

Genome-wide analyses of early-onset acute myocardial infarction identify 29 novel loci by whole genome sequencing

Yeonsu Jeon^{1,2†}, Sungwon Jeon^{3†}, Whan-Hyuk Choi^{1,2}, Kyungwhan An^{1,2}, HanSol Choi^{1,2}, Byoung-Chul Kim³, Weon Kim⁴, Sang Yeob Lee⁵, Jang-Whan Bae⁵, Jin-Yong Hwang⁶, Min Gyu Kang⁶, Yeonkyung Kim³, Younghui Kang¹, Yeo Jin Kim³, Byung Chul Kim³, Jong Bhak^{1,2,3,7*}, and Eun-Seok Shin^{7,8*}

¹ Korean Genomics Center (KOGIC), Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea

² Department of Biomedical Engineering, College of Information-Bio Convergence Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea

³ Clinomics Inc., Ulsan 44919, Republic of Korea

⁴ Division of Cardiology, Department of Internal Medicine, Kyung Hee University Hospital, Kyung Hee University, Seoul 02447, Republic of Korea

⁵ Department of Internal Medicine, Chungbuk National University Hospital, College of Medicine, Chungbuk National University, Cheongju 28644, Republic of Korea

⁶ Department of Internal Medicine, Gyeongsang National University School of Medicine and Gyeongsang National University Hospital, Jinju 52727, Republic of Korea

⁷ Personal Genomics Institute (PGI), Genome Research Foundation (GRF), Osong, 28160, Republic of Korea

⁸ Department of Cardiology, Ulsan University Hospital, University of Ulsan College of Medicine, Ulsan, 44033, Republic of Korea

† These authors contributed equally to this work.

*** Correspondence to:**

Jong Bhak, PhD, Department of Biomedical Engineering, College of Information-Bio Convergence Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea, Email: jongbhak@genomics.org, Tel:+82-10-4644-6754, Fax:+82-52-217-3229

Eun-Seok Shin, MD, PhD, Department of Cardiology, Ulsan University Hospital, University of Ulsan College of Medicine, Ulsan, 44033, Republic of Korea, Email: sesim1989@gmail.com, Tel:+82-10-6319-4025, Fax:+82-52-250-7058

1 **Abstract**

2 Early-onset acute myocardial infarction (AMI) may have a higher genetic predisposition than
3 late-onset AMI does. The present study aimed to identify and characterize germline variants
4 that affect early-onset AMI using whole-genome sequencing (WGS). We performed a genome-
5 wide association study based on WGS of 1,239 Koreans, including 596 early-onset AMI
6 patients and 643 healthy individuals. Patients with AMI who underwent percutaneous coronary
7 intervention (PCI) caused by atherothrombotic occlusive lesions were included in the study. A
8 total of 29 novel loci were found to be associated with early-onset AMI. These loci are involved
9 in thrombosis, fibrinolysis, inflammation, and lipid metabolism. One of the associated single
10 nucleotide variants (SNVs), rs1614576, located upstream of *PRKCB*, is known to be associated
11 with thrombus formation. Additionally, the results revealed a novel locus, rs78631167, located
12 upstream of *PLAUR* which plays a critical role in regulating plasminogen activation and is
13 related to fibrinolysis. The association between early-onset AMI and rs9357455, which is
14 located upstream of *PHACTR1* that regulates inflammation in AMI, was found. Moreover, we
15 could identify a lipid metabolism related genetic risk locus, rs5072, in the *APOA1-AS* gene.
16 This study provides new evidence supporting the genetic association between early-onset AMI
17 and thrombosis and fibrinolysis, as well as inflammation and lipid metabolism, by analyzing
18 the whole-genome of 596 patients with early-onset AMI who have been treated with PCI. Our
19 findings highlight potential genetic markers for the prediction and management of AMI, as
20 well as for understanding the etiology of AMI.

