI group.
547 Our network analysis reveals distinct gene interactions for each group. Of the 24 genes uniquely associated
548 with single MI, 21 share connections with 110 genes in the co-expression network. However, the two genes
549 associated with recurrent MI have seven highly co-expressed genes. This indicates more extensive genetic
550 interconnections in single MI cases. Interestingly, the shared neighbouring genes for single MI, and those
551 common to both single and recurrent MI, show functional enrichment in several biological processes. In
552 contrast, the shared neighboring genes for recurrent MI do not show significant functional enrichment. This
553 disparity suggests distinct biological pathways involved in single versus recurrent MI events. Moreover,
554 only one function is common between the shared genes for both single and recurrent MI, suggesting unique
555 biological mechanisms specific to single MI events.

556 The results suggest a greater genetic influence in atrial fibrillation (AF) compared to myocardial infarction
557 (MI), but several factors could affect this perception. Clinically, it's expected that more genes would increase
558 the risk of recurrent AF, a pattern observed in this study. In contrast, the findings for MI are opposite,
559 potentially influenced by their higher ages: MI becomes less common in younger individuals, and in
560 older populations, comorbidities often overshadow genetic risk factors. There could also be physiological
561 reasons behind these observations. For instance, MI in younger individuals might more frequently result
562 from genetic factors related to platelet aggregation or atherosclerosis, conditions that are generally more
563 responsive to treatment. In older individuals, MI might be more associated with broader age-related issues,
564 reducing the relative impact of genetics. The reporting of these conditions could also influence the results,
565 with AF potentially being under-reported compared to MI, which itself is possibly over-reported. This
566 discrepancy could explain why phenocodes related to neurodegenerative disorders and cerebral ischemia
567 emerge as significant only in single MI cases in the PheWAS results. These findings might be influenced
568 by diagnostic practices where MI is often recorded as a cause of death, even when other diseases are the
569 actual cause, or in cases where individuals die before a recurrent event occurs. The latter has been adjusted
570 for by excluding individuals with single events with less than 7 years between the event and the censoring
571 date, being either death or the end of the study. Still, even with this adjustment, some single MI event cases
572 might be censored out before a second event, possibly influencing the the case groups and thereby the
573 results.

574 The observed differences in clinical characteristics for both diseases might be even more pronounced if
575 the HUNT and UK Biobank (UKBB) studies had used similar questionnaire formats. In the HUNT study,
576 participants involved in both HUNT2 and HUNT3 provided cholesterol and blood pressure measurements
577 taken eleven years apart. We opted to use the measurements recorded closest to the first AF or MI event,
578 as we believe these offer the most relevant information. However, for the UKBB participants, most were
579 measured only once, eliminating the option to choose measurements nearest to the event. This discrepancy
580 in data collection methodology could explain why significant differences in cholesterol and blood pressure
581 measurements are observed between single and recurrent event groups in the HUNT population but not in
582 the UKBB group. The lack of longitudinal data in the UKBB may obscure potential differences that are
583 more apparent in the HUNT study due to its repeated measurements.

584 In any GWAS study, larger sample sizes, diverse ancestry representation, and result replication are crucial.
585 The HUNT study, comprised solely of individuals of European ancestry, led us to select only European
586 ancestry participants from the UKBB for consistency. However, for global applicability, conducting similar
587 analyses across all ancestries is essential. Regarding sample sizes, while the combined participant pool
588 of HUNT and UKBB exceeds 500,000, the specific filtering and subgrouping in our study result in some
589 case/control groups having fewer than 2,000 individuals. This reduced size may limit our ability to achieve
590 robust significant findings. The small sample sizes might explain the prevalence of significant singleton
591 SNP hits (one SNP per region) found only in the HUNT study, with the same SNPs with too rare allele
592 frequencies in the UKBB population. Increasing sample sizes or including an additional population would
593 likely enhance the reliability of these findings, particularly since these variants are reported as common in
594 the general European population. While increasing the sample sizes could verify/dispute or even identify
595 additional regions, some genetic differences observed between recurrent and single cases might diminish
596 with larger sample sizes. Although many spikes in the Manhattan plots (Fig.1 and Fig.2) indicate clear
597 differences between single and recurrent cases, certain regions almost reach significance for the opposite
598 group, such as the hit on chromosome 8 for recurrent AF. This suggests that some observed genetic
599 distinctions might be less pronounced with a more substantial and diverse sample.

