UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Inframe insertion and splice site variants in MFGE8 associate with protection against coronary atherosclerosis

Estonian Biobank Research Team; BioBank Japan Project; FinnGen; Ruotsalainen, Sanni E.; Surakka, Ida ...

2022-08-17
Springer
http://hdl.handle.net/10138/347614
Estonian Biobank Research Team, BioBank Japan Project, FinnGen, Ruotsalainen, S E, Surakka, I, Mars, N, Karjalainen, J, Kurki, M, Kanai, M, Krebs, K, Graham, S, Mishra, P P, Mishra, B H, Sinisalo, J, Palta, P, Lehtimäki, T, Raitakari, O, Milani, L, Okada, Y, Palotie, A, Widen, E, Daly, M J, Ripatti, S, Tienari, P & Kaprio, J 2022, 'Inframe insertion and splice site variants in MFGE8 associate with protection against coronary atherosclerosis', Communications Biology, vol. 5, no. 1, 802, pp. 802. https://doi.org/10.1038/s42003-022-03552-0

Downloaded from Helda, University of Helsinki institutional repository. https://helda.helsinki.fi
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

communications biology

ARTICLE

https://doi.org/10.1038/s42003-022-03552-0 OPEN

Inframe insertion and splice site variants in *MFGE8* associate with protection against coronary atherosclerosis

Sanni E. Ruotsalainen[1], Ida Surakka[2], Nina Mars[1], Juha Karjalainen[3], Mitja Kurki[3], Masahiro Kanai[3,4,5], Kristi Krebs[6], Sarah Graham[2], Pashupati P. Mishra[7,8,9], Binisha H. Mishra[7,8,9], Juha Sinisalo[10], Priit Palta[1,6], Terho Lehtimäki[7,8,9], Olli Raitakari[11,12,13], Estonian Biobank Research Team*, Lili Milani[6], The Biobank Japan Project*, Yukinori Okada[5,14], FinnGen*, Aarno Palotie[1,3], Elisabeth Widen[1], Mark J. Daly[1,3,4] & Samuli Ripatti[1,3,15]✉

Cardiovascular diseases are the leading cause of premature death and disability worldwide, with both genetic and environmental determinants. While genome-wide association studies have identified multiple genetic loci associated with cardiovascular diseases, exact genes driving these associations remain mostly uncovered. Due to Finland's population history, many deleterious and high-impact variants are enriched in the Finnish population giving a possibility to find genetic associations for protein-truncating variants that likely tie the association to a gene and that would not be detected elsewhere. In a large Finnish biobank study FinnGen, we identified an association between an inframe insertion rs534125149 in *MFGE8* (encoding lactadherin) and protection against coronary atherosclerosis. This variant is highly enriched in Finland, and the protective association was replicated in meta-analysis of BioBank Japan and Estonian biobank. Additionally, we identified a protective association between splice acceptor variant rs201988637 in *MFGE8* and coronary atherosclerosis, independent of the rs534125149, with no significant risk-increasing associations. This variant was also associated with lower pulse pressure, pointing towards a function of *MFGE8* in arterial aging also in humans in addition to previous evidence in mice. In conclusion, our results suggest that inhibiting the production of lactadherin could lower the risk for coronary heart disease substantially.

[1] Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. [2] Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA. [3] The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4] Analytic and Translational Genetics Unit, Masfsachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [5] Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. [6] Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. [7] Department of Clinical Chemistry, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [8] Finnish Cardiovascular Research Centre, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [9] Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. [10] Heart and Lung Center, Helsinki University Hospital and Helsinki University, Helsinki, Finland. [11] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [12] Centre for Population Health Research, University of Turku, Turku University Hospital, Turku, Finland. [13] Department of Clinical Physiology and Nuclear Medicine, University of Turku, Turku, Finland. [14] Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [15] Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland. *Lists of authors and their affiliations appear at the end of the paper. ✉email: samuli.ripatti@helsinki.fi

Cardiovascular disease (CVD) is the leading cause of premature death and disability worldwide, with both genetic and environmental determinants[1,2]. The most common cardiovascular disease is coronary heart disease (CHD), including coronary atherosclerosis and myocardial infarction, among others. While genome-wide association studies (GWAS) have identified multiple genetic loci associated with cardiovascular diseases, exact genes driving these associations remain mostly uncovered[3].

Owing to Finland's population history, many deleterious and high-impact variants are enriched in the Finnish population giving a possibility to find genetic associations that would not be detected elsewhere[4]. Many studies have reported high-impact loss-of-function (LoF) variants associated with risk factors for CVD, such as blood lipid levels, thus impacting on the CVD risk remarkably. For example, high-impact LoF variants in genes *LPA*[4], *PCSK9*[5], *APOC3*[6], and *ANGPTL4*[7] have been shown to be associated with Lipoprotein(a), LDL-cholesterol (LDL-C), or triglyceride levels, and lowering the CVD risk.

Besides blood lipids, other risk factors for CVD include hypertension, smoking and the metabolic syndrome cluster components. The mechanism that links these risk factors to atherogenesis, however, remains incompletely elucidated. Many, if not all, of these risk factors, however, also participate in the activation of inflammatory pathways, and inflammation in turn can alter the function of artery wall cells in a manner that drives atherosclerosis[8].

Using data from a sizeable Finnish biobank study FinnGen ($n = 260{,}405$), we identified an association with an inframe insertion rs534125149 in *MFGE8* and protection against coronary atherosclerosis and other representations of major coronary heart disease (CHD), such as myocardial infarction (MI). This variant is highly enriched in Finland, 70-fold compared to Non-Finnish Europeans (NFE) in the gnomAD genome reference database[9] with AF of 3% in Finland. This association was also replicated in BioBank Japan (BBJ) and Estonian Biobank (EstBB). We also identified a splice acceptor variant rs201988637 in the same gene, which is also associated with protection against coronary atherosclerosis and other representations of major CHD, indicating that rs534125149 has very similar effect on CHD as a splice acceptor variant in *MFGE8*. Associations of both of these two variants in *MFGE8* were specific to CHD, and they did not significantly ($p < 1.75 \times 10^{-5}$) increase risk for any other disease, highlighting *MFGE8* as a potential drug target candidate.

Results

GWAS results for coronary atherosclerosis. We identified a total of 2 302 variants associated (GWS, $p < 5 \times 10^{-8}$) with coronary atherosclerosis (detailed description of the definition of the endpoint is in Supplementary Note 1). These variants were located in 38 distinct genetic loci (a minimum of 0.5 Mb distance from each other; Fig. 1 and Supplementary Table 1). Out of the 38 GWS loci, four (within or near genes *MFGE8, TMEM200A, PRG3*, and *FHL1*) have not been previously reported to associate with any CVD-related endpoints or risk factor for CVD in GWAS Catalog[10] [https://www.ebi.ac.uk/gwas/]. Lead variants in these loci and their characteristics are listed in Table 1 and locus zoom plots for each of the loci are in Supplementary Fig. 1.

Among these four previously unreported loci for coronary atherosclerosis, the locus near *MFGE8* had the strongest association (p-value $= 2.63 \times 10^{-16}$ for top variant rs534125149). The lead variant is an inframe insertion located in the sixth exon in the *MFGE8* gene (Supplementary Fig. 2) and it is highly enriched in the Finnish population compared to NFSEEs (Non-Finnish, Estonian or Swedish Europeans). Interestingly, *MFGE8* is mainly expressed in coronary and tibial arteries according to data from GTEx v8 (Supplementary Fig. 3), and furthermore the expression of *MFGE8* is highest in aorta. In addition, previously identified common variants in *MFGE8* locus that have been associated with decreased expression of *MFGE8* in tibial artery and aorta have also been associated with decreased risk of CHD[11].

In addition to *MFGE8*, we identified three additional previously unreported loci to be associated with coronary atherosclerosis, *TMEM200A*, *PRG3* and *FHL1* being the nearest genes of the lead variants. *TMEM200A* and *PRG3* loci had one non-coding low-frequency variant reaching the genome-wide significance threshold, and *FHL1* had 11. All variants in the credible sets of all these associations were either intergenic or intronic variants and had no reported significant GWAS associations with any trait in the GWAS Catalog or significant eQTL associations in GTEx. The one variant (rs118042209) in the credible set of *TMEM200A* locus was associated with multiple disease endpoints representing major coronary heart disease (CHD) in FinnGen, including coronary atherosclerosis, ischemic heart disease and angina pectoris, whereas the lead variant in the *PRG3* locus was associated with cardiomyopathy. All variants in the credible set of *FHL1* were associated with multiple disease endpoints representing major CHD in FinnGen, including angina pectoris and ischemic heart disease. *TMEM200A* have been reported to be associated with ten traits (including height and trauma exposure) and *PRG3* with two traits (eosinophil count and eosinophil percentage of white cells) in the GWAS Catalog. *FHL1* gene had no reported associations in GWAS Catalog.

Replication. Association between rs534125149 in *MFGE8* locus with CHD was replicated in Biobank Japan[12,13] (BBJ) and the Estonian Biobank (EstBB)[14] (35,644 cases and 328 461 controls total: OR = 0.752 [0.67–0.84], $p = 4.37 \times 10^{-7}$). Association results for rs534125149 with CHD and MI across different cohorts are in Fig. 2. Post hoc power calculations for each cohort were performed (probability that the test will reject the null hypothesis H0 at GWS threshold) and the results as the function of effect size are in Supplementary Fig. 4. From these calculations we can see that in FinnGen the power to detect the variant as GWS is remarkably greater than in EstBB or BBJ, even with similar effect sizes and sample sizes. Therefore, this boost in power in FinnGen seems to be mainly due to a different allele frequencies, since this variant is highly enriched to Finland.

In addition to *MFGE8*, meta-analysis across FinnGen, UKBB, EstBB, and BBJ was performed for the lead variants in the three other previously unreported loci for CHD (*TMEM200A, PRG3*, and *FHL1*), where available. Lead variant in *PRG3* locus is highly enriched to Finland and absent in all other cohorts, and thus replication efforts for that variant were not possible. The two other loci that were meta-analyzed (*TMEM200A* and *FHL1*) did not replicate (p-value in the combined meta-analysis of the replication cohorts (meta-analysis without FinnGen) is smaller than $0.05/4 = 0.0125$ and all effect size estimates are in same direction). Association results for rs534125149 with CHD and MI across different cohorts for *TMEM200A* and *FHL1* variants are in Fig. 3. Post hoc power calculations for each cohort were performed and the results as the function of effect size are in Supplementary Fig. 5. From those results we can see that the lack of replication in UKBB, EstBB and BBJ does not appear to be due to lack of power. Therefore, we identified and replicated one novel locus for CHD (*MFGE8*).

