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

Sanni E. Ruotsalainen, Ida Surakka, Nina Mars, Juha Karjalainen, Mitja Kurki, Masahiro Kanai, Kristi Krebs, Sarah Graham, Pashupati P. Mishra, Binisha H. Mishra, Juha Sinisalo, Priit Palta, Terho Lehtimäki, Olli Raitakari, Estonian Biobank Research Team, Lili Milani, The Biobank Japan Project, Yukinori Okada, FinnGen, Aarno Palotie, Elisabeth Widen, Mark J. Daly & Samuli Ripatti

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.
Cardiovascular disease (CVD) is the leading cause of premature death and disability worldwide, with both genetic and environmental determinants. 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.

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. 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, PCSK9, APOC3, and ANGPTLA 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.

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 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 × 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 × 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 FHDL1) have not been previously reported to associate with any CVD-related endpoints or risk factor for CVD in GWAS Catalog with any CVD-related endpoints or risk factor for CVD in GWAS Catalog. 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 × 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 NSFEs (Non-Finnish, Estonian or Swedish Europeans). Interestingly, MFGE8 is mainly expressed in coronary and tibial arteries according to data from GTEx v8.

**Replication.** Association between rs534125149 in MFGE8 locus with CHD was replicated in Biobank Japan (BBJ) and the Estonian Biobank (EstBB) (35,644 cases and 328,461 controls total; OR = 0.752 [0.67–0.84], p = 4.37 × 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 FHDL1), 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 FHDL1) 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 FHDL1 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 × 10^-16) and
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], \(p = 7.94 \times 10^{-10}\)) 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 inframe insertion rs534125149 and splice acceptor variant rs201988637 showed also significant association in survival analysis when analyzing survival time from birth to first diagnosis 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\(^{14,15}\). 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 | AF (p-value) | Info (Post-pr) | Nearest gene chrom:pos:ref_alt (rsid) | AF (p-value) | Info (Post-pr) | Enrichment (NFE) | p-value | rsid |
|--------------------------------------|------------------------|--------------|--------------|---------------|--------------------------------------|--------------|---------------|----------------|---------|-------|
| chr15:88901702:C_CTGT (rs534125149)  | Inframe insertion      | MFGE8        | 0.0029       | 0.99          | chr11:57380633:A:G (rs118042209)   | 0.0003       | 0.0010         | 0.87            | 0.91    | rs118042209 |
| chrX:136194941:C_G (rs5974585)      | Intron variant         | FHL1         | 0.250         | 0.62          | chr8:24548925:T:G (rs764568652)    | 0.003        | 0.098          | 0.89            | 0.99    | rs764568652 |

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 × 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
An association of a splice acceptor variant rs201988637 in MFGE8 with lower pulse pressure, a potential biomarker for arterial stiffness, 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. Lactadherin is known to play important roles in several other biological processes, including apoptotic cell clearance and adaptive immunity, 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, 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, and cancers, such as breast, bladder, esophageal, and colorectal cancer. Recently, not only has MFG-E8 expression emerged as a molecular hallmark of adverse cardiovascular remodeling with age, but MFG-E8 signaling has also been found to mediate the vascular outcomes of arterial and matrix responses to the hostile stresses associated with hypertension, diabetes, and atherosclerosis.

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 inflammatory signaling network. 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 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 phenotype-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 and APOC3, 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. Based on the phenotype-wide association profile for the splice acceptor variant rs201988637, we hypothesize that inhibiting MFGE8 could lower the CHD risk, if the variant can be proven to be loss-of-function in MFGE8.