600 While our meta-analysis did not yield any GWAS significant results when directly comparing recurrent
601 to single AF or MI cases, the data still indicate genetic differences between these groups. Notably, some
602 genetic regions were identified in both single and recurrent groups (termed 'common'), while a substantial
603 number were unique to each group. This could suggest that the common regions may play a role in the
604 general susceptibility to AF and MI, whereas the unique regions could confer specific genetic risks for the
605 disease's form (single or recurrent events). Further studies are needed to test the hypotheses generated from
606 this study, in particular investigate the effect of the novel SNPs and genes identified and their involvement
607 in either single or recurrent AF/MI. While many of the identified genes and their related function might not
608 have an direct confirmed effect on AF/MI, we do see some similar functions of the novel genes as for the
609 known AF/MI genes, showing the potential of identification and understanding of new genetic functions of
610 the diseases.

611 There has been considerable research aimed at uncovering genetic causes for AF and MI. Our findings
612 underscore the importance of genetic studies focused on disease subgroups, as conducted here. Both AF
613 and MI are widespread diseases with varied impacts on individuals' lives. The progression and outcomes of
614 these diseases are not uniform across all patients. By examining subgroups within these diseases, we could
615 gain new insights into their mechanisms, potentially leading to more effective prevention and treatment
616 strategies tailored to different patient profiles.

DATA AVAILABILITY STATEMENT

617 The Trøndelag Health Study (HUNT) has invited persons aged 13 – 100 years to four surveys between
618 1984 and 2019. Comprehensive data from more than 140,000 persons having participated at least once
619 and biological material from 78,000 persons were collected. The data are stored in HUNT databank and
620 the biological material in HUNT biobank. HUNT Research Centre has permission from the Norwegian
621 Data Inspectorate to store and handle these data. The key identification in the data base is the personal
622 identification number given to all Norwegians at birth or immigration, whilst de-identified data are sent to
623 researchers upon approval of a research protocol by the Regional Ethical Committee and HUNT Research
624 Centre. To protect participants' privacy, HUNT Research Centre aims to limit storage of data outside
625 HUNT databank, and cannot deposit data in open repositories. HUNT databank has precise information on
626 all data exported to different projects and are able to reproduce these on request. There are no restrictions
627 regarding data export given approval of applications to HUNT Research Centre. For more information see:
628 <http://www.ntnu.edu/hunt/data>.

629 The UKBB characteristics and genetic data are available for all researchers upon requested.
630 To identify the number of events per participant, the HESIN-tables were applied. See
631 <https://biobank.ndph.ox.ac.uk/showcase/> for further information.

632 PheWas analysis were conducted on the Pan UKBB resource (24). See <https://pan.ukbb.broadinstitute.org>.
633 The mapping from ICD9 and ICD10 codes are available at <https://phewascalog.org/phecodes>.

634 Gene expression matrices were downloaded from the GTEx portal (30), V8. See <https://www.gtexportal.org>.

ETHICS STATEMENT

635 Participation in the HUNT Study is based on informed consent, and the study has been approved by the
636 Norwegian Data Inspectorate and the Regional Ethics Committee for Medical Research in Norway (REK:
637 2014/144). All participants from both UKBB and HUNT who have contributed in this study have provided
638 their written informed consent to participate.

AUTHOR CONTRIBUTIONS

639 IS identified the research topic. MH and AHS conceived and designed the study, with that defining the traits
640 and cohorts, planned the analysis and analysed the characteristics of the cohorts. MH conducted the analysis
641 and interpreted the results. MH wrote the manuscript. MH, EA, AHS, IS and HD edited and revised the
642 manuscript. MH, AHS, IS, HD and EA approved the final version of the manuscript.

FUNDING

643 M.H. and E.A. thank the K. G. Jebsen Foundation for Grant SKGJ-MED-015.

ACKNOWLEDGMENTS

644 The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Centre (Faculty of
645 Medicine and Health Sciences, Norwegian University of Science and Technology NTNU), Trøndelag
646 County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health.
647 The genetic investigations of the HUNT Study, is a collaboration between researchers from the K.G.
648 Jebsen center for genetic epidemiology and University of Michigan Medical School and the University

649 of Michigan School of Public Health. The K.G. Jebsen Center for Genetic Epidemiology is financed by
650 Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU, Norway. We would
651 like to thank clinicians and other employees at Nord-Trøndelag Hospital Trust for their support and for
652 contributing to data collection in this research project. We would also like to thank all participants that
653 have contributed in the UKBB study.