21

22

1 INTRODUCTION

2 Acute myocardial infarction (AMI) is a leading cause of death with more than one million
3 deaths per year worldwide (1). The majority of AMI cases occur in individuals >65 years of
4 age. Early-onset AMI is rare with a prevalence of 1-5%, and genetic markers discovered from
5 them are probably more strongly associated than the late-onset counterpart (2-4). Until now,
6 many AMI studies have found loci associated with lipid and cholesterol metabolism which is
7 associated with atherosclerotic plaque development and progression (2, 5-8). Atherosclerotic
8 plaque causes coronary artery disease; however, a thrombus consisting of platelet aggregates
9 and fibrin strands ultimately leads to the clinical manifestations of AMI (2, 5, 9, 10). However,
10 the genetic factors that contribute to atherothrombotic coronary occlusion leading to AMI are
11 not clearly understood (11). In addition, recent clinical studies have shown that the occurrence
12 of AMI may be caused by an imbalance between thrombosis and fibrinolysis, a conflicting
13 mechanism (12). Fibrinolysis is involved in plaque healing and has emerged as a critical
14 mechanism of AMI (12). However, few studies have been conducted on the genetic factors
15 contributing to AMI by fibrinolysis (13).

16 The present study aimed to identify novel genomic loci associated with atherothrombotic
17 coronary occlusion, leading to the clinical manifestations of AMI. We report a genome-wide
18 association study (GWAS) based on whole-genome sequencing (WGS) in patients with early-
19 onset AMI who were eventually treated with percutaneous coronary intervention (PCI), and
20 patients with non-obstructive AMI were excluded.

21

22 RESULTS

1 **Novel Genomic Loci Associated with Early-Onset AMI**

2 Twenty-nine novel genomic loci (85 suggestive variants) for early-onset AMI were identified
3 using WGS-based GWAS over the genome-wide suggestive threshold (P -value cutoff = $1 \times$
4 10^{-5} ; Fig. 1, Table 1, 2, and Supplementary Material, Fig. S1, S2, Table S1). Approximately
5 86% of the novel loci were not covered in SNP chips such as Illumina human exome beadchip,
6 Humanhap 610 Quad Chip, and Affymetrix 6.0 GeneChip (Supplementary Material, Table S2
7 and S3) (2, 5, 14). The 29 novel loci included loci that are known to be related to thrombosis
8 and fibrinolysis. Locus rs1614576, located upstream of protein kinase C beta (*PRKCB*) gene,
9 had a suggestive association with early-onset AMI (OR = 1.646, $P = 3.41 \times 10^{-6}$; Fig. 2A). It
10 is involved in platelet responses implicated in thrombus formation (15). Another novel locus,
11 rs78631167 (OR = 0.473, $P = 2.93 \times 10^{-6}$) located upstream of the *PLAUR* (plasminogen
12 activator, urokinase receptor), was identified (Fig. 2B), and it is involved in fibrinolysis (16).
13 This locus was predicted to be located in a regulatory region with a RegulomeDB rank of 4
14 (17), representing transcription factor (TF) binding and DNase peak site. *PLAUR* encodes
15 CD87, which binds pro-urokinase to the cell surface and converts plasminogen to plasmin,
16 resulting in clot lysis (18). CD87 insufficiency results in impaired fibrinolysis and reduced clot
17 dissolution (18). In addition, a novel locus related to inflammation, rs9357455, located
18 upstream of the *PHACTR1* (phosphatase and actin regulator 1) was also identified, suggesting
19 an association with early-onset AMI (OR = 0.666, $P = 9.73 \times 10^{-6}$; Fig. 2C). This locus was
20 predicted to have regulatory functions, with a RegulomeDB rank of 5 (17), including regulation
21 of oxidative stress and inflammation in coronary artery endothelial cells via interactions with
22 NF- κ B/p65 (19). Previous myocardial infarction studies have reported that variants, such as
23 rs12526453 and rs9349379, in the *PHACTR1* gene are associated with myocardial infarction
24 (MI) (Supplementary Material, Table S4) (2, 20). The results of this study revealed an