| Cohort                      | N cases / controls | AF (%) | OR [CI]               | P- Value |
|-----------------------------|--------------------|--------|-----------------------|----------|
| CHD                         |                    |        |                       |          |
| FinnGen                     | 28 598 / 222 551   | 2.9    | 0.754 [0.706–0.806]   | 1.63e–16 |
| EstBB                       | 6 564 / 176 815    | 0.6    | 0.628 [0.622–1.102]   | 0.194    |
| BBJ                         | 29 080 / 149 646   | 1.2    | 0.74 [0.657–0.834]    | 6.62e–07 |
| Meta-analysis               | 64 242 / 551 012   |        | 0.754 [0.712–0.798]   | 2.69e–22 |
| Meta-analysis without FinnGen| 35 644 / 328 461   |        | 0.752 [0.674–0.84]    | 4.37e–07 |
| MI                          |                    |        |                       |          |
| FinnGen                     | 14 305 / 222 551   | 2.9    | 0.742 [0.679–0.81]    | 1.95e–11 |
| EstBB                       | 3 110 / 190 450    | 0.6    | 0.643 [0.444–0.932]   | 0.0197   |
| BBJ                         | 14 992 / 146 214   | 1.2    | 0.771 [0.662–0.896]   | 0.00928  |
| Meta-analysis               | 32 407 / 559 215   |        | 0.745 [0.691–0.802]   | 9.99e–15 |
| Meta-analysis without FinnGen| 18 102 / 336 664   |        | 0.751 [0.652–0.865]   | 7.15e–05 |

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.
Increase within the arterial walls and hearts of various species, including rats, humans, and monkeys, and MFG-E8 is markedly up-regulated in rat aortic walls with aging. High levels of MFG-E8 have also been detected within endothelial cells, SMC, and macrophages of atherosclerotic aortae in both mice and humans. 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, as well as inducing recovery from ischemia by facilitating angiogenesis.

In addition, expression of MFG-E8 is highly enriched to tissues relevant to the reported association, such as aorta. Genes nearby MFG-E8, including ABHD2 and HAPLN3, are, however similarly to MFG-E8 enriched to arteries. 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 MFG-E8.

Our study does, however, have a few limitations. First, our primary association results come from Finnish population with considerable elevation in allele frequency in MFG-E8 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.

**Table 1**

| Cohort | N cases / controls | AF (%) | OR [CI] | P- Value |
|--------|--------------------|--------|---------|----------|
| CHD    |                    |        |         |          |
| FinnGen | 28 598 / 222 551  | 0.92   | 0.697 [0.62–0.78] | 1.91e–09 |
| UKBB   | 25 054 / 318 559   | 1.4    | 1.081 [0.995–1.175] | 0.066   |
| EstBB  | 6 564 / 178 815    | 1.22   | 1.137 [1.019–1.407] | 0.238   |
| Meta-analysis | 60 216 / 719 925    | 0.949 [0.889–1.013] | 0.115   |
| Meta-analysis without FinnGen | 31 618 / 497 374    | 1.088 [1.006–1.177] | 0.034   |

**Table 2**

| Cohort | N cases / controls | AF (%) | OR [CI] | P- Value |
|--------|--------------------|--------|---------|----------|
| MI     |                    |        |         |          |
| FinnGen | 14 305 / 222 551  | 0.92   | 0.766 [0.66–0.88] | 0.0099  |
| UKBB   | 7 030 / 335 495   | 1.4    | 1.034 [0.983–1.197] | 0.655   |
| EstBB  | 3 110 / 190 450   | 1.22   | 0.922 [0.705–1.206] | 0.557   |
| Meta-analysis | 24 945 / 748 496    | 0.893 [0.811–0.984] | 0.023   |
| Meta-analysis without FinnGen | 10 640 / 626 945    | 1.007 [0.885–1.145] | 0.191   |

**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.
**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.finnngen.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 × 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 × 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         | SNP ID          | Most severe consequence | Most severe consequence | Original GWAS results | Conditional results |
|-------------------|-----------------|-------------------------|-------------------------|-----------------------|--------------------|
|                   |                 | OR [CI]                 | p-value                 | OR [CI]               | p-value            |
| Coronary infarction | chr15:88901702:C:CTGT | Inframe insertion       | 0.75 [0.70–0.80]        | 0.80 [0.73–0.85]     | 7.68 × 10⁻⁶        |
|                   |                 |                         |                         |                       |                    |
| Myocardial infarction, strict | chr15:88899813:T:G (rs201988637) | Splice acceptor variant | 0.72 [0.64–0.81]      | 0.69 [0.58–0.83]     | 2.63 × 10⁻⁶        |

This table presents 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.

aConditional on rs201988637.
bConditional on rs534125149.