CONFLICT OF INTEREST STATEMENT

654 The authors confirm that there are no conflicts of interest. The authors declare that the research was
655 conducted in the absence of any commercial or financial relationships that could be construed as a potential
656 conflict of interest.

SUPPLEMENTAL DATA

657 Additional File 1 - Fig. S1-S14. S1: QQ-plot of single AF vs. healthy controls in HUNT. S2: QQ-plot of
658 single AF vs. healthy controls in UKBB. S3: QQ-plot of recurrent AF vs. healthy controls in HUNT. S4:
659 QQ-plot of recurrent AF vs. healthy controls in UKBB. S5: QQ-plot of single MI vs. healthy controls in
660 HUNT. S6: QQ-plot of single MI vs. healthy controls in UKBB. S7: QQ-plot of recurrent MI vs. healthy
661 controls in HUNT. S8: QQ-plot of recurrent MI vs. healthy controls in UKBB. S9: Tissue specificity of the
662 18 genes uniquely identified for recurrent AF. Test results in each row are Bonferroni corrected and red
663 bars show tissues where the set of genes are significantly up-regulated (top), down-regulated (middle), and
664 both-sided (bottom). S10: Tissue specificity for the two genes identified uniquely for single AF. Test results
665 in each row are Bonferroni corrected and red bars show tissues where the set of genes are significantly
666 up-regulated (top), down-regulated (middle), and both-sided (bottom). S11: Tissue specificity for the 16
667 genes identified for both single and recurrent AF. Test results in each row are Bonferroni corrected and red
668 bars show tissues where the set of genes are significantly up-regulated (top), down-regulated (middle), and
669 both-sided (bottom). S12: Tissue specificity for the 24 genes identified uniquely for single MI. Test results
670 in each row are Bonferroni corrected and red bars show tissues where the set of genes are significantly
671 up-regulated (top), down-regulated (middle), and both-sided (bottom). S13: Tissue specificity for the two
672 genes identified uniquely for recurrent MI. Test results in each row are Bonferroni corrected and red bars
673 show tissues where the set of genes are significantly up-regulated (top), down-regulated (middle), and
674 both-sided (bottom). S14: Tissue specificity for the four genes identified for both single and recurrent MI.
675 Test results in each row are Bonferroni corrected and red bars show tissues where the set of genes are
676 significantly up-regulated (top), down-regulated (middle), and both-sided (bottom). S15-S39: Regional
677 plot of single significant hit identified uniquely for recurrent AF, single MI or recurrent MI.

678 Additional File 2 - Excel file containing all phenocodes found to be significantly associated with the set
679 of SNPs associated with both single and recurrent AF/MI or uniquely for one of them.

680 Additional File 3 - Excel file containing genes shared between the GWAS associated genes (target genes)
681 in the network analysis for genes associated with both single and recurrent AF/MI or uniquely for one of
682 them.

683 Additional File 4 - Excel file containing allele frequency of SNPs identified in only one of the populations
684 (listed as for instance –? in Tab. 2 and 3). The columns named 'AF HUNT', 'AF UKBB' and 'AF
685 European' shows the allele frequency in the HUNT population, the UKBB population and as reported in
686 gnomAD (37)(<https://gnomad.broadinstitute.org/>), respectively. The 'HUNT/UKBB' columns shows if the

687 SNP is identified in HUNT or UKBB, and the 'imputation info' columns shows the imputation info given
688 in the genotyped UKBB files.