Phenome-wide association results for rs534125149. We observed a highly protective association for the Finnish enriched inframe insertion rs534125149 in the *MFGE8* gene and multiple disease endpoints, all representing major CHD, including coronary atherosclerosis (OR = 0.75 [0.71–0.81], $p = 2.63 \times 10^{-16}$) and

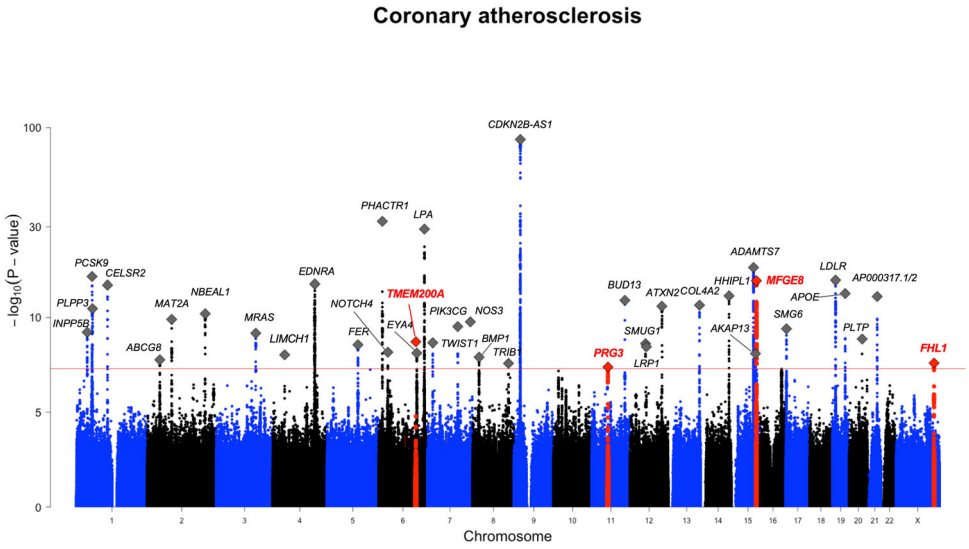


Fig. 1 GWAS results for coronary atherosclerosis in FinnGen. Total number of independent genome-wide significant associations (GWS; $p < 5 \times 10^{-8}$) is 38, the lead variant in each marked with diamonds. Four previously unreported associations for CVD-related phenotypes are highlighted with ±750 Mb around the lead variant in the region as red and the lead variant marked with red diamond.

myocardial infarction (MI) (OR = 0.74 [0.68–0.81], $p = 1.95 \times 10^{-11}$). In total, this variant was associated (PWS) with 14 disease endpoints, all representing major CHD (Fig. 4). Majority of them are highly overlapping, and thus similar associations to all of them is expected. Thus, we pruned the 14 PWS disease endpoints down to six disease endpoints (coronary atherosclerosis, coronary revascularization, ischemic heart diseases, major coronary heart disease event, myocardial infarction, and statin medication) that have fundamental characteristics for further analyses. For the inframe insertion rs534125149 in *MFGE8*, we did not detect other phenome-wide significant associations among the 2 861 endpoints in our data.

Splice acceptor variant rs201988637 in *MFGE8*. In addition to inframe insertion rs53412514, we identified a splice acceptor variant (rs201988637) in MFGE8 to be associated with coronary atherosclerosis (OR = 0.72 [0.63–0.83], $p = 7.94 \times 10^{-06}$) and multiple disease endpoints representing major CHD. The splice acceptor variant had very similar PheWAS profile as the inframe insertion (Supplementary Fig. 6) and furthermore the two variants had very similar protective effect sizes for the endpoints (Fig. 5 and Supplementary Table 2). Similar to rs534125149, this variant is also highly enriched in Finland (37-fold compared to NFE), allele frequency in Finland being 0.6%. Moreover, both the splice acceptor and the inframe insertion variants were enriched to Eastern Finland (Supplementary Fig. 7).

These two variants (rs534125149 and rs201988637) are in low linkage disequilibrium (LD, $r^2 = 0.00015$) and did not have any effect on the other variant's associations with coronary atherosclerosis or MI (Table 2 and Supplementary Fig. 8). This indicates that they both are independently associated with these endpoints.

Survival analysis. In addition to protection against coronary atherosclerosis and myocardial infarction, both the infame insertion rs534125149 and splice acceptor variant rs201988637 showed also significant association in survival analysis when analyzing survival time from birth to first diagnose of coronary atherosclerosis (HR = 0.78 [0.74–0.93]), $p = 1.67 \times 10^{-17}$ and HR = 0.77 [0.69–0.88], $p = 5.08 \times 10^{-05}$, respectively) and myocardial infarction (HR = 0.86 [0.80–0.93], $p = 2.63 \times 10^{-10}$ and HR = 0.72 [0.61–0.85], $p = 8.16 \times 10^{-05}$). In addition, when combining the heterozygous and homozygous carriers of both rs534125149 and rs201988637 together, carriers get the first diagnose significantly later than non-carriers (HR = 0.81 [0.77–0.85], $p = 6.4 \times 10^{-16}$ for coronary atherosclerosis and HR = 0.78 [0.72–0.85], $p = 1.16 \times 10^{-11}$ for MI) (Fig. 6).

In addition, as a sensitivity analysis we performed the similar Cox model for first event of MI by adding different risk factors for CHD as covariates in the model to see if any of these risk factors (BMI, Type 2 Diabetes, smoking, statin use or sex) have impact on the observed association. Risk factors were added to the model both individually and together. As a result, we saw only a small change in the effect size when adjusting for these risk factors (Supplementary Table 3). The change was more noticeable on *p*-values where the missing data in the added covariates lead to decreased statistical power.

Associations with risk factors for CVD. We then tested for possible associations between the *MFGE8* variants and risk factors for CVD. The splice acceptor variant rs201988637 was associated with pulse pressure in analyses across four cohorts with pulse pressure measurement and variant rs201988637 available, with the risk lowering allele associated with lower pulse pressure ($p = 1.7 \times 10^{-04}$, $\beta = -0.13$ [−0.2 to −0.06]) (Fig. 7). Association with pulse pressure was also tested for inframe insertion rs534125149 and previously reported common variant in the locus, rs8042271 across all where the variants were available. We saw consistent effect sizes across the cohorts, and significant ($p < 0.05$) meta-analysis *p*-values for both variants (Supplementary Fig. 9).

In addition, in recent studies for blood pressure measurements (systolic and diastolic blood pressure and pulse pressure), genome-wide significant association have been reported in the region[15,16]. To assess whether these reflects the same signal, we performed colocalization analysis in the region ±200 kB around rs53412514 using Coloc package in R[17] with coronary atherosclerosis results from FinnGen and pulse pressure GWAS results from Evangelou et al.[16] The probability for shared signal (PP4) was 97.1%, further validating *MFGE8* locus is associated with pulse pressure (Supplementary Fig. 10).

In addition to pulse pressure associations in the region, rs534125149 was significantly associated with height, but further

Table 1 Lead variants in previously unreported loci for coronary atherosclerosis.

Lead variant chrom:pos:ref_alt (rsid)	Most severe consequence	Nearest gene	FIN enrichment (NFE)	AF	OR (p-value)	Info	# cs (Post-pr)	#coding in cs(s)
chr15:88901702:C_CTGT (rs534125149)	Inframe insertion	*MFGE8*	70.59	0.029	0.75 (2.60×10^{-16})	0.99	2 (0.705)	1
chr6:130483492:A_G (rs118042209)	Intergenic variant	*TMEM200A*	0.87	0.010	0.7 (1.90×10^{-9})	0.91	1 (0.904)	0
chr11:57380633:A:G (rs764568652)	Intron variant	*PRG3*	[a]	0.0003	7.72 (4.12×10^{-8})	0.89	1 (0.583)	0
chrX:136194941:C_G (rs5974585)	Intron variant	*FHL1*	1.25	0.49	0.95 (2.55×10^{-8})	0.99	1 (0.692)	0

[a]Variant not present in NFE in gnomAD.

analysis pointed this signal to be reflecting the association of a known association of *ACAN* with height, located near *MFGE8* (Supplementary Fig. 11). No associations with other risk factors were observed.

In the Corogene cohort ($n = 4896$), rs534125149 was significantly ($p < 0.05$) associated with lower risk for acute coronary syndrome and stable coronary heart disease (RR = 0.87 and 0.83, respectively) compared to healthy controls, but not with myocardial infarction without coronary artery occlusion (Supplementary Fig. 12). These results are in line with our findings regarding the specificity of the association of variants in *MFGE8* on atherosclerotic cardiovascular disorders. The *p*-value for the difference of the AFs of rs534125149 among patients with acute coronary syndrome or stable coronary heart disease and among MINOCA was, however, not significant ($p = 0.78$), which may due to lack of power. In addition, the cohort is very heterogeneous.

Previously reported common variants near *MFGE8*. Previously, common intergenic variant (rs8042271) near *MFGE8* has been reported to associate with coronary heart disease (CHD) risk[3,18]. We replicate this association (OR = 0.90, $p = 3.69 \times 10^{-10}$ for coronary atherosclerosis) in FinnGen. LD between the common variant rs8042271 and the inframe insertion rs534125149 is 0.154. The LD characteristics for all three variants in *MFGE8* (rs534125149, rs201988637 and rs8042271) in FinnGen are in Supplementary Table 4. Common variant rs8042271 was in the 95% credible set for MI with the causal probability of 0.003 but was not included in the 95% credible sets for coronary atherosclerosis (Supplementary Tables 5 and 6). The conditional analyses of all three *MFGE8* variants showed that the association of the previously reported common variant rs8042271 can be explained by the inframe insertion variant rs534125149, but not vice versa, and that the association of the splice acceptor variant rs201988637 is independent of both rs534125149 and rs8042271. (Supplementary Table 7). This was the case also with previously reported common variant rs734780, showing very similar LD with rs534125149 (0.112) as rs8042271 (0.154).

Fine-mapping of the *MFGE8* locus. In our fine-mapping analyses, MI had most probably one credible set (set of causal variants) of 32 variants with the highest posterior probability (posterior probability = 0.62), and coronary atherosclerosis had two credible sets of 6 and 45 variants, respectively, with the highest posterior probability (posterior probability = 0.74). For both MI and coronary atherosclerosis, rs534125149 had the highest probability of being causal (probability of being causal = 0.250 and 0.318, respectively) and was included in the first credible set (Supplementary Tables 5 and 6; and Supplementary Fig. 13). Splice acceptor variant rs201988637 was not included in the credible sets for either MI or coronary atherosclerosis, whereas previously reported common variant rs8042271 was included in the credible set for MI with the probability of being causal = 0.003 (Supplementary Table 6).

Protein modeling. We predicted the impact of the insertion variant rs534125149 on the protein structure of *MFGE8* using AlphaFold[19]. The predicted conformational changes were localized to a loop region within the C2 domain, ~20 Å away from the key amino acids involved in membrane binding (Supplementary Fig. 14)[20,21]. This loop contains Asn238, which is known to be glycosylated[22]. It is possible that the insertion of an additional asparagine may lead to impaired glycosylation, which is important for protein folding, among other cellular processes[23]. The role of this region in the function of *MFGE8* hasn't been

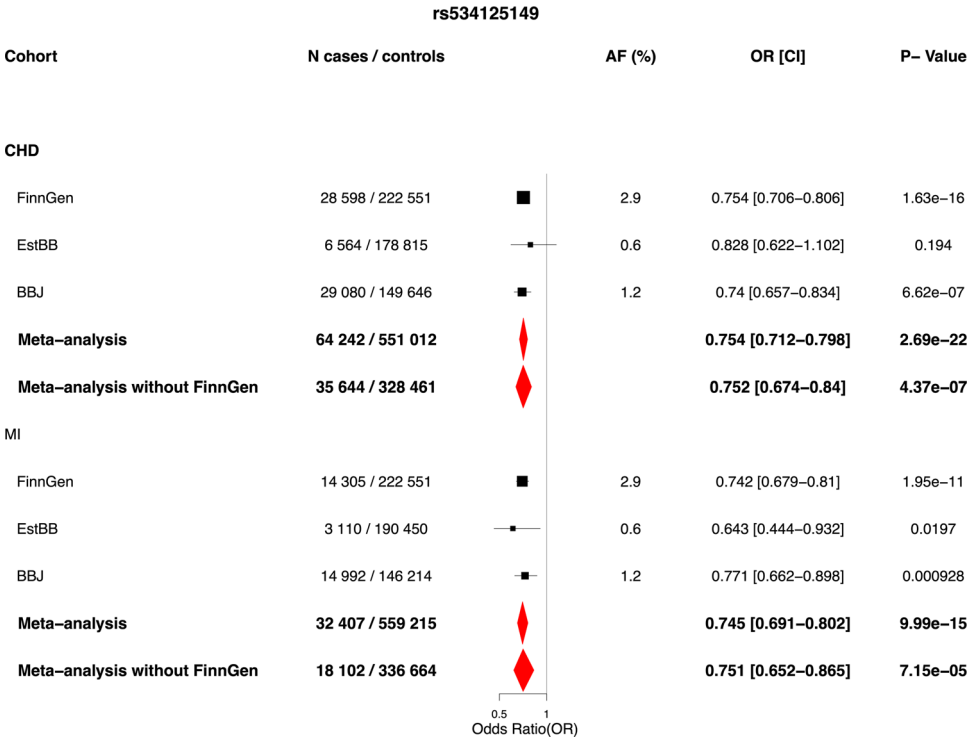


Fig. 2 Results for rs534125149 against coronary heart disease and myocardial infarction across cohorts where available and meta-analysis results. Logistic regression has been applied, adjusted for age and sex. Meta-analysis was performed using inverse-variance weighted fixed-effects meta-analysis method. Black dots represents odds ratios, and lines 95% confidence interval from the the single cohorts and red diamonds represent the results from meta-analysis ends of the diamonds representing the ends of the 95% confidence interval. Source data for the figure is in Supplementary Data 1.

previously described and it is therefore unclear how this variant would otherwise lead to an impact on *MFGE8* function. Thus, further experimental work is necessary to understand the mechanism by which this variant leads to protection against coronary atherosclerosis.