Methods

Study cohort and data. We studied total of 2,861 disease endpoints in Finnish biobank study FinnGen (n = 260,405) (Table 3). FinnGen 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.finnngen.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/protopocols.io.nqdlwn. 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 × 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).
**Pulse Pressure: rs201988637**

| Cohort     | N  | AF (%) | P-value | Beta [95% CI] |
|------------|----|--------|---------|---------------|
| FINRISK    | 26666 | 0.64 | 0.02 | -0.12 [-0.22, -0.02] |
| GeneRISK   | 7318  | 0.74 | 0.341 | -0.09 [-0.28, 0.10] |
| YFS        | 1934  | 0.60 | 0.8 | -0.05 [-0.40, 0.30] |
| EstBB      | 51380 | 0.21 | 0.00708 | -0.17 [-0.29, -0.05] |
| UKBB       | 456102 | 0.01 | 0.135 | -0.39 [-0.91, 0.12] |
| Meta_analysis | 543400 | -   | 0.00017 | -0.13 [-0.20, -0.06] |

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.

| N (%)  | Females | Males  |
|--------|---------|--------|
| Age (mean (sd)) | 260,405 | 27.29 (5.36) | 27.72 (4.76) |
| Statin use (N (%)) | 86,466 (32.2%) | 40,422 (27.48%) | 46,044 (40.62%) |
| Hypertension (N (%)) | 68,005 (26.11%) | 33,420 (22.72%) | 34,585 (30.51%) |
| Smoking (N (%)) | 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%) |

| All | Females | Males  |
|-----|---------|--------|
| BMI (mean (sd)) | 27.29 (5.36) | 27.72 (4.76) |
| Hypertension | 68,005 (26.11%) | 33,420 (22.72%) | 34,585 (30.51%) |
| Smoking | 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%) |

**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 saddle-point approximation (SPA) to calibrate unbalanced case-control ratios\(^5\). Accurate Implementation of Generalized mixed model (SAIGE), which uses sad-

**Survival analysis.** Survival analysis for coronary atherosclerosis and myocardial infarction was performed using GATE\(^5\), 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 EPA to estimate the null distribution of the score statistic. For coronary atherosclerosis and myocardial infarction, survival time from birth to first diagnosis 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 frame 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 subcohorts of FinnGen, the population-based national FINRISK study\(^{65} (n = 26,717)\) and GeneRISK\(^{66} (n = 7,239)\). 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 (EutEB)\(^{68,69}\), BioBank Japan (BBJ)\(^{70,71}\), and UK Biobank (UKBB)\(^{69}\). In EutEB (\(n = 51,388-137,722\)) we tested the association of both variants with body mass index (BMI), systolic and diastolic blood pressure (SBP and DBP) and pulse pressure (PP), in BBJ 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\(^{72,73}\). 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 rs3151425194 or splice acceptor variant rs201988637. We tested both variants in the same model, and the lead variant for height (rs11630187) in the region, the smallest \(p\)-value before conditioning \(\sim 1.19 \times 10^{-29}\), whereas conditioning on the lead variant for height (rs11630187) in the region, the smallest \(p\)-value in the region was \(1.9 \times 10^{-15}\) (for variant rs28564751). In addition, conditioning on either known height-associated variant rs16492341 or lead variant for height in FinnGen (rs163187) did not affect on rs534125149's association with height \((p < 0.1)\) 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).

### Identifying causal variants

We used FASTMAP\(^{59}\) 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. FASTMAP 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/Q84831). 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\(^{23}\) (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\(^{23}\).

### 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.ingenion.fi) and also at Open Targets website. The Finnish biobank data can be accessed through the Fingensious\(^{5}\) services (web link: https://site.fingensious.fi/en/, email: contact@fnbfi) managed by FINRB. 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 BBI summary statistics are available at the National Bioscience Data Center (NBDC) Human Database (accession code: hum0197) and at the GWAS catalog (https://www.ebi.ac.uk/gwas/home). They are also browsable 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 BBI genotype data is accessible on request at the Japanese Genotype–phenotype Archive (http://trace.ddbj.nig.ac.jp/sga/index_e.html) with accession code JGAD0000000123 and JGAS0000000114. 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.nmmndc5e. 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. Cardioi. 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 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. Karzeczki, 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–S17 (2018).
13. Sakae, 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).
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.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 Sarl, 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 delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biobank), TRL Biobank (www.thl.fi/biobank), Helsinki Biobank (www. helsinki.fi/biobank), THL Biobank (www.thl. fi/biobank), Biobank of Eastern Finland (www.itu-suomenbiobankki.fi/en), Central Finland Biobank (www.kszhp.fi/fin-fl/Potilaat/Biobankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biobankki/tietoa- ja-käytäntöä) and Terveystalo Biobank (www.terveystalo.com/t/tarvikkeet/terveytalo-biobankki/biobankki/), Biobank Borealis of Northern Finland (www.winter-biobank.fi/en). All Finnish Biobanks are members of BMBL8 infrastructure (www.bmbli. 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 material in this article are included in the article 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 Esko6, Andres Metspalu6, Reedik Mägi6 & Mari Nelis6

The Biobank Japan Project

Koichi Matsuda16,17, Yuji Yamanashi18, Yoichi Fukikawa19, Takayuki Morisaki19, Yoshinori Murakami20, Yoichiro Kamatani17,21, Kaori Muto22, Akiko Nagai22, Wataru Obara23, Ken Yamaji24, Kazuhisa Takahashi25, Satoshi Asai26,27, Yasuo Takahashi27, Takao Suzuki28, Nobuaki Sinozaki28, Hiroki Yamaguchi29, Shiro Minami30, Shigeo Murayama31, Kozo Yoshimori32, Satoshi Nagayama33, Daisuke Obata34, Masahiko Higashiyama35, Akihide Masumoto36 & Yuukiho Koretsune37

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

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.

Estonian Biobank Research Team

Tõnu Esko6, Andres Metspalu6, Reedik Mägi6 & Mari Nelis6

The Biobank Japan Project

Koichi Matsuda16,17, Yuji Yamanashi18, Yoichi Fukikawa19, Takayuki Morisaki19, Yoshinori Murakami20, Yoichiro Kamatani17,21, Kaori Muto22, Akiko Nagai22, Wataru Obara23, Ken Yamaji24, Kazuhisa Takahashi25, Satoshi Asai26,27, Yasuo Takahashi27, Takao Suzuki28, Nobuaki Sinozaki28, Hiroki Yamaguchi29, Shiro Minami30, Shigeo Murayama31, Kozo Yoshimori32, Satoshi Nagayama33, Daisuke Obata34, Masahiko Higashiyama35, Akihide Masumoto36 & Yuukiho Koretsune37

ARTICLE COMMUNICATIONS BIOLOGY | https://doi.org/10.1038/s42003-022-03552-0

12 COMMUNICATIONS BIOLOGY | (2022)5:802 | https://doi.org/10.1038/s42003-022-03552-0 | www.nature.com/commsbio
**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 Hakonen\(^51\), Terhi Kilpi\(^52\), Markus Perola\(^52\), Jukka Partanen\(^53\), Anne Pitkäranta\(^54\), Juhani Junnila\(^55\), 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 Färkkilä63, Jukka Koskela63, Sampsa Pikkarainen63, Airi Jussila62, Katri Kaukinen62, Timo Blomster61, Mikko Kiviniemi60, Markku Voutilainen64, Ali Abbasi38, Graham Heap38, Jeffrey Waring38, Nizar Smaoui38, Fedik Rahimov38, Anne Lehtonen38, Keith Usiskin41, Tim Lu’42, Natalie Bowers42, Danny Oh42, Sarah Pendergrass42, Kirsi Kalpala45, Melissa Miller45, Xinli Hu45, Linda McCarthy43, Onuralp Soylemez44 & Mark Daly1

Rheumatology Group Kari Eklund63, Antti Palomäki64, Pia Isomäki62, Laura Pirilä64, Olli Kaipiainen-Seppänen60, Johanna Huhtakangas61, Ali Abbasi38, Jeffrey Waring38, Fedik Rahimov38, Apinya Lertratanakul38, Nizar Smaoui38, Anne Lehtonen38, David Close39, Marla Hochfeld41, Natalie Bowers42, Sarah Pendergrass42, Onuralp Soylemez44, Kirsi Kalpala45, Nan Bing45, Xinli Hu45, Jorge Esparza Gordillo43, Kirsi Auro43, Dawn Waterworth48 & Nina Mars1

Pulmonology Group Tarja Laitinen62, Margit Pelkonen60, Paula Kauppi63, Hannu Kankaanranta62, Terttu Harju61, Riitta Lahesmaa64, Nizar Smaoui38, Alex Mackay39, Glenda Lassi39, Susan Eaton40, Steven Greenberg41, Hubert Chen42, Sarah Pendergrass42, Natalie Bowers42, Joanna Betts43, Soumitra Ghosh43, Kirsi Auro43, Rajashree Mishra43 & Sina Rüeger1

Cardiometabolic Diseases Group Teemu Niiranen65, Felix Vaura65, Veikko Salomaa65, Markus Juonala64, Kai Metsärinne64, Mika Kähön62, Juhani Junntila61, Markku Laakso60, Jussi Pihlajamäki60, Daniel Gordin63, Juha Sinisalo63, Marja-Riitta Taskinen63, Tiinaamija Tuomi63, Jari Laukkana66, Benjamin Challis39, Dirk Paul39, Julie Hunkapiller42, Natalie Bowers42, Sarah Pendergrass42, Onuralp Soylemez44, Jaakko Parkkinen45, Melissa Miller45, Russell Miller45, Audrey Chu43, Kirsi Auro43, Keith Usiskin41, Amanda Elliott1,3, Joel Rämö1, Samuli Ripatti1, Mary Pat Reeve1 & Sanni Ruotsalainen1

Oncology Group Tuomo Meretoja63, Heikki Joensuu63, Olli Carpén63, Lauri Aaltonen63, Johanna Mattson63, Annika Auranen62, Peeter Karihtala61, Salla Kauppila61, Päivi Auvinen60, Klaus Elenius64, Johanna Schleutker64, Relja Popovic38, Jeffrey Waring38, Bridget Riley-Gillis38, Anne Lehtonen38, Jennifer Schutzman42, Julie Hunkapiller42, Natalie Bowers42, Sarah Pendergrass42, Andrey Loboda44, Aparna Chhibber44, Heli Lehtonen45, Stefan McDonough45, Marika Crohns46, Sauli Vuoti46, Diptee Kulkarni43, Kirsi Auro43, Esa Pitkänen1, Nina Mars1 & Mark Daly1

Ophthalmology Group Kai Kaarniranta60, Joni A. Turunen63, Terhi Olli63, Sanna Seitsonen63, Hannu Uusitalo62, Vesa Aaltonen64, Hannele Uusitalo-Järven62, Marja Luodonpää61, Nina Hautala61, Mengzhen Liu38, Heiko Runz40, Stephanie Loomis40, Erich Strauss42, Natalie Bowers42, Hao Chen42, Sarah Pendergrass42, Anna Podgornaia44, Juha Karjalainen1,3 & Esa Pitkänen1

Dermatology Group Kaisa Tasanen61, Laura Huilaja61, Katarina Hannula-Jouppi63, Teea Salmi62, Sirkku Peltonen64, Leena Koulu64, Kirsi Kalpala45, Ying Wu45, David Choy42, Sarah Pendergrass42, Nizar Smaoui38, Fedik Rahimov38, Anne Lehtonen38 & Dawn Waterworth48

Odontology Group Pirkko Pussinen63, Aino Salminen63, Tuula Salo63, David Rice63, Pekka Nieminen63, Ulla Palotie63, Juha Sinisalo63, Maria Siponen60, Liisa Suominen60, Päivi Mäntylä60, Ulvi Gursoy64, Vuokko Anttonen61, Kirsi Sipilä61 & Sarah Pendergrass42
Sample Collection Coordination  Anu Loukola

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

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

Sequencing Informatics  Priti Palta, Kalle Pärn, Shuang Luo & Vishal Sinha

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

Data Protection Officer  Tero Jyrhämä

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

38 Abbvie, Chicago, IL, USA. 39 AstraZeneca, 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 GlaxoSmit...