REFERENCES

- 689 1 .Shoemaker MB, Bollmann A, Lubitz SA, Ueberham L, Saini H, Montgomery J, et al. Common genetic
690 variants and response to atrial fibrillation ablation. *Circulation: Arrhythmia and Electrophysiology* **8**
691 (2015) 296–302. doi:10.1161/CIRCEP.114.001909.
- 692 2 .Official Statistics of Sweden H, Care M. Statistics on Myocardial Infarctions 2020. Tech. rep. (2021).
- 693 3 .[Dataset] Nielsen JB, Thorolfsdottir RB, Fritsche LG, Zhou W, Skov MW, Graham SE, et al. Biobank-
694 driven genomic discovery yields new insight into atrial fibrillation biology (2018). doi:10.1038/
695 s41588-018-0171-3.
- 696 4 .Christophersen IE, Rienstra M, Roselli C, Yin X, Geelhoed B, Barnard J, et al. Large-scale analyses of
697 common and rare variants identify 12 new loci associated with atrial fibrillation. *Nature Genetics* **49**
698 (2017) 946–952. doi:10.1038/ng.3843.
- 699 5 .Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000
700 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics* **47**
701 (2015) 1121–1130. doi:10.1038/ng.3396.
- 702 6 .Hartiala JA, Han Y, Jia Q, Hilser JR, Huang P, Gukasyan J, et al. Genome-wide analysis identifies
703 novel susceptibility loci for myocardial infarction. *European Heart Journal* **42** (2021) 919–933.
704 doi:10.1093/eurheartj/ehaa1040.
- 705 7 .Husser D, Adams V, Piorkowski C, Hindricks G, Bollmann A. Chromosome 4q25 Variants and Atrial
706 Fibrillation Recurrence After Catheter Ablation. *Journal of the American College of Cardiology* **55**
707 (2010) 747–753. doi:10.1016/j.jacc.2009.11.041.
- 708 8 .Wutzler A, Kestler C, Perrot A, Loehr L, Huemer M, Parwani AS, et al. Variations in the human
709 soluble epoxide hydrolase gene and recurrence of atrial fibrillation after catheter ablation. *International*
710 *Journal of Cardiology* **168** (2013) 3647–3651. doi:10.1016/j.ijcard.2013.05.010.
- 711 9 .Zhang X, Huang J, Li J, Lu Q, Huang Y, Lu D, et al. Association Between TCF21 Gene Polymorphism
712 with the Incidence of Paroxysmal Atrial Fibrillation and the Efficacy of Radiofrequency Ablation for
713 Patients with Paroxysmal Atrial Fibrillation. *International Journal of General Medicine* **Volume 15**
714 (2022) 4975–4983. doi:10.2147/ijgm.s366956.
- 715 10 .Ulus T, Dural M, Mese P, Yetmis F, Mert KU, Gorenek B, et al. Genetic variants associated with atrial
716 fibrillation and long-Term recurrence after catheter ablation for atrial fibrillation in Turkish patients.
717 *Anatolian Journal of Cardiology* **25** (2021) 129–138. doi:10.14744/AnatolJCardiol.2020.44082.
- 718 11 .Verweij N, Eppinga RN, Hagemeyer Y, van der Harst P. Identification of 15 novel risk loci for coronary
719 artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. *Scientific Reports*
720 **7** (2017) 2761. doi:10.1038/s41598-017-03062-8.
- 721 12 .Tragante V, Doevendans PA, Nathoe HM, van der Graaf Y, Spiering W, Algra A, et al. The impact of
722 susceptibility loci for coronary artery disease on other vascular domains and recurrence risk. *European*
723 *Heart Journal* **34** (2013) 2896–2904. doi:10.1093/eurheartj/ehs222.
- 724 13 .Wauters E, Carruthers KF, Buyschaert I, Dunbar DR, Peuteman G, Belmans A, et al. Influence of
725 23 coronary artery disease variants on recurrent myocardial infarction or cardiac death: the GRACE
726 Genetics Study. *European Heart Journal* **34** (2013) 993–1001. doi:10.1093/eurheartj/ehs389.
- 727 14 .Åsvold BO, Langhammer A, Rehn TA, Kjelvik G, Grøntvedt TV, Sørgerd EP, et al. Cohort Profile
728 Update: The HUNT Study, Norway. *International Journal of Epidemiology* (2022). doi:10.1093/ije/
729 dyac095.