Discussion

Here, we show that a Finnish enriched inframe insertion in *MFGE8* is associated with substantially lower risk of diseases representing major CHD, including myocardial infarction and coronary atherosclerosis. This variant was associated with CHD specifically, and no significant association was observed to other diseases in a phenome-wide search, even if this can be due to lower statistical power in rare disease endpoints. Splice acceptor variant rs201988637 in *MFGE8* was also associated with lower pulse pressure, but not with blood lipids, blood pressure or other known coronary heart disease risk factors.

Our findings allow us to draw several conclusions. First, *MFGE8* is a potential intervention target with specific effects on coronary heart disease. Specific protective association with the variants in *MFGE8* and CHD shows potential for efficacy of a treatment targeting MFGE8 protein or downstream products. Second, the lack of risk elevation in other diseases provide evidence on the potential safety of the intervention. Previously, the protective effect of loss-of-function variants have been reported for example for *PCSK9*[5] and *APOC3*[6], and in phase I, II and III trials, inhibition of PCSK9 have led to significantly decreased LDL-C levels, and in short-term trials, PCSK9 inhibitors have been well-tolerated and have had a low incidence of adverse effects[24] Based on the phenome-wide association profile for the splice acceptor variant rs201988637, we hypothesize that inhibiting *MFGE8* could lower the CHD risk, if the variant can be proved to be loss-of-function in *MFGE8*.

An association of a splice acceptor variant rs201988637 in *MFGE8* with lower pulse pressure, a potential biomarker for arterial stiffness[25], are very much in line with previous studies on MFGE8 and the inflammatory aging process of the arteries, highlighting the possible role of *MFGE8* in arterial aging and stiffness. The *MFGE8* gene encodes Milk-fat globule-EGF 8 (MFGE8), or lactadherin, which is an integrin-binding glycoprotein implicated in vascular smooth muscle cell (VSMC) proliferation and invasion, and the secretion of pro-inflammatory molecules[26,27]. Lactadherin is known to play important roles in several other biological processes, including apoptotic cell clearance and adaptive immunity[28], which are known to contribute to the pathogenesis of ischemic stroke. Initially lactadherin was identified as a bridging molecule between apoptotic cells and phagocytic macrophages[29–31], but growing evidence has indicated that it is a secreted inflammatory mediator that orchestrates diverse cellular interactions involved in the pathogenesis of various diseases, including vascular metabolic disorders and some tumors[32–36] and cancers, such as breast[34,37], bladder[38], esophageal[39] and colorectal cancer[40]. Recently, not only has MFG-E8 expression emerged as a molecular hallmark of adverse cardiovascular remodeling with age[41–44], but MFG-E8 signaling has also been found to mediate the vascular outcomes of cellular and matrix responses to the hostile stresses associated with hypertension, diabetes, and atherosclerosis[45–49].

Arterial inflammation and remodeling are linked to the pathogenesis of age-associated arterial diseases, such as atherosclerosis. Recently, lactadherin has been identified as a novel local biomarker for aging arterial walls by high-throughput proteomic screening, and it has been shown to also be an element of the arterial inflammatory signaling network[50]. The transcription, translation, and signaling levels of MFG-E8 are increased in aged, atherosclerotic, hypertensive, and diabetic arterial walls in vivo, as well as activated VSMCs and a subset of macrophages in vitro. During aging, both MFG-E8 transcription and translation

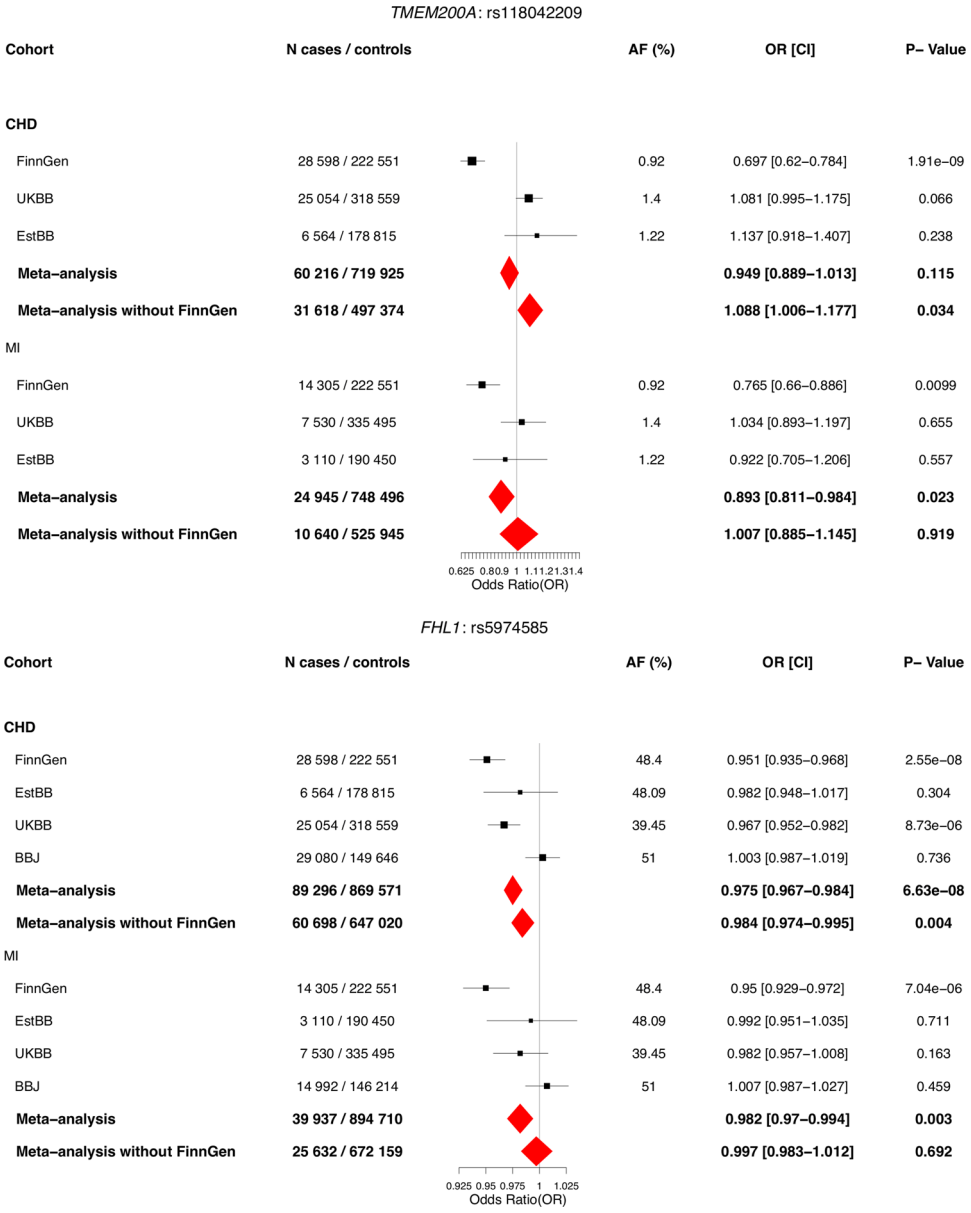


Fig. 3 Results for rs118042209 in *TMEM200A* and rs5974585 in *FHL1* against coronary heart disease and myocardial infarction across different cohorts across cohorts where available. Logistic regression has been applied, adjusted for age and sex. Meta-analysis was performed using inverse-variance weighted fixed-effects meta-analysis method. Black dots represent odds ratios, and lines 95% confidence interval from the single cohorts and red diamonds represent the results from meta-analysis ends of the diamonds representing the ends of the 95% confidence interval. Source data for the figure is in Supplementary Data 1.

increase within the arterial walls and hearts of various species, including rats, humans, and monkeys[44,51–53], and MFG-E8 is markedly up-regulated in rat aortic walls with aging[44]. High levels of MFG-E8 have also been detected within endothelial cells, SMC, and macrophages of atherosclerotic aortae in both mice and humans[49,54]. Furthermore, in the advanced atherosclerotic plaques found in murine models, decreased macrophage MFG-E8 levels are associated with an inhibition of apoptotic cell engulfment, leading to the accumulation of cellular debris during the pathogenesis of atherosclerosis. Lactadherin has, however, in contrast shown tissue protection in various models of organ injury, including suppression of inflammation and apoptosis in intestinal ischemia in mice[55], as well as inducing recovery from ischemia by facilitating angiogenesis[56].

In addition, expression of *MFGE8* is highly enriched to tissues relevant to the reported association, such as aorta. Genes nearby *MFGE8*, including *ABHD2* and *HAPLN3*, are, however similarly to *MFGE8* enriched to arteries[18]. Therefore, they could play a role in atherosclerosis via coordinated gene network. In addition, recent studies have pointed toward the fact that lncRNA, called CARMAL, may regulate the expression of *MFGE8*[57].

Our study does, however, have a few limitations. First, our primary association results come from Finnish population with considerable elevation in allele frequency in *MFGE8* variants among Finns. Therefore, the replication of the association in other populations has reduced statistical power. However, there were enough carriers combined in Japanese, Estonian and UK samples to replicate robustly both the protective association with coronary heart disease and for pulse pressure. Secondly, although our data shows association with pulse pressure, which has previously been linked to arterial stiffness, the direct effect of the genetic variants on arterial stiffness and arterial aging needs further evidence.

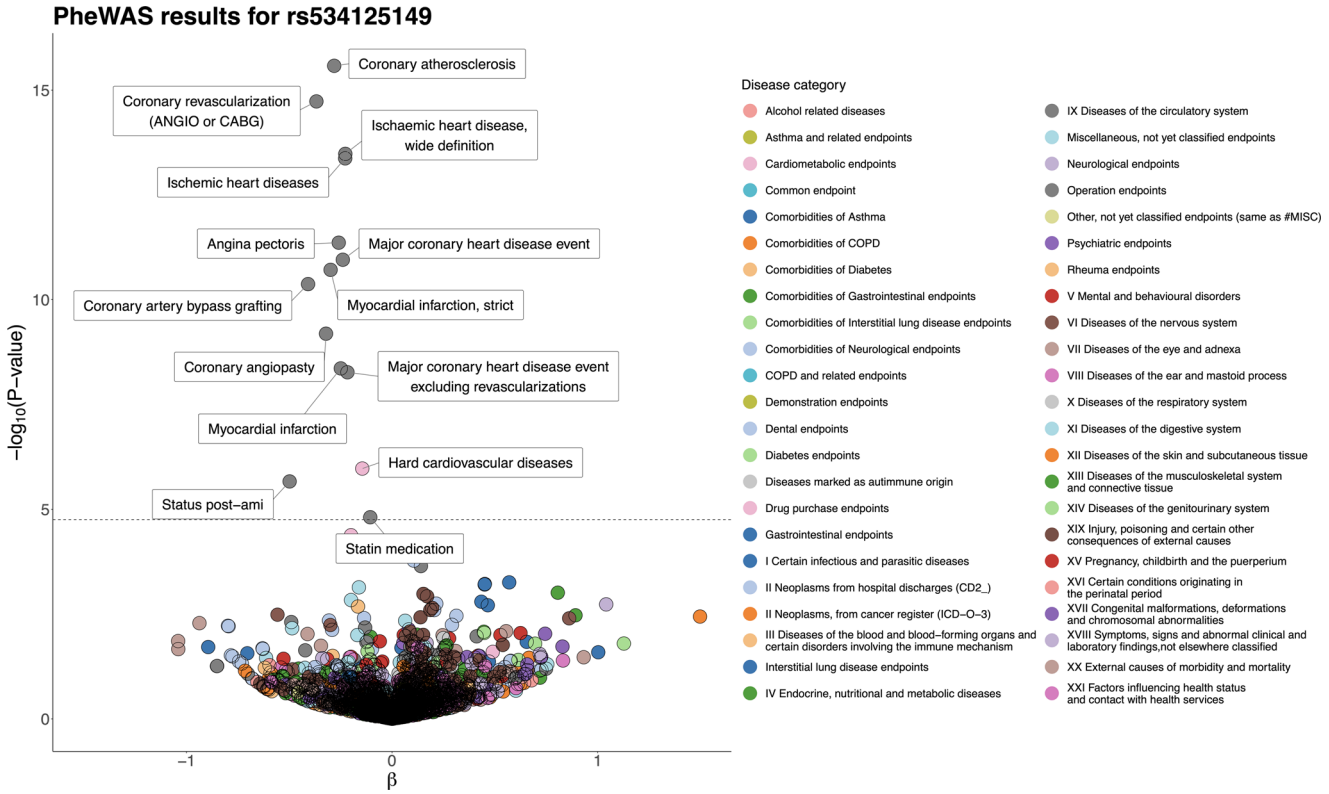


Fig. 4 Phenome-wide association study (PheWAS) results for rs534125149. Total number of tested endpoints is 2861 (A complete list of endpoints analyzed and their definitions is available at https://www.finngen.fi/en/researchers/clinical-endpoints). The dashed line represents the phenome-wide significance threshold, multiple testing corrected by the number of endpoints = $0.05/2861 = 1.75 \times 10^{-5}$. All endpoints reaching that threshold are labeled in the figure.

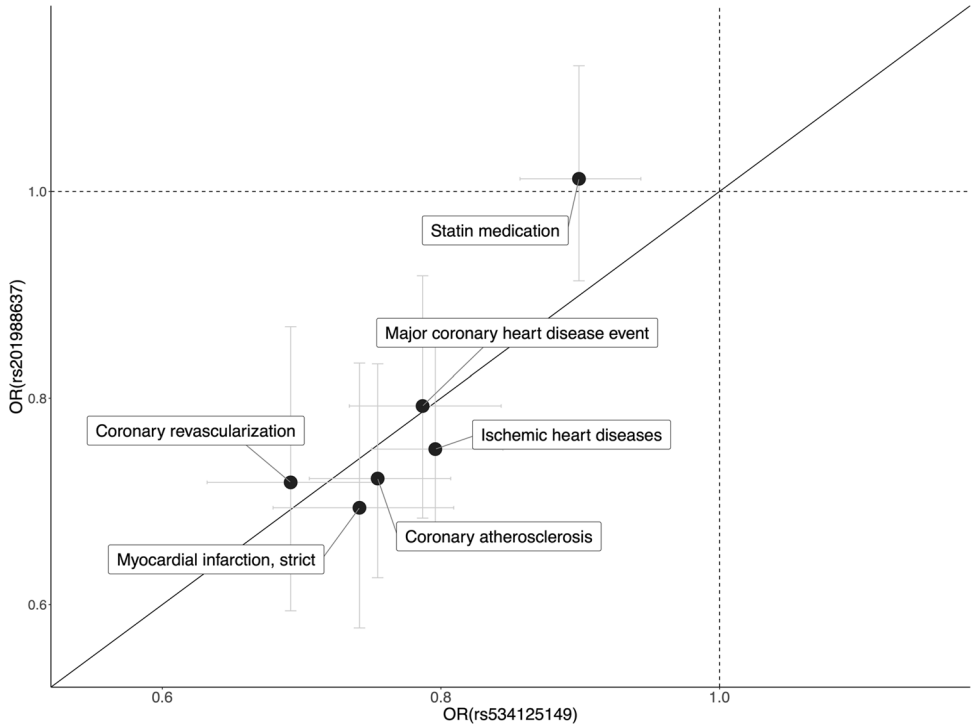


Fig. 5 Effect size comparison. Comparison of the effects (OR) of rs534125149 and rs201988637 for 14 endpoints with p-value $< 1.75 \times 10^{-5}$ (PWS) for rs534125149 in FinnGen R6. 95% confidence intervals represented as gray lines.

Table 2 Results of the conditional analysis on MI and coronary atherosclerosis.

Phenotype	SNPID [chr:position:ref:alt] (rsid)	Most severe consequence	Original GWAS results		Conditional results	
			OR [CI]	*p*-value	OR [CI]	*p*-value
Coronary atherosclerosis	chr15:88901702:C:CTGT (rs534125149)	Inframe insertion	0.75 [0.71–0.81]	2.63×10^{-16}	0.75 [0.70–0.80][a]	7.68×10^{-15a}
	chr15:88899813:T:G (rs201988637)	Splice acceptor variant	0.72 [0.63–0.83]	7.94×10^{-6}	0.73 [0.64–0.85][b]	1.99×10^{-5b}
Myocardial infarction, strict	chr15:88901702:C:CTGT (rs534125149)	Inframe insertion	0.74 [0.68–0.81]	1.95×10^{-11}	0.79 [0.73–0.85][a]	1.92×10^{-10a}
	chr15:88899813:T:G (rs201988637)	Splice acceptor variant	0.69 [0.58–0.83]	9.62×10^{-5}	0.71 [0.59–0.85][b]	4.03×10^{-4b}

This table present the conditional analysis results for coronary atherosclerosis and MI (strict definition, only primary diagnoses accepted) where the association has been conditioned on rs534125149 and rs201988637, separately.
[a]Conditional on rs201988637.
[b]Conditional on rs534125149.

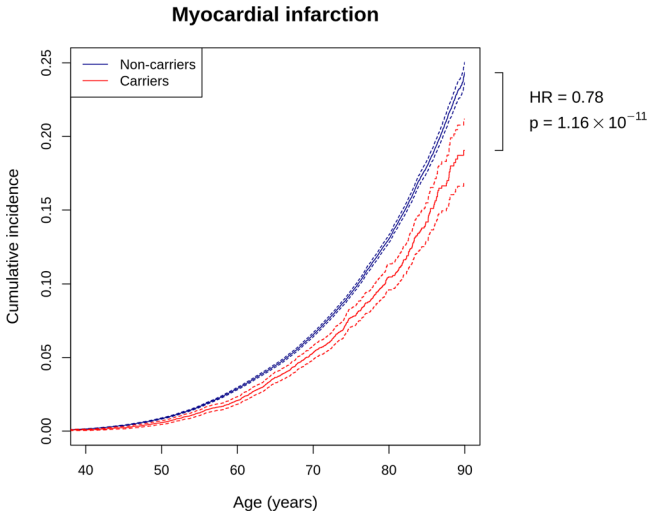


Fig. 6 Cumulative incidence plots for first event of myocardial infarction in FinnGenR6. Red line represents carriers (homo- or heterozygous) for either rs534125149 or rs201988637 ($n = 17,838$), and blue line represent non-carriers ($n = 242,567$). Hazard ratio and *p*-value are from cox-proportional hazards model. Dashed lines represent 95% confidence intervals.

Lastly, with our dataset, we have not been able to demonstrate that the two variants (rs534125149 and rs201988637) in *MFGE8* are loss-of-function variants, and thus further experimental work is required to validate our findings.

In conclusion, our results suggests that inhibiting production of lactadherin could reduce the risk for coronary atherosclerosis substantially and thus present *MFGE8* as a potential therapeutical target for atherosclerotic cardiovascular disease. Our study also highlights the potential of FinnGen, as a large-scale biobank study in isolated population to identify high-impact variants either very rare or absent in other populations.

Methods

Study cohort and data. We studied total of 2 861 disease endpoints in Finnish biobank study FinnGen ($n = 260\,405$) (Table 3). FinnGen (https://www.finngen.fi/en) is a large biobank study that aims to genotype 500,000 Finns and combine this data with longitudinal registry data, including national hospital discharge, death, and medication reimbursement registries, using unique national personal identification numbers. FinnGen includes prospective epidemiological and disease-based cohorts as well as hospital biobank samples.

Definition of disease endpoints. All the 2861 disease-endpoint analyzed in FinnGen have been defined based on registry linkage to national hospital discharge, death, and medication reimbursement registries. Diagnoses are based on International Classification of Diseases (ICD) codes and have been harmonized over ICD codes 8, 9, and 10. More detailed lists of the ICD codes used for the disease-endpoints myocardial infarction and coronary atherosclerosis, which are discussed more in this study, are in Supplementary Note 1. A complete list of endpoints analyzed, and their definitions is available at https://www.finngen.fi/en/researchers/clinical-endpoints.

Genotyping and imputation. FinnGen samples were genotyped with multiple Illumina and Affymetrix arrays (Thermo Fisher Scientific, Santa Clara, CA, USA). Genotype calls were made with GenCall and zCall algorithms for Illumina and AxiomGT1 algorithm for Affymetrix chip genotyping data batchwise. Genotyping data produced with previous chip platforms were lifted over to build version 38 (GRCh38/hg38) following the protocol described here: dx.doi.org/10.17504/protocols.io.nqtddwn. Samples with sex discrepancies, high-genotype missingness (>5%), excess heterozygosity (±4SD) and non-Finnish ancestry were removed. Variants with high missingness (>2%), deviation from Hardy–Weinberg equilibrium ($p < 1 \times 10^{-6}$) and low minor allele count (MAC < 3) were removed.

Pre-phasing of genotyped data was performed with Eagle 2.3.5 (https://data.broadinstitute.org/alkesgroup/Eagle/) with the default parameters, except the number of conditioning haplotypes was set to 20,000. Imputation of the genotypes was carried out by using the population-specific Sequencing Initiative Suomi (SISu) v3 imputation reference panel with Beagle 4.1 (version 08Jun17.d8b,

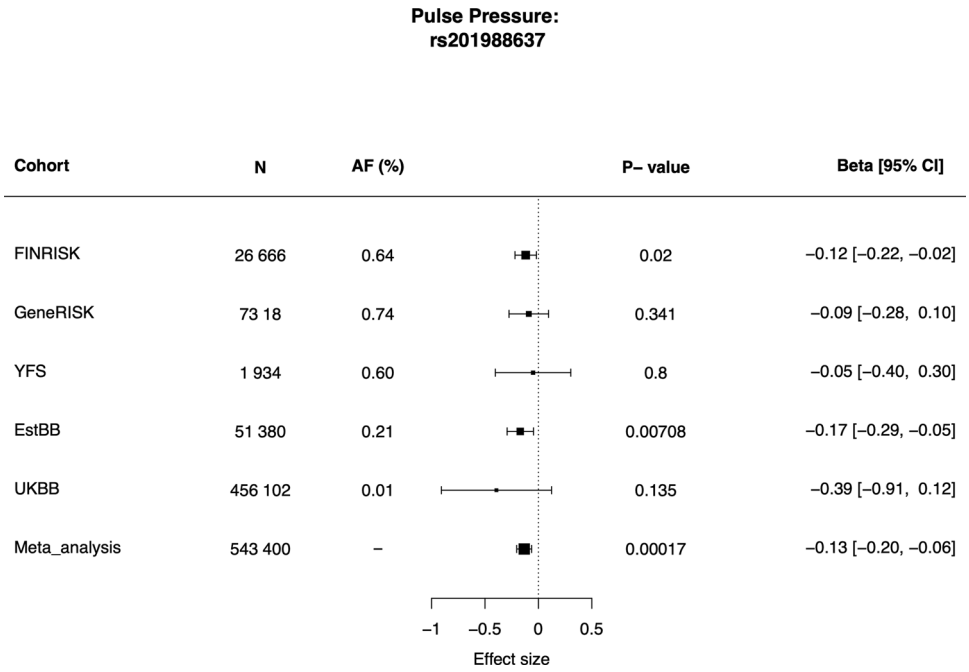


Fig. 7 Results for pulse pressure association across all cohorts with splice acceptor variant rs201988637 available (FINRISK, GeneRISK, YFS, EstBB, and UKBB). Size of the boxes represent the sample size of the cohorts, and the lines the 95% confidence interval. Associations were tested using linear regression, adjusting for age and sex Pulse pressure phenotypes were inverse-rank normalized prior analysis. Source data for the figure is in Supplementary Data 1.

Table 3 Basic characteristics of the study cohort.

	All	Females	Males
N (%)	260,405	147,061 (56.47%)	113,344 (43.53%)
Age (mean (sd))	53.15 (17.55)	51.84 (17.71)	54.85 (17.19)
BMI (mean (sd))[a]	27.29 (5.36)	27.21 (5.83)	27.38 (4.76)
Statin use (*N* (%))	86,466 (33.2%)	40,422 (27.48%)	46,044 (40.62%)
Hypertension (*N* (%))	68,005 (26.11%)	33,420 (22.72%)	34,585 (30.51%)
Smoking (*N* (%))[b]	1733 (1.07%)	901 (0.96%)	832 (1.22%)
Coronary atherosclerosis	28,598 (11.38%)	9252 (6.87%)	19,346 (17.86%)
Myocardial infarction	14,305 (6.04%)	3958 (2.87%)	10,347 (10.42%)

[a]BMI is available only from 178,966 individuals.
[b]Smoking information is available only from 98,654 individuals.

https://faculty.washington.edu/browning/beagle/b4_1.html) as described in the following protocol: dx.doi.org/10.17504/protocols.io.nmndc5e. SISu v3 imputation reference panel was developed using the high-coverage (25–30x) whole-genome sequencing data generated at the Broad Institute of MIT and Harvard and at the McDonnell Genome Institute at Washington University, USA; and jointly processed at the Broad Institute. Variant callset was produced with Genomic Analysis Toolkit (GATK) HaplotypeCaller algorithm by following GATK best practices for variant calling. Genotype-, sample- and variant-wise quality control was applied in an iterative manner by using the Hail framework v0.2. The resulting high-quality WGS data for 3775 individuals were phased with Eagle 2.3.5 as described above. As a post-imputation quality control, variants with INFO score <0.7 were excluded.

Association testing and replication. A total of 260,405 samples from FinnGen Data Freeze 6 with 2861 disease endpoints were analyzed using Scalable and Accurate Implementation of Generalized mixed model (SAIGE), which uses saddlepoint approximation (SPA) to calibrate unbalanced case-control ratios[58]. Models were adjusted for age, sex, genotyping batch and first ten principal components. All variants reaching genome-wide significance p-value threshold of 5×10^{-8} are considered as genome-wide significant (GWS), and all disease-endpoints reaching multiple testing corrected (for the number of endpoints tested = 2861) p-value threshold of $0.05/2861 = 1.75 \times 10^{-5}$ were considered as phenome-wide significant (PWS).

Independent GWS loci for atherosclerosis were determined as adding ±0.5 Mb around each variant that reached the genome-wide significance threshold, overlapping regions were merged. The publicly available summary statistics from CARDIoGRAMplusC4D, a large meta-analysis of CHD involving 60,801 cases and 123,504 controls[3] was used for assessing whether the locus has been previously reported to associate with CHD. In addition, NHGRI-EBI GWAS Catalog[10] was used for assessing whether the locus has been previously reported to associate with any CVD-related endpoint or traditional risk factor for CVD, such as blood lipids, BMI and blood pressure. All loci that had not been reported to associate with CVD were fine-mapped using FINEMAP[59] to determine the credible sets in each signal, and meta-analyzed across the cohorts (UKBB, EstBB and BBJ) where available to test their novelty.

In Corogene[60] ($n = 5300$), a sub-cohort of FinnGen where participants have been collected as patients with coronary heart disease (CHD) and other related heart diseases, we tested the association of rs534125149 with sub-types of coronary heart disease: acute coronary syndrome, stable coronary heart disease (CHD) and MINOCA[61] (myocardial infarction no coronary artery occlusion), by which we refer to patients that have had symptoms, ECG-changes and cardiac enzyme or troponine release suggesting acute coronary syndrome, but did not have coronary stenosis. The acute coronary syndrome was further divided into unstable Angina pectoris, non-ST segment elevation myocardial infarction (NSTEMI) and ST segment elevation myocardial infarction (STEMI). Associations were tested by calculating risk ratios (RR) for carriers vs. non-carriers of rs534125149 using non-CHD group always as controls and excluding the other tested groups from the analysis. p-values were calculated using χ^2-test, and p-values < 0.05 were considered significant.

Survival analysis. Survival analysis for coronary atherosclerosis and myocardial infarction was performed using GATE[62], which accounts for both population

structure and sample relatedness and controls type I error rates even for phenotypes with extremely heavy censoring. GATE transforms the likelihood of a multivariate Gaussian frailty model to a modified Poisson generalized linear mixed model (GLMM[63,64]), and to obtain well-calibrated *p*-values for heavily censored phenotypes, GATE uses the SPA to estimate the null distribution of the score statistic. For coronary atherosclerosis and myocardial infarction, survival time from birth to first diagnose was analyzed for both rs534125149 and rs201988637. Models were adjusted for age, sex, genotyping batch and first ten principal components, similarly to original GWAS analyses. In addition, cox-proportional hazards model was used for survival analysis for coronary atherosclerosis and myocardial infarction using a binary variable (carrier or non-carrier) for either inframe insertion rs534125149 or splice acceptor variant rs201988637.

Biomarker analyses. We tested the association of the two *MFGE8* variants (rs534125149 and rs201988637) with quantitative measurements of cardiometabolic relevance or known risk factors for CVD in two sub-cohorts of FinnGen, the population-based national FINRISK study[65] ($n = 26{,}717$) and GeneRISK[66] ($n = 7239$). The associations were tested across 66 quantitative measurements of cardiometabolic relevance in FINRISK, and for 158 sub-lipid species in GeneRISK. In Young Finns Study (YFS)[67] cohort ($n = 1934$), we tested the association of the two variants with three measurements of arterial relevance (carotid artery distensibility, pulse wave velocity, and pulse pressure).

In addition to Finnish cohorts described above, we tested the association of the two variants in Estonian Biobank data (EstBB)[14,68], BioBank Japan (BBJ)[12,13], and UK Biobank (UKBB)[69]. In EstBB ($n = 51{,}388$–$137{,}722$) we tested the associations of both variants with body mass index (BMI), systolic and diastolic blood pressure (SBP and DBP) and pulse pressure (PP), in BBJ in we tested the association of rs534125149 with 17 known quantitative risk factors for CVD and lastly, in the UKBB we tested the association of rs201988637 with 79 measurements of cardiometabolic relevance. In all of these biomarker analyses, a linear regression model adjusted for age and sex was used and for all quantitative risk factors rank-based inverse normal transformation was applied prior to analysis. Bonferroni corrected *p*-value threshold for the number of phenotypes tested was used to assess the significance of resulting associations in each cohort.

For biomarkers that showed significant association in any of the cohorts, we performed a meta-analysis across all cohorts the measurement was available. Meta-analysis was performed using inverse-variance weighted fixed-effects meta-analysis method[70,71]. Bonferroni corrected *p*-value for number of traits tested ($n = 2$) was used to assess the significance of resulting associations in meta-analysis.

Height association. To assess whether the association of rs534125149 with height was due to the *MFGE8* gene, we first performed conditional analysis of height conditioning the association for rs534125149, the lead variant in FinnGen height GWAS (rs11630187) and for previously known height-associated variant in the locus, rs16942341[72], separately. Conditioning the height association on rs534125149 did not have much effect on the association of the lead variant for height (rs11630187) in the region (*p*-value before conditioning $= 5.07 \times 10^{-34}$ and after conditioning $= 1.19 \times 10^{-26}$), whereas when conditioning on the lead variant for height (rs11630187) in the region, the smallest *p*-value in the region was 1.39×10^{-15} (for variant rs28564751). In addition, conditioning on either known height-associated variant rs16942341 or lead variant for height in FinnGen (rs11630187) did not affect on rs534125149's association with height (*p*-value before conditioning $= 8.04 \times 10^{-13}$ and after conditioning $= 3.14 \times 10^{-12}$ and 2.75×10^{-05}, respectively)

In addition, to assess whether the association of rs534125149 with atherosclerotic cardiovascular disease and height reflect the same signal, we performed colocalization analysis in the region ±200kB around rs53412514 using Coloc package in R. The probability for shared signal (PP4) was 9.22×10^{-13}, whereas probability for two independent (PP3) signals was 1, indicating two independent signals for height and coronary atherosclerosis in the locus.

Identifying causal variants. We used FINEMAP[59] on the GWAS summary statistics to identify causal variants underlying the associations for MI (strict definition, i.e., only primary diagnoses accepted) and coronary atherosclerosis. FINEMAP analyses were restricted to a ±1.5 Mb region around the rs534125149. We assessed variants in the top 95% credible sets, i.e., the sets of variants encompassing at least 95% of the probability of being causal (causal probability) within each causal signal in the genomic region. Credible sets were filtered if minimum linkage disequilibrium (LD, r^2) between the variants in the credible set was <0.1, i.e., not clearly representing one signal.

Protein modeling. The predicted structure of lactadherin was obtained from AlphaFold[19] (https://alphafold.ebi.ac.uk/entry/Q08431). Model confidence for the domain containing the variant of interest was scored mostly as very high and was structurally similar to the crystal structure of bovine lactadherin[21] (PDB ID:2PQS). The structure of the insertion variant rs534125149 was predicted using the AlphaFold Colab notebook (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb). Protein structures were visualized using PyMOL[73].

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Full GWAS results are publicly available through FinnGen PheWEB browser (r6.finngen.fi) and also at Open Targets website. The Finnish biobank data can be accessed through the Fingenious® services (web link: https://site.fingenious.fi/en/, email: contact@finbb.fi) managed by FINBB. The UK Biobank resource is available to bona fide researchers for health-related research in the public interest at https://www.ukbiobank.ac.uk/researchers/. The BBJ summary statistics are available at the National Bioscience Database Center (NBDC) Human Database (accession code: hum0197) and at the GWAS catalog (https://www.ebi.ac.uk/gwas/home). They are also browseable at our PheWeb website (https://pheweb.jp/). The variant rs534125149 was originally excluded from the publicly available GWAS summary statistics. Its associations were reported in Supplementary Fig. 4. The BBJ genotype data is accessible on request at the Japanese Genotype–phenotype Archive (http://trace.ddbj.nig.ac.jp/jga/index_e.html) with accession code JGAD00000000123 and JGAS00000000114. Genotype and phenotype data from the Estonian Biobank are available (https://genomics.ut.ee/en/biobank.ee/data-access) upon request. The dataset supporting the conclusions of this article were obtained from the Cardiovascular Risk in Young Finns Study, which comprises health-related participant data. The use of data is restricted under the regulations on professional secrecy (Act on the Openness of Government Activities, 612/1999) and on sensitive personal data (Personal Data Act, 523/1999, implementing the EU data protection directive 95/46/EC). Owing to these restrictions, the data cannot be stored in public repositories or otherwise made publicly available. Data access may be permitted on a case-by-case basis upon request only. Data sharing outside the group is done in collaboration with YFS group and requires a data-sharing agreement. Investigators can submit an expression of interest to the chairman of the publication committee Professor Mika Kähönen (Tampere University, Finland) or Professor Terho Lehtimäki (Tampere University, Finland).

Code availability

The full genotyping and imputation protocol for FinnGen is described at dx.doi.org/10.17504/protocols.io.nmndc5e. The code used for the analyses in this paper are available from the corresponding author upon reasonable request.

Received: 23 June 2021; Accepted: 6 June 2022;
Published online: 17 August 2022

References

1. Kessler, T., Erdmann, J. & Schunkert, H. Genetics of coronary artery disease and myocardial infarction-2013. *Curr. Cardiol. Rep.* **15**, 368 (2013).
2. O'Donnell, C. J. & Nabel, E. G. Genomics of cardiovascular disease. *N. Engl. J. Med.* **365**, 2098–2109 (2011).
3. Nikpay, M. et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121 (2015).
4. Lim, E. T. et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* **10**, e1004494 (2014).
5. Cohen, J. C., Boerwinkle, E., Mosley, T. H. Jr & Hobbs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
6. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N. Engl. J. Med.* **371**, 22–31 (2014).
7. Dewey, F. E. et al. Inactivating variants in ANGPTL4 and risk of coronary artery disease. *N. Engl. J. Med.* **374**, 1123–1133 (2016).
8. Libby, P. et al. Atherosclerosis. *Nat. Rev. Dis. Primers* **5**, 56 (2019).
9. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
10. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2018).
11. Nurnberg, S. T. et al. Genomic profiling of human vascular cells identifies TWIST1 as a causal gene for common vascular diseases. *PLoS Genet.* **16**, e1008538 (2020).
12. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
13. Sakaue, S. et al. A global atlas of genetic associations of 220 deep phenotypes. Preprint at *medRxiv* https://doi.org/10.1101/2020.10.23.20213652 (2021).

14. Leitsalu, L. et al. Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).
15. Hoffmann, T. J. et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* **49**, 54–64 (2017).
16. Evangelou, E. et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
17. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, 1–15 (2014).
18. Soubeyrand, S. et al. Regulation of MFGE8 by the intergenic coronary artery disease locus on 15q26. 1. *Atherosclerosis* **284**, 11–17 (2019).
19. AlQuraishi, M. AlphaFold at CASP13. *Bioinformatics* **35**, 4862–4865 (2019).
20. Andersen, M. H., Graversen, H., Fedosov, S. N., Petersen, T. E. & Rasmussen, J. T. Functional analyses of two cellular binding domains of bovine lactadherin. *Biochemistry (N. Y.)* **39**, 6200–6206 (2000).
21. Lin, L., Huai, Q., Huang, M., Furie, B. & Furie, B. C. Crystal structure of the bovine lactadherin C2 domain, a membrane binding motif, shows similarity to the C2 domains of factor V and factor VIII. *J. Mol. Biol.* **371**, 717–724 (2007).
22. Picariello, G., Ferranti, P., Mamone, G., Roepstorff, P. & Addeo, F. Identification of N-linked glycoproteins in human milk by hydrophilic interaction liquid chromatography and mass spectrometry. *Proteomics* **8**, 3833–3847 (2008).
23. Helenius, A. & Aebi, M. Intracellular functions of N-linked glycans. *Science* **291**, 2364–2369 (2001).
24. Dadu, R. T. & Ballantyne, C. M. Lipid lowering with PCSK9 inhibitors. *Nat. Rev. Cardiol.* **11**, 563 (2014).
25. Benetos, A. et al. Mortality and cardiovascular events are best predicted by low central/peripheral pulse pressure amplification but not by high blood pressure levels in elderly nursing home subjects: the PARTAGE (Predictive Values of Blood Pressure and Arterial Stiffness in Institutionalized Very Aged Population) study. *J. Am. Coll. Cardiol.* **60**, 1503–1511 (2012).
26. Oshima, K. et al. Lactation-dependent expression of an mRNA splice variant with an exon for a multiplyO-glycosylated domain of mouse milk fat globule glycoprotein MFG-E8. *Biochem. Biophys. Res. Commun.* **254**, 522–528 (1999).
27. Aoki, N. et al. Immunologically cross-reactive 57 kDa and 53 kDa glycoprotein antigens of bovine milk fat globule membrane: isoforms with different N-linked sugar chains and differential glycosylation at early stages of lactation. *Biochim. Biophys. Acta (BBA)* **1200**, 227–234 (1994).
28. Deroide, N. et al. MFGE8 inhibits inflammasome-induced IL-1β production and limits postischemic cerebral injury. *J. Clin. Invest.* **123**, 1176–1181 (2013).
29. Hanayama, R. et al. Identification of a factor that links apoptotic cells to phagocytes. *Nature* **417**, 182–187 (2002).
30. Hanayama, R. et al. Autoimmune disease and impaired uptake of apoptotic cells in MFG-E8-deficient mice. *Science* **304**, 1147–1150 (2004).
31. Yoshida, H. et al. Phosphatidylserine-dependent engulfment by macrophages of nuclei from erythroid precursor cells. *Nature* **437**, 754–758 (2005).
32. Tahara, H. et al. Emerging concepts in biomarker discovery; the US-Japan Workshop on Immunological Molecular Markers in Oncology. *J. Transl. Med.* **7**, 1–25 (2009).
33. Neutzner, M. et al. MFG-E8/lactadherin promotes tumor growth in an angiogenesis-dependent transgenic mouse model of multistage carcinogenesis. *Cancer Res.* **67**, 6777–6785 (2007).
34. Taylor, M. R., Couto, J. R., Scallan, C. D., Ceriani, R. L. & Peterson, J. A. Lactadherin (formerly BA46), a membrane-associated glycoprotein expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp (RGD)-dependent cell adhesion. *DNA Cell Biol.* **16**, 861–869 (1997).
35. Jinushi, M. et al. Milk fat globule EGF-8 promotes melanoma progression through coordinated Akt and twist signaling in the tumor microenvironment. *Cancer Res.* **68**, 8889–8898 (2008).
36. Raymond, A., Ensslin, M. A. & Shur, B. D. SED1/MFG-E8: a bi-motif protein that orchestrates diverse cellular interactions. *J. Cell. Biochem.* **106**, 957–966 (2009).
37. Yu, L. et al. MFG-E8 overexpression is associated with poor prognosis in breast cancer patients. *Pathol. Res. Pract.* **215**, 490–498 (2019).
38. Sugano, G. et al. Milk fat globule—epidermal growth factor—factor VIII (MFGE8)/lactadherin promotes bladder tumor development. *Oncogene* **30**, 642–653 (2011).
39. Kanemura, T. et al. Immunoregulatory influence of abundant MFG-E8 expression by esophageal cancer treated with chemotherapy. *Cancer Sci.* **109**, 3393–3402 (2018).
40. Jia, M. et al. Prognostic correlation between MFG-E8 expression level and colorectal Cancer. *Arch. Med. Res.* **48**, 270–275 (2017).
41. Wang, M., Khazan, B. & Lakatta, G. E. Central arterial aging and angiotensin II signaling. *Curr. Hypertens. Rev.* **6**, 266–281 (2010).
42. Wang, M., Monticone, R. E. & Lakatta, E. G. Arterial aging: a journey into subclinical arterial disease. *Curr. Opin. Nephrol. Hypertens.* **19**, 201 (2010).
43. Ortiz, A. et al. Clinical usefulness of novel prognostic biomarkers in patients on hemodialysis. *Nat. Rev. Nephrol.* **8**, 141 (2012).
44. Fu, Z. et al. Milk fat globule protein epidermal growth factor-8: a pivotal relay element within the angiotensin II and monocyte chemoattractant protein-1 signaling cascade mediating vascular smooth muscle cells invasion. *Circ. Res.* **104**, 1337–1346 (2009).
45. Bagnato, C. et al. Proteomics analysis of human coronary atherosclerotic plaque: a feasibility study of direct tissue proteomics by liquid chromatography and tandem mass spectrometry. *Mol. Cell. Proteomics* **6**, 1088–1102 (2007).
46. Li, X. et al. Proteomics approach to study the mechanism of action of grape seed proanthocyanidin extracts on arterial remodeling in diabetic rats. *Int. J. Mol. Med.* **25**, 237–248 (2010).
47. Strøm, C. C. et al. Identification of a core set of genes that signifies pathways underlying cardiac hypertrophy. *Comp. Funct. Genomics* **5**, 459–470 (2004).
48. Lin, Y. et al. Comparative proteomic analysis of rat aorta in a subtotal nephrectomy model. *Proteomics* **10**, 2429–2443 (2010).
49. Ait-Oufella, H. et al. Clinical perspective. *Circulation* **115**, 2168–2177 (2007).
50. Wang, M., Wang, H. & Lakatta, G. Milk fat globule epidermal growth factor VIII signaling in arterial wall remodeling. *Curr. Vasc. Pharmacol.* **11**, 768–776 (2013).
51. Peng, S., Glennert, J. & Westermark, P. Medin-amyloid: a recently characterized age-associated arterial amyloid form affects mainly arteries in the upper part of the body. *Amyloid* **12**, 96–102 (2005).
52. Peng, S. et al. Medin and medin-amyloid in ageing inflamed and non-inflamed temporal arteries. *J. Pathol.* **196**, 91–96 (2002).
53. Grant, J. E. et al. Quantification of protein expression changes in the aging left ventricle of Rattus norvegicus. *J. Proteome Res.* **8**, 4252–4263 (2009).
54. Bagnato, C. et al. Proteomics analysis of human coronary atherosclerotic plaque: a feasibility study of direct tissue proteomics by liquid chromatography and tandem mass spectrometry. *Mol. Cell. Proteom.* **6**, 1088–1102 (2007).
55. Cheyuo, C. et al. Recombinant human MFG-E8 attenuates cerebral ischemic injury: its role in anti-inflammation and anti-apoptosis. *Neuropharmacology* **62**, 890–900 (2012).
56. Silvestre, J. et al. Lactadherin promotes VEGF-dependent neovascularization. *Nat. Med.* **11**, 499–506 (2005).
57. Soubeyrand, S. et al. CARMAL Is a Long Non-coding RNA Locus That Regulates MFGE8 Expression. *Front. Genet.* **11**, 631 (2020).
58. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
59. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
60. Vaara, S. et al. Cohort profile: the Corogene study. *Int. J. Epidemiol.* **41**, 1265–1271 (2012).
61. Collet, J. et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: the Task Force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). *Eur. Heart J.* **42**, 1289–1367 (2021).
62. Dey, R. et al. An efficient and accurate frailty model approach for genome-wide survival association analysis controlling for population structure and relatedness in large-scale biobanks. *bioRxiv* (2020).
63. Breslow, N. E. & Clayton, D. G. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* **88**, 9–25 (1993).
64. Chen, H. et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am. J. Hum. Genet.* **98**, 653–666 (2016).
65. Borodulin, K. et al. Forty-year trends in cardiovascular risk factors in Finland. *Eur. J. Public Health* **25**, 539–546 (2014).
66. Widén, E. et al. How communicating polygenic and clinical risk for atherosclerotic cardiovascular disease impacts health behavior: an observational follow-up study. *Circulation: Genomic and Precision Medicine* **15**, e003459 (2022).
67. Raitakari, O. T. et al. Cohort profile: the cardiovascular risk in Young Finns Study. *Int. J. Epidemiol.* **37**, 1220–1226 (2008).
68. Mitt, M. et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* **25**, 869 (2017).
69. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203 (2018).
70. Cooper, H., Hedges, L. V. & Valentine, J. C. in *The Handbook of Research Synthesis and Meta-analysis* (Russell Sage Foundation, 2019).
71. Evangelou, E. & Ioannidis, J. P. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **14**, 379–389 (2013).
72. Lango, H. A. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).

73. DeLano, W. L. Pymol: an open-source molecular graphics tool. *CCP4 Newslet. Protein Crystallogr.* **40**, 82–92 (2002).

Acknowledgements

We would like to thank all participants of all study cohorts for their generous participation. We also want to thank Dr. Kaoru Ito at RIKEN Center for Integrative Medical Sciences for supporting the study. This work was supported by the Academy of Finland Center of Excellence in Complex Disease Genetics (Grant No 312062 and 336820 to S.R.), the Finnish Foundation for Cardiovascular Research, the Sigrid Juselius Foundation, University of Helsinki HiLIFE Fellow, Grand Challenge grants and Horizon 2020 Research and Innovation Programme (grant number 101016775 "INTERVENE" to S.R.), Academy of Finland grant number 331671 to N.M., the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.16-0125 to L.M.), the Estonian Research Council grant PRG184 [to L.M.] and the Doctoral Programme in Population Health, University of Helsinki [to S.E.R.]. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Celgene Corporation, Celgene International II Sàrl, Genentech Inc., Merck Sharp & Dohme Corp, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, and Novartis AG. Following biobanks are acknowledged for deliverig biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta) and Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi) and FinBB (https://finbb.fi/).

Author contributions

S.E.R. and S.R. designed the study; S.E.R. and J.K. performed the analyses; S.E.R., I.S., N.M., E.W., M.J.D., and S.R. performed interpretation of data; S.E.R., I.S., and S.R. drafted the manuscript; all authors read the manuscript before submission; M. Kanai, K.K., P.P.M., B.H.M. performed analysis for replication cohorts, S.G. performed the protein modeling, M.J.D., J.S., P.P., T.L., O.R., Y.O., L.M., and A.P. provided administrative or material support.

Competing interests

The authors declare no competing interests.

Ethics

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) approved the FinnGen study protocol Nr HUS/990/2017. The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019, THL/1524/5.05.00/2020, and THL/2364/14.02/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020 and Statistics Finland (permit numbers: TK-53-1041-17 and TK-53-90-20). The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 6 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auria Biobank AB17-5154, Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018, Finnish Clinical Biobank Tampere MH0004, Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001. The FINRISK data used for the study were obtained from THL Biobank with application number BB2015_55.1 and UKBB data using the UK Biobank Resource with application number 22627. The study was approved by the Estonian Committee on Bioethics and Human Research (approval number 1.1-12/624).

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-022-03552-0.

Correspondence and requests for materials should be addressed to Samuli Ripatti.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. This article has been peer reviewed as part of Springer Nature's Guided Open Access initiative.

Reprints and permission information is available at http://www.nature.com/reprints

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

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

© The Author(s) 2022

Estonian Biobank Research Team

Tõnu Esko[6], Andres Metspalu[6], Reedik Mägi[6] & Mari Nelis[6]

The Biobank Japan Project

Koichi Matsuda[16,17], Yuji Yamanashi[18], Yoichi Furukawa[19], Takayuki Morisaki[19], Yoshinori Murakami[20], Yoichiro Kamatani[17,21], Kaori Muto[22], Akiko Nagai[22], Wataru Obara[23], Ken Yamaji[24], Kazuhisa Takahashi[25], Satoshi Asai[26,27], Yasuo Takahashi[27], Takao Suzuki[28], Nobuaki Sinozaki[28], Hiroki Yamaguchi[29], Shiro Minami[30], Shigeo Murayama[31], Kozo Yoshimori[32], Satoshi Nagayama[33], Daisuke Obata[34], Masahiko Higashiyama[35], Akihide Masumoto[36] & Yukihiro Koretsune[37]

[16]Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [17]Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [18]Division of Genetics, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [19]Division of Molecular Pathology, IMSUT Hospital Department of Internal Medicine, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [20]Department of Cancer Biology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [21]Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [22]Department of Public Policy, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [23]Department of Urology, Iwate Medical University, Iwate, Japan. [24]Department of Internal Medicine and Rheumatology, Juntendo University Graduate School of Medicine, Tokyo, Japan. [25]Department of Respiratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan. [26]Division of Pharmacology, Department of Biomedical Science, Nihon University School of Medicine, Tokyo, Japan. [27]Division of Genomic Epidemiology and Clinical Trials, Clinical Trials Research Center, Nihon University. School of Medicine, Tokyo, Japan. [28]Tokushukai Group, Tokyo, Japan. [29]Departmentof Hematology, Nippon Medical School, Tokyo, Japan. [30]Department of Bioregulation, Nippon Medical School, Kawasaki, Japan. [31]Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan. [32]Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo, Japan. [33]The Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo, Japan. [34]Center for Clinical Research and Advanced Medicine, Shiga University of Medical Science, Shiga, Japan. [35]Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan. [36]IIZUKA HOSPITAL, Fukuoka, Japan. [37]National Hospital Organization Osaka National Hospital, Osaka, Japan.

FinnGen

Steering Committee Aarno Palotie[1] & Mark Daly[1]

Pharmaceutical Companies Bridget Riley-Gills[38], Howard Jacob[38], Dirk Paul[39], Heiko Runz[40], Sally John[40], Robert Plenge[41], Mark McCarthy[42], Julie Hunkapiller[42], Meg Ehm[43], Kirsi Auro[43], Caroline Fox[44], Anders Mälarstig[45], Katherine Klinger[46], Deepak Raipal[46], Tim Behrens[47], Robert Yang[48] & Richard Siegel[49]

University of Helsinki & Biobanks Tomi Mäkelä[50], Jaakko Kaprio[1], Petri Virolainen[51], Antti Hakanen[51], Terhi Kilpi[52], Markus Perola[52], Jukka Partanen[53], Anne Pitkäranta[54], Juhani Junttila[55], Raisa Serpi[55], Tarja Laitinen[56], Johanna Mäkelä[56], Veli-Matti Kosma[57] & Urho Kujala[58]

Other Experts/Non-Voting Members Outi Tuovila[59] & Raimo Pakkanen[59]

Scientific Committee

Pharmaceutical Companies Jeffrey Waring[38], Ali Abbasi[38], Mengzhen Liu[38], Ioanna Tachmazidou[39], Chia-Yen Chen[40], Heiko Runz[40], Shameek Biswas[41], Julie Hunkapiller[42], Meg Ehm[43], Neha Raghavan[44], Adriana Huertas-Vazquez[44], Anders Mälarstig[45], Xinli Hu[45], Katherine Klinger[46], Matthias Gossel[46], Robert Graham[47], Tim Behrens[47], Beryl Cummings[47], Wilco Fleuren[48], Dawn Waterworth[48], Nicole Renaud[49] & Ma´en Obeidat[49]

University of Helsinki & Biobanks Samuli Ripatti[1], Johanna Schleutker[51], Markus Perola[52], Mikko Arvas[53], Olli Carpén[54], Reetta Hinttala[55], Johannes Kettunen[55], Johanna Mäkelä[56], Arto Mannermaa[57], Jari Laukkanen[58] & Urho Kujala[58]

Clinical Groups

Neurology Group Reetta Kälviäinen[60], Valtteri Julkunen[60], Hilkka Soininen[60], Anne Remes[61], Mikko Hiltunen[60], Jukka Peltola[62], Pentti Tienari[63], Juha Rinne[64], Roosa Kallionpää[64], Ali Abbasi[38], Adam Ziemann[38], Jeffrey Waring[38], Sahar Esmaeeli[38], Nizar Smaoui[38], Anne Lehtonen[38], Susan Eaton[40], Heiko Runz[40], Sanni Lahdenperä[40], Janet van Adelsberg[41], Shameek Biswas[41], Julie Hunkapiller[42], Natalie Bowers[42], Edmond Teng[42], Sarah Pendergrass[42], Onuralp Soylemez[44], Kari Linden[45], Fanli Xu[43], David Pulford[43], Kirsi Auro[43], Laura Addis[43], John Eicher[43], Minna Raivio[63], Sarah Pendergrass[42], Beryl Cummings[47] & Juulia Partanen[1]

Gastroenterology Group Martti Färkkilä[63], Jukka Koskela[63], Sampsa Pikkarainen[63], Airi Jussila[62], Katri Kaukinen[62], Timo Blomster[61], Mikko Kiviniemi[60], Markku Voutilainen[64], Ali Abbasi[38], Graham Heap[38], Jeffrey Waring[38], Nizar Smaoui[38], Fedik Rahimov[38], Anne Lehtonen[38], Keith Usiskin[41], Tim Lu[42], Natalie Bowers[42], Danny Oh[42], Sarah Pendergrass[42], Kirsi Kalpala[45], Melissa Miller[45], Xinli Hu[45], Linda McCarthy[43], Onuralp Soylemez[44] & Mark Daly[1]

Rheumatology Group Kari Eklund[63], Antti Palomäki[64], Pia Isomäki[62], Laura Pirilä[64], Oili Kaipiainen-Seppänen[60], Johanna Huhtakangas[61], Ali Abbasi[38], Jeffrey Waring[38], Fedik Rahimov[38], Apinya Lertratanakul[38], Nizar Smaoui[38], Anne Lehtonen[38], David Close[39], Marla Hochfeld[41], Natalie Bowers[42], Sarah Pendergrass[42], Onuralp Soylemez[44], Kirsi Kalpala[45], Nan Bing[45], Xinli Hu[45], Jorge Esparza Gordillo[43], Kirsi Auro[43], Dawn Waterworth[48] & Nina Mars[1]

Pulmonology Group Tarja Laitinen[62], Margit Pelkonen[60], Paula Kauppi[63], Hannu Kankaanranta[62], Terttu Harju[61], Riitta Lahesmaa[64], Nizar Smaoui[38], Alex Mackay[39], Glenda Lassi[39], Susan Eaton[40], Steven Greenberg[41], Hubert Chen[42], Sarah Pendergrass[42], Natalie Bowers[42], Joanna Betts[43], Soumitra Ghosh[43], Kirsi Auro[43], Rajashree Mishra[43] & Sina Rüeger[1]

Cardiometabolic Diseases Group Teemu Niiranen[65], Felix Vaura[65], Veikko Salomaa[65], Markus Juonala[64], Kaj Metsärinne[64], Mika Kähönen[62], Juhani Junttila[61], Markku Laakso[60], Jussi Pihlajamäki[60], Daniel Gordin[63], Juha Sinisalo[63], Marja-Riitta Taskinen[63], Tiinamaija Tuomi[63], Jari Laukkanen[66], Benjamin Challis[39], Dirk Paul[39], Julie Hunkapiller[42], Natalie Bowers[42], Sarah Pendergrass[42], Onuralp Soylemez[44], Jaakko Parkkinen[45], Melissa Miller[45], Russell Miller[45], Audrey Chu[43], Kirsi Auro[43], Keith Usiskin[41], Amanda Elliott[1,3], Joel Rämö[1], Samuli Ripatti[1], Mary Pat Reeve[1] & Sanni Ruotsalainen[1]

Oncology Group Tuomo Meretoja[63], Heikki Joensuu[63], Olli Carpén[63], Lauri Aaltonen[63], Johanna Mattson[63], Annika Auranen[62], Peeter Karihtala[61], Saila Kauppila[61], Päivi Auvinen[60], Klaus Elenius[64], Johanna Schleutker[64], Relja Popovic[38], Jeffrey Waring[38], Bridget Riley-Gillis[38], Anne Lehtonen[38], Jennifer Schutzman[42], Julie Hunkapiller[42], Natalie Bowers[42], Sarah Pendergrass[42], Andrey Loboda[44], Aparna Chhibber[44], Heli Lehtonen[45], Stefan McDonough[45], Marika Crohns[46], Sauli Vuoti[46], Diptee Kulkarni[43], Kirsi Auro[43], Esa Pitkänen[1], Nina Mars[1] & Mark Daly[1]

Opthalmology Group Kai Kaarniranta[60], Joni A. Turunen[63], Terhi Ollila[63], Sanna Seitsonen[63], Hannu Uusitalo[62], Vesa Aaltonen[64], Hannele Uusitalo-Järvinen[62], Marja Luodonpää[61], Nina Hautala[61], Mengzhen Liu[38], Heiko Runz[40], Stephanie Loomis[40], Erich Strauss[42], Natalie Bowers[42], Hao Chen[42], Sarah Pendergrass[42], Anna Podgornaia[44], Juha Karjalainen[1,3] & Esa Pitkänen[1]

Dermatology Group Kaisa Tasanen[61], Laura Huilaja[61], Katariina Hannula-Jouppi[63], Teea Salmi[62], Sirkku Peltonen[64], Leena Koulu[64], Kirsi Kalpala[45], Ying Wu[45], David Choy[42], Sarah Pendergrass[42], Nizar Smaoui[38], Fedik Rahimov[38], Anne Lehtonen[38] & Dawn Waterworth[48]

Odontology Group Pirkko Pussinen[63], Aino Salminen[63], Tuula Salo[63], David Rice[63], Pekka Nieminen[63], Ulla Palotie[63], Juha Sinisalo[63], Maria Siponen[60], Liisa Suominen[60], Päivi Mäntylä[60], Ulvi Gursoy[64], Vuokko Anttonen[61], Kirsi Sipilä[61] & Sarah Pendergrass[42]

Women's Health and Reproduction Group Hannele Laivuori[1], Venla Kurra[62], Oskari Heikinheimo[63], Ilkka Kalliala[63], Laura Kotaniemi-Talonen[62], Kari Nieminen[62], Päivi Polo[64], Kaarin Mäkikallio[64], Eeva Ekholm[64], Marja Vääräsmäki[61], Outi Uimari[61], Laure Morin-Papunen[61], Marjo Tuppurainen[60], Katja Kivinen[1], Elisabeth Widen[1], Taru Tukiainen[1], Mary Pat Reeve[1], Mark Daly[1], Liu Aoxing[1], Eija Laakkonen[67], Niko Välimäki[68], Lauri Aaltonen[63], Johannes Kettunen[61], Mikko Arvas[69], Jeffrey Waring[38], Bridget Riley-Gillis[38], Mengzhen Liu[38], Janet Kumar[43], Kirsi Auro[43], Andrea Ganna[1] & Sarah Pendergrass[42]

FinnGen Analysis Working Group Justin Wade Davis[38], Bridget Riley-Gillis[38], Danjuma Quarless[38], Fedik Rahimov[38], Sahar Esmaeeli[38], Slavé Petrovski[39], Eleonor Wigmore[39], Adele Mitchell[40], Benjamin Sun[40], Ellen Tsai[40], Denis Baird[40], Paola Bronson[40], Ruoyu Tian[40], Stephanie Loomis[40], Yunfeng Huang[40], Joseph Maranville[41], Shameek Biswas[41], Elmutaz Mohammed[41], Samir Wadhawan[41], Erika Kvikstad[41], Minal Caliskan[41], Diana Chang[42], Julie Hunkapiller[42], Tushar Bhangale[42], Natalie Bowers[42], Sarah Pendergrass[42], Kirill Shkura[44], Victor Neduva[44], Xing Chen[45], Åsa Hedman[45], Karen S. King[43], Padhraig Gormley[43], Jimmy Liu[43], Clarence Wang[46], Ethan Xu[46], Franck Auge[46], Clement Chatelain[46], Deepak Rajpal[46], Dongyu Liu[46], Katherine Call[46], Tai-He Xia[46], Beryl Cummings[47], Matt Brauer[47], Huilei Xu[49], Amy Cole[49], Jonathan Chung[49], Jaison Jacob[49], Katrina de Lange[49], Jonas Zierer[49], Mitja Kurki[1,3], Samuli Ripatti[1], Mark Daly[1], Juha Karjalainen[1,3], Aki Havulinna[1], Juha Mehtonen[1], Priit Palta[1], Shabbeer Hassan[1], Pietro Della Briotta Parolo[1], Wei Zhou[3], Mutaamba Maasha[3], Susanna Lemmelä[1], Manuel Rivas[70], Aarno Palotie[1], Arto Lehisto[1], Andrea Ganna[1], Vincent Llorens[1], Hannele Laivuori[1], Mari E. Niemi[1], Taru Tukiainen[1], Mary Pat Reeve[1], Henrike Heyne[1], Nina Mars[1], Kimmo Palin[68], Javier Garcia-Tabuenca[71], Harri Siirtola[71], Tuomo Kiiskinen[1], Jiwoo Lee[1,3], Kristin Tsuo[1,3], Amanda Elliott[1,3], Kati Kristiansson[52], Mikko Arvas[53], Kati Hyvärinen[69], Jarmo Ritari[69], Miika Koskinen[54], Olli Carpén[54], Johannes Kettunen[55], Katri Pylkäs[72], Marita Kalaoja[72], Minna Karjalainen[72], Tuomo Mantere[55], Eeva Kangasniemi[56], Sami Heikkinen[73], Arto Mannermaa[57], Eija Laakkonen[67], Samuel Heron[74], Dhanaprakash Jambulingam[74], Venkat Subramaniam Rathinakannan[74] & Nina Pitkänen[51]

Biobank Directors Lila Kallio[51], Sirpa Soini[52], Jukka Partanen[53], Eero Punkka[54], Raisa Serpi[55], Johanna Mäkelä[56], Veli-Matti Kosma[57] & Teijo Kuopio[58]

FinnGen Teams

Administration Anu Jalanko[1], Huei-Yi Shen[1], Risto Kajanne[1] & Mervi Aavikko[1]

Analysis Mitja Kurki[1,3], Juha Karjalainen[1,3], Pietro Della Briotta Parolo[1], Arto Lehisto[1], Juha Mehtonen[1], Wei Zhou[3], Masahiro Kanai[3] & Mutaamba Maasha[3]

Clinical Endpoint Development Hannele Laivuori[1], Aki Havulinna[1], Susanna Lemmelä[1], Tuomo Kiiskinen[1] & L. Elisa Lahtela[1]

Communication Mari Kaunisto[1]

E-Science Elina Kilpeläinen[1], Timo P. Sipilä[1], Georg Brein[1], Oluwaseun Alexander Dada[1], Awaisa Ghazal[1] & Anastasia Shcherban[1]

Genotyping Kati Donner[1] & Timo P. Sipilä[1]

Sample Collection Coordination Anu Loukola[54]

Sample Logistics Päivi Laiho[52], Tuuli Sistonen[52], Essi Kaiharju[52], Markku Laukkanen[52], Elina Järvensivu[52], Sini Lähteenmäki[52], Lotta Männikkö[52] & Regis Wong[52]

Registry Data Operations Hannele Mattsson[52], Kati Kristiansson[52], Susanna Lemmelä[1], Sami Koskelainen[52], Tero Hiekkalinna[52] & Teemu Paajanen[52]

Sequencing Informatics Priit Palta[1], Kalle Pärn[1], Shuang Luo[1] & Vishal Sinha[1]

Trajectory Tarja Laitinen[62], Mary Pat Reeve[1], Harri Siirtola[71], Javier Gracia-Tabuenca[71], Mika Helminen[71], Tiina Luukkaala[71] & Iida Vähätalo[71]

Data Protection Officer Tero Jyrhämä[1]

FinBB—Finnish Biobank Cooperative Marco Hautalahti[75], Laura Mustaniemi[75], Mirkka Koivusalo[75], Sarah Smith[75] & Tom Southerington[75]

[38]Abbvie, Chicago, IL, USA. [39]Astra Zeneca, Cambridge, UK. [40]Biogen, Cambridge, MA, USA. [41]Celgene, Summit, NJ, United States/Bristol Myers Squibb, New York, NY, USA. [42]Genentech, San Francisco, CA, USA. [43]GlaxoSmithKline, Brentford, UK. [44]Merck, Kenilworth, NJ, USA. [45]Pfizer, New York, NY, USA. [46]Sanofi, Paris, France. [47]Maze Therapeutics, San Francisco, CA, USA. [48]Janssen Biotech, Beerse, Belgium. [49]Novartis, Basel, Switzerland. [50]HiLIFE, University of Helsinki, Finland, Finland. [51]Auria Biobank/University of Turku/Hospital District of Southwest Finland, Turku, Finland. [52]THL Biobank/The National Institute of Health and Welfare Helsinki, Helsinki, Finland. [53]Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland. [54]Helsinki Biobank/Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [55]Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland. [56]Finnish Clinical Biobank Tampere/University of Tampere/Pirkanmaa Hospital District, Tampere, Finland. [57]Biobank of Eastern Finland/University of Eastern Finland/Northern Savo Hospital District, Kuopio, Finland. [58]Central Finland Biobank/University of Jyväskylä/Central Finland Health Care District, Jyväskylä, Finland. [59]Business Finland, Helsinki, Finland. [60]Northern Savo Hospital District, Kuopio, Finland. [61]Northern Ostrobothnia Hospital District, Oulu, Finland. [62]Pirkanmaa Hospital District, Tampere, Finland. [63]Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [64]Hospital District of Southwest Finland, Turku, Finland. [65]The National Institute of Health and Welfare Helsinki, Helsinki, Finland. [66]Central Finland Health Care District, Jyväskylä, Finland. [67]University of Jyväskylä, Jyväskylä, Finland. [68]University of Helsinki, Helsinki, Finland. [69]Finnish Red Cross Blood Service, Helsinki, Finland. [70]University of Stanford, Stanford, CA, USA. [71]University of Tampere, Tampere, Finland. [72]University of Oulu, Oulu, Finland. [73]University of Eastern Finland, Kuopio, Finland. [74]University of Turku, Turku, Finland. [75]Finnish Biobank Cooperative—FINBB, Helsinki, Finland.