1 association between early-onset AMI and a novel variant called rs9357455 in the *PHACTR1*
2 gene (Fig. 2C). Moreover, rs5072, a genetic risk locus for early-onset AMI (OR = 1.545, $P =$
3 8.61×10^{-6}) located in the intron of the *APOA1-AS* (apolipoprotein A1 antisense RNA) gene
4 which is involved in lipid metabolism, was also identified (Fig. 2D). It regulates the expression
5 of *APOA1* whose protein is a component of HDL cholesterol and a major participant in the
6 regulation of reverse cholesterol transport from peripheral tissues to the liver (21). Protein-
7 protein interaction (PPI) analysis revealed that dissolution of fibrin clot (HSA-75205),
8 complement and coagulation cascades (hsa04610), and cholesterol metabolism (hsa04979)
9 pathways were enriched from the genes on which the 29 novel loci are located (Supplementary
10 Material, Fig. S3 and Table S5-S7).

11

12 **Quantitative Trait Loci Mapping of the Novel Loci**

13 Fourteen of the twenty-nine lead variants tended to have larger effect sizes in this study
14 compared to the European (UK biobank; <http://www.nealelab.is/uk-biobank/>) and Japanese MI
15 (Biobank of Japan) studies (20) (Table 2 and Supplementary Material, Fig. S4 and S5). The
16 rest of the lead variants were not reported in previous SNP chip-based studies, implying that
17 only limited genomic regions had been analyzed (22) and that WGS-based GWAS can identify
18 novel loci using wider coverage of target genomic loci compared to SNP chip-based GWAS.
19 Among the 29 novel loci for early-onset AMI, five loci appeared as eQTL based on the GTEx
20 database (Fig. 3A and Supplementary Material, Fig. S6) (23, 24). Rs1614576 located upstream
21 of the *PRKCB* had the most significant eQTL signals in whole blood (eQTL Bonferroni
22 adjusted $P = 4 \times 10^{-207}$) by the eQTLGen database and artery tibial ($P = 2.75 \times 10^{-14}$; Fig. 3B)
23 and artery aorta ($P = 1.22 \times 10^{-9}$) in the GTEx database (Supplementary Material, Table S8).
24 The *PRKCB* gene affected by the rs1614576 loci is known to be involved in platelet responses

1 implicated in thrombus formation (15). Rs591012 located in *TTK* gene was not only an eQTL
2 ($P = 4.34 \times 10^{-17}$ in artery tibial; Fig. 3C), which affects the expression of the *ELOVL4* (ELOVL
3 fatty acid elongase 4), involved in lipid metabolism, but also an sQTL that changes the
4 alternative splicing of the gene (Supplementary Material, Table S8 and S9). Rs5072, located
5 in *APOAI-AS*, exhibited eQTL signals that changed the expression of *PAFAH1B2* (platelet
6 activating factor acetylhydrolase 1b catalytic subunit 2) ($P = 1.68 \times 10^{-4}$; Fig. 3D) and *PCSK7*
7 (proprotein convertase subtilisin/kexin type 7) ($P = 3.96 \times 10^{-11}$) in cells cultured fibroblasts in
8 the GTEx database (Supplementary Material, Table S8). In particular, *PAFAH1B2* encodes a
9 subunit of PAFAH that degrades and inactivates platelet-activating factor (PAF), a potent
10 platelet agonist (25). In addition, rs10232190 affects the expression of *LINC01003* ($P = 1.00 \times$
11 10^{-6} and 1.16×10^{-6} in the heart atrial appendage and whole blood, respectively; Fig. 3E and
12 Supplementary Material, Table S8). *LINC01003* can alter the expression of *PIMI* (proto-
13 oncogene serine/threonine-protein kinase Pim-1) whose inhibition suppresses vascular smooth
14 muscle cell (VSMC) proliferation and the development of atherosclerosis (26, 27).

15

16 **DISCUSSION**

17 This study aimed to identify new genomic loci associated with early-onset AMI and determine
18 their functional mechanisms using GWAS, QTL mapping, and protein-protein interaction
19 analyses. The results revealed 29 novel loci associated with early-onset AMI, of which two loci
20 are related to thrombosis and fibrinolysis, and confirmed the role of inflammation and lipid
21 metabolism in the development of early-onset AMI.

22 Consistent with a previous study showing that lipid metabolism-associated genes such as *LDLR*
23 and *PCSK9* are critical for early-onset AMI (2), the present study also identified a novel lipid

1 metabolism associated locus in *APOAI-AS* that regulates the expression of *APOAI* gene which
2 encodes a component of HDL cholesterol and protects against atherosclerosis (21). The eQTL
3 analysis based on the GTEx dataset showed that the expression of the *ELOVL4* (ELOVL fatty
4 acid elongase 4) gene, which is involved in lipid metabolism, was affected by the novel early-
5 onset AMI risk loci identified in this study. *ELOVL4* encodes a crucial enzyme that plays a role
6 in the production of very-long-chain fatty acids (VLC-SFA and VLC-PUFA) (28).

7 Few genetic markers associated with thrombus formation and fibrinolysis in AMI have been
8 discovered (5, 13). In 2021, Hartiala *et al.* discovered a thrombosis associated locus located in
9 the *SLC44A3* (Solute Carrier Family 44 Member 3) gene using GWAS for AMI (5). XU *et al.*
10 suggested that fibrinolysis related genes that encodes urokinase-type plasminogen activator
11 (uPA) which might influence the susceptibility to AMI; however, the study did not scan whole-
12 genome (13). Additionally, several studies have reported loci that are mainly associated with
13 lipid and cholesterol metabolism involving in plaque development and progression (6, 8, 29);
14 however, the results of the present study reveal novel loci related to thrombosis and fibrinolysis,
15 which are the leading cause of clinical manifestation of AMI. This result was supported by the
16 PPI analysis of the 29 novel loci, showing that the interacting genes belong to fibrinolysis
17 related pathways, such as dissolution of fibrin clot (HSA-75205) and complement and
18 coagulation cascades (hsa04610) (Supplementary Material, Table S5-7). Previous clinical
19 studies have focused mostly on the role of thrombosis in AMI formation (30). The spontaneous
20 lysis of a coronary thrombus, which is a natural protective mechanism against lasting occlusion
21 and myocardial infarction, is an important mechanism in the development of AMI (31).
22 Although arterial thrombosis is responsible for most AMI cases, endogenous fibrinolysis is
23 also a powerful defense mechanism against lasting arterial thrombotic occlusion (12). Recent
24 clinical studies have reported that impaired endogenous fibrinolysis can be detected in a

1 significant number of patients with the acute coronary syndrome, and this appears to be a novel
2 risk factor that is independent of traditional cardiovascular risk factors and unaffected by
3 antiplatelet therapy (12, 16, 32-34). Since only patients with early-onset AMI with definite
4 AMI phenotypes were included in the study, the present study implies that patients with early-
5 onset AMI may have susceptible genomic loci in the thrombosis related *PRKCB* gene and
6 protective loci in the fibrinolysis related *PLAUR* gene, suggesting the association of these
7 mechanisms with early-onset AMI.

8 Our study had three key strengths: clearly defined sample collection, whole-genome coverage,
9 and strict removal of batch effects. This study collected early-onset AMI samples that
10 underwent PCI for atherothrombotic occlusive lesions to prevent misclassification of AMI
11 phenotype. Notably, AMI samples with non-obstructive coronary arteries were excluded to
12 identify the cause of atherothrombosis, which has allowed us to acquire a considerable number
13 of newly detected loci in our GWAS, despite the sample size. Additionally, previous studies
14 on AMI utilized SNP chips such as Illumina human exome beadchip, Humanhap 610 Quad
15 Chip, or Affymetrix 6.0 GeneChip (2, 5, 14) and only limited genomic regions could be covered
16 by the SNP chip-based experiments (22). In 2019, Amit *et al.* studied the whole-genome of
17 2,081 patients with early-onset AMI; however, their study focused on analyzing the polygenic
18 risk score rather than determining early-onset AMI-specific markers (35). In the present study,
19 WGS data were analyzed to characterize the entire genome to identify early-onset AMI-
20 specific markers, revealing 86% loci and 77% variants associated with early AMI that were not
21 covered in SNP chips (Supplementary Material, Table S2 and S3). Moreover, the variants were
22 strictly filtered to remove false positives caused by the batch effect which can lead to false
23 disease association (36, 37). A batch effect was confirmed according to the sequencing year,
24 and it was not successfully filtered by common variant filtering criteria such as missing

1 genotype rate, minor allele frequency, and Hardy-Weinberg equilibrium (Supplementary
2 Material, Fig. S7). The batch effect was removed by filtering the variants with allele balance
3 bias, which identified systematic genotyping errors (36) (Supplementary Material, Fig. S8 and
4 S9).

5 This study had two potential limitations. First, the sample size of this study was only one-third
6 of that in the previous study by Amit *et al.* (35); however, our samples were clinically approved
7 samples requiring PCI for atherothrombotic AMI, and our genetic analysis covered the whole
8 genome. Second, we analyzed early-onset AMI in the Korean population. Therefore, the results
9 need to be validated in multiracial populations. Nevertheless, the genomic loci identified in this
10 study have four critical functions: thrombosis, fibrinolysis, inflammation, and lipid metabolism,
11 which are known to be related to AMI. Therefore, these loci are generally applicable to other
12 ethnic groups. We aim to verify the candidate loci using additional samples in functional
13 studies with additional multi-omics data, including RNA and protein expression levels and
14 epigenetic markers. These multi-layer omics data can provide researchers with detailed and
15 conclusive understanding on AMI development.

16 In conclusion, this study identified 29 novel loci that are involved in all the steps of AMI
17 development, including thrombosis and fibrinolysis, as well as previously known metabolic
18 pathways, such as inflammation and lipid metabolism, using WGS data accompanied by
19 GWAS, PPI, and QTL mapping. These results provide novel insights into the etiology of AMI.
20 These novel genetic markers are potential biomarkers for the prediction, prevention, and
21 personalized treatment of AMI in individuals who carry the loci.

22

23 **MATERIALS AND METHODS**

24 **Data Sources and Study Population**

1 In this study, 596 patients with early-onset AMI from four teaching hospitals in Korea were
2 enrolled and 643 healthy subjects from the Korean Genome Project (KGP) were selected as
3 control (38, 39). Patients with early-onset AMI were hospitalized with a diagnosis and
4 treatment of an ST-segment elevation myocardial infarction (STEMI) or non-ST-segment
5 elevation myocardial infarction (NSTEMI) caused by atherothrombotic occlusive lesions
6 which were treated with PCI (40). Patients with early-onset AMI included in this study were
7 men and women under 50 and 60 years of age, respectively. Population-based control subjects
8 were derived from the KGP (38, 39). KGP is the largest Korean Genome Project and currently
9 includes about 4,000 human genomes sequenced in Korea. Information regarding the KGP data
10 can be found on the Korean Genome Project web page (<http://koreangenome.org>). The
11 genomes of these 596 patients were compared with those of 643 control subjects in Korea, who
12 were free of cardiovascular disease, between 2018 and 2019. All study participants provided
13 written informed consent. Sample collection and sequencing were approved by the institutional
14 review board (IRB) of the Ulsan National Institute of Science and Technology (UNISTIRB-
15 15-19-A). Analyses were performed using Python version 3.7.7 and R version 3.5.0 (41, 42).

16

17 **Whole-genome sequencing by Illumina NovaSeq sequencer**

18 WGS was performed using the Illumina NovaSeq 6000 platform and clinical information from
19 the KGP was matched (38, 39). Of the 596 patient samples, 265 had already been sequenced
20 in a previous study (40), and the remaining 331 samples were sequenced for this study (Table
21 1). Genomic DNA was isolated from the plate using a DNeasy Blood & Tissue kit (Qiagen,
22 Germany) according to the manufacturer's protocol. The extracted DNA was quantified using
23 a Quant-iT BR assay kit (Invitrogen, USA). High-molecular-weight genomic DNA was
24 sheared using a Covaris S2 ultrasonicator system to obtain fragments of appropriate sizes.

1 Libraries with short 350–base pair (bp) inserts for paired-end reads were prepared using the
2 TruSeq Nano DNA sample prep kit in accordance with the manufacturer’s protocol for
3 Illumina-based sequencing. The products were quantified using a Bioanalyzer 2100 (Agilent,
4 Santa Clara, CA, USA) and raw data were generated using an Illumina NovaSeq 6000 platform
5 (Illumina). Clusters were generated using paired-end 2×150 -bp cycle sequencing reads via
6 resequencing. The quality of the bases in the reads was checked using FastQC (ver. 0.11.5;
7 www.bioinformatics.babraham.ac.uk/projects/fastqc/) software. Additional details regarding
8 genomic variant identification are provided in the Supplementary Methods in the
9 Supplementary Material.

10

11 **Genome-Wide Association Study (GWAS)**

12 GWAS was performed using logistic regression with an additive genetic model using PLINK
13 (ver. 1.9b) (43). Additional details regarding batch effect removal and individual and variant
14 filtering are provided in the Supplementary Methods in the Supplement Material. A total of
15 7,869,081 SNPs and indels were tested. Sex and the top ten principal components were
16 included in the model as covariates. The genome-wide suggestive threshold was determined to
17 be 1×10^{-5} for suggestive variants. Suggestive variants were considered novel according to the
18 following criteria: (i) single nucleotide variants (SNVs) with no overlap with previously
19 reported loci for myocardial infarction and (ii) SNVs with lead variants that do not have LD
20 variants ($r^2 > 0.1$) in previously reported loci for myocardial infarction. Variants were grouped
21 into the loci with the ‘--clump-r2 0.1 --clump-window-size 250’ option with PLINK (ver. 1.9b)
22 (43).

23

1 **Quantitative Trait Loci (QTL) Mapping**

2 Variants were mapped to expression QTL (eQTL), splicing QTL (sQTL), and methylation QTL
3 (mQTL) using publicly available datasets. For eQTL, which explains a fraction of the genetic
4 variance of a gene expression phenotype (44), we mapped the lead variants from the GWAS to
5 the eQTL from GTEx and eQTLGen dataset (23, 24). For sQTL, sQTL was determined using
6 the GTEx dataset (23). For mQTL, mQTLdb, a large-scale genome-wide DNA methylation
7 analysis of 1,000 mother-child pairs at series time points across the life-course (45), was used.
8 QTLs for suggestive lead variants were reported.

9

10 **DATA AVAILABILITY**

11 Full summary statistics relating to the GWAS analysis will be deposited on the Korean Genome
12 Project web page (<http://koreangenome.org/Cardiomics>). All other relevant data are available
13 upon request from the authors.

14

15 **FUNDING**

16 This work was supported by the Promotion of Innovative Businesses for Regulation-Free
17 Special Zones funded by the Ministry of SMEs and Startups (MSS, Korea) [grant number
18 P0016195, P0016193 (1425156792, 1425157301) (2.220035.01, 2.220036.01)]. This work
19 was also supported by the Establishment of Demonstration Infrastructure for Regulation-Free
20 Special Zones fund (MSS, Korea) [grant number P0016191 (2.220037.01) (1425157253)] by
21 the Ministry of SMEs and Startups.

22

1

2 **ACKNOWLEDGEMENTS**

3 We thank all the genome donating participants. The biospecimens for this study were provided
4 by Ulsan Medical Center and the Biobanks of Gyeongsang National University Hospital,
5 Chungbuk National University Hospital (18-27, 20-04), and Kyung Hee University Hospital
6 (2018-4, 2019-4, 2019-6), the members of the National Biobank of Korea, which is supported
7 by the Ministry of Health, Welfare and Family Affairs. We thank the Korea Institute of Science
8 and Technology Information (KISTI) which provided us with the Korea Research Environment
9 Open NETwork (KREONET).

10 The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the
11 Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA,
12 NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained
13 from the GTEx Portal on 11/25/2021.

14

15 **CONFLICT OF INTEREST**

16 SJ, BK, YK, and YJK were employed by Clinomics Inc. BCK and JB are the CEOs of
17 Clinomics Inc. The remaining authors declare that the research was conducted in the absence
18 of any commercial or financial relationships that could be constructed as a potential conflict of
19 interest.

20

21

1 REFERENCES

- 2 1 Mechanic, O.J. and Grossman, S.A. (2017) Acute myocardial infarction. in press.
- 3 2 Consortium, M.I.G. (2009) Genome-wide association of early-onset myocardial
4 infarction with common single nucleotide polymorphisms, common copy number variants, and
5 rare copy number variants. *Nature genetics*, **41**, 334.
- 6 3 Members, W.G., Rosamond, W., Flegal, K., Furie, K., Go, A., Greenlund, K., Haase,
7 N., Hailpern, S.M., Ho, M. and Howard, V. (2008) Heart disease and stroke statistics—2008
8 update: a report from the American Heart Association Statistics Committee and Stroke
9 Statistics Subcommittee. *Circulation*, **117**, e25-e146.
- 10 4 Control, C.f.D. and Prevention. (2007) Prevalence of heart disease--United States,
11 2005. *MMWR. Morbidity and mortality weekly report*, **56**, 113-118.
- 12 5 Hartiala, J.A., Han, Y., Jia, Q., Hilser, J.R., Huang, P., Gukasyan, J., Schwartzman,
13 W.S., Cai, Z., Biswas, S. and Trégouët, D.-A. (2021) Genome-wide analysis identifies novel
14 susceptibility loci for myocardial infarction. *European Heart Journal*, **42**, 919-933.
- 15 6 Tajima, T., Morita, H., Ito, K., Yamazaki, T., Kubo, M., Komuro, I. and Momozawa,
16 Y. (2018) Blood lipid-related low-frequency variants in LDLR and PCSK9 are associated with
17 onset age and risk of myocardial infarction in Japanese. *Scientific reports*, **8**, 1-9.
- 18 7 Do, R., Stitzel, N.O., Won, H.-H., Jørgensen, A.B., Duga, S., Angelica Merlini, P.,
19 Kiezun, A., Farrall, M., Goel, A. and Zuk, O. (2015) Exome sequencing identifies rare LDLR
20 and APOA5 alleles conferring risk for myocardial infarction. *Nature*, **518**, 102-106.
- 21 8 Libby, P., Buring, J., Badimon, L. and Hansson, G. (2019) Dean eld J, Bittencourt MS,
22 Tokgözoğlu L, Lewis EF. *Atherosclerosis. Nat Rev Dis Primers*, **5**, 56.
- 23 9 Asada, Y., Yamashita, A., Sato, Y. and Hatakeyama, K. (2020) Pathophysiology of
24 atherothrombosis: Mechanisms of thrombus formation on disrupted atherosclerotic plaques.

- 1 *Pathology International*, **70**, 309-322.
- 2 10 Fuster, V., Moreno, P.R., Fayad, Z.A., Corti, R. and Badimon, J.J. (2005)
- 3 Atherothrombosis and high-risk plaque: part I: evolving concepts. *Journal of the American*
- 4 *College of Cardiology*, **46**, 937-954.
- 5 11 Palasubramaniam, J., Wang, X. and Peter, K. (2019) Myocardial Infarction—From
- 6 Atherosclerosis to Thrombosis: Uncovering New Diagnostic and Therapeutic Approaches.
- 7 *Arteriosclerosis, thrombosis, and vascular biology*, **39**, e176-e185.
- 8 12 Vergallo, R. and Crea, F. (2020) Atherosclerotic plaque healing. *New England Journal*
- 9 *of Medicine*, **383**, 846-857.
- 10 13 Xu, J., Li, W., Bao, X., Ding, H., Chen, J., Zhang, W., Sun, K., Wang, J., Wang, X.
- 11 and Wang, H. (2010) Association of putative functional variants in the PLAU gene and the
- 12 PLAUR gene with myocardial infarction. *Clinical science*, **119**, 353-359.
- 13 14 Han, Y., Dorajoo, R., Chang, X., Wang, L., Khor, C.-C., Sim, X., Cheng, C.-Y., Shi,
- 14 Y., Tham, Y.C. and Zhao, W. (2017) Genome-wide association study identifies a missense
- 15 variant at APOA5 for coronary artery disease in Multi-Ethnic Cohorts from Southeast Asia.
- 16 *Scientific reports*, **7**, 1-11.
- 17 15 Heemskerk, J.W., Harper, M.T., Cosemans, J.M. and Poole, A.W. (2011) Unravelling
- 18 the different functions of protein kinase C isoforms in platelets. *FEBS letters*, **585**, 1711-1716.
- 19 16 Gorog, D.A. and Lip, G.Y. (2019) Impaired spontaneous/endogenous fibrinolytic
- 20 status as new cardiovascular risk factor? JACC review topic of the week. *Journal of the*
- 21 *American College of Cardiology*, **74**, 1366-1375.
- 22 17 Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M.,
- 23 Karczewski, K.J., Park, J., Hitz, B.C. and Weng, S. (2012) Annotation of functional variation
- 24 in personal genomes using RegulomeDB. *Genome research*, **22**, 1790-1797.

- 1 18 Hill, A., Kelly, R.J. and Hillmen, P. (2013) Thrombosis in paroxysmal nocturnal
2 hemoglobinuria. *Blood, The Journal of the American Society of Hematology*, **121**, 4985-4996.
- 3 19 Zhang, Z., Jiang, F., Zeng, L., Wang, X. and Tu, S. (2018) PHACTR1 regulates
4 oxidative stress and inflammation to coronary artery endothelial cells via interaction with NF-
5 κ B/p65. *Atherosclerosis*, **278**, 180-189.
- 6 20 Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshihara, S., Narita,
7 A., Konuma, T., Yamamoto, K. and Akiyama, M. (2021) A cross-population atlas of genetic
8 associations for 220 human phenotypes. *Nature genetics*, **53**, 1415-1424.
- 9 21 Shioji, K., Mannami, T., Kokubo, Y., Goto, Y., Nonogi, H. and Iwai, N. (2004) An
10 association analysis between ApoA1 polymorphisms and the high-density lipoprotein (HDL)
11 cholesterol level and myocardial infarction (MI) in Japanese. *Journal of human genetics*, **49**,
12 433-439.
- 13 22 Dreger, D.L., Rimbault, M., Davis, B.W., Bhatnagar, A., Parker, H.G. and Ostrander,
14 E.A. (2016) Whole-genome sequence, SNP chips and pedigree structure: building
15 demographic profiles in domestic dog breeds to optimize genetic-trait mapping. *Disease*
16 *Models & Mechanisms*, **9**, 1445-1460.
- 17 23 Battle, A., Brown, C.D., Engelhardt, B.E. and Montgomery, S.B. (2017) Genetic
18 effects on gene expression across human tissues. *Nature*, **550**, 204-213.
- 19 24 Vösa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten,
20 H., Saha, A., Kreuzhuber, R. and Brugge, H. (2021) Large-scale cis-and trans-eQTL analyses
21 identify thousands of genetic loci and polygenic scores that regulate blood gene expression.
22 *Nature genetics*, **53**, 1300-1310.
- 23 25 Tselepis, A