- 730 **15** .Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access
731 Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age.
732 *PLoS Medicine* **12** (2015). doi:10.1371/journal.pmed.1001779.
- 733 **16** .Brumpton BM, Graham S, Surakka I, Skogholt AH, Løset M, Fritsche LG, et al. The HUNT Study: a
734 population-based cohort for genetic research. *medRxiv* (2021). doi:10.1101/2021.12.23.21268305.
- 735 **17** .Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with
736 deep phenotyping and genomic data. *Nature* **562** (2018) 203–209. doi:10.1038/s41586-018-0579-z.
- 737 **18** .Zhao Lq, Zhang Gb, Wen Zj, Huang Ck, Wu Hq, Xu J, et al. Common variants predict recurrence after
738 nonfamilial atrial fibrillation ablation in Chinese Han population. *International Journal of Cardiology*
739 **227** (2017) 360–366. doi:10.1016/j.ijcard.2016.11.057.
- 740 **19** .McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of
741 64,976 haplotypes for genotype imputation. *Nature Genetics* **48** (2016) 1279–1283. doi:10.1038/ng.
742 3643.
- 743 **20** .Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling
744 for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature*
745 *Genetics* **50** (2018) 1335–1341. doi:10.1038/s41588-018-0184-y.
- 746 **21** .Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association
747 scans. *Bioinformatics* **26** (2010) 2190–2191. doi:10.1093/bioinformatics/btq340.
- 748 **22** .Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic
749 associations with FUMA. *Nature Communications* **8** (2017). doi:10.1038/s41467-017-01261-5.
- 750 **23** .Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog:
751 knowledgebase and deposition resource. *Nucleic Acids Research* (2022). doi:10.1093/nar/gkac1010.
- 752 **24** .[Dataset] PAN-UKB team. <https://pan.ukbb.broadinstitute.org> (2020).
- 753 **25** .Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the
754 PANTHER classification system. *Nature Protocols* **8** (2013) 1551–1566. doi:10.1038/nprot.2013.092.
- 755 **26** .Stuart JM, Segal E, Koller D, Kim SK. A Gene-Coexpression Network for Global Discovery of
756 Conserved Genetic Modules. *Science* **302** (2003) 249–255. doi:10.1126/science.1087447.
- 757 **27** .Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, et al. Analysis of oncogenic signaling
758 networks in glioblastoma identifies β -catenin as a molecular target. *Proceedings of the National*
759 *Academy of Sciences* **103** (2006) 17402–17407. doi:10.1073/pnas.0608396103.
- 760 **28** .Nowick K, Gernat T, Almaas E, Stubbs L. Differences in human and chimpanzee gene expression
761 patterns define an evolving network of transcription factors in brain. *Proceedings of the National*
762 *Academy of Sciences of the United States of America* **106** (2009) 22358–63. doi:10.1073/pnas.
763 0911376106.
- 764 **29** .van Dam S, Vösa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for
765 functional classification and gene-disease predictions. *Briefings in Bioinformatics* **19** (2018) 575–592.
766 doi:10.1093/bib/bbw139.
- 767 **30** .Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, Graubert A, et al. The GTEx Consortium atlas of
768 genetic regulatory effects across human tissues. *Science* **369** (2020) 1318–1330. doi:10.1126/science.
769 aaz1776.
- 770 **31** .Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC*
771 *Bioinformatics* **9** (2008). doi:10.1186/1471-2105-9-559.
- 772 **32** .Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal Complex*
773 *Systems* (2006) 1695.

- 774 **33** .Csárdi G, Nepusz T, Traag V, Horvát S, Zanini F, Noom D, et al. {igraph}: Network Analysis and
775 Visualization in R. Tech. rep. (2023). doi:10.5281/zenodo.7682609.
- 776 **34** .Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software
777 environment for integrated models of biomolecular interaction networks. *Genome research* **13** (2003)
778 2498–2504.
- 779 **35** .Husser D, Büttner P, Stübner D, Ueberham L, Platonov PG, Dinov B, et al. PR interval associated
780 genes, atrial remodeling and rhythm outcome of catheter ablation of atrial fibrillation-A gene-based
781 analysis of GWAS data. *Frontiers in Genetics* **8** (2017). doi:10.3389/fgene.2017.00224.
- 782 **36** .Hong M, Hwang I, Yu HT, Kim TH, Uhm JS, Joung B, et al. Potential causal association
783 of a prolonged PR interval and clinical recurrence of atrial fibrillation after catheter ablation: a
784 Mendelian randomization analysis. *Journal of Human Genetics* **65** (2020) 813–821. doi:10.1038/
785 s10038-020-0774-2.
- 786 **37** .Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational
787 constraint map using variation in 76,156 human genomes. *Nature* **625** (2024) 92–100. doi:10.1038/
788 s41586-023-06045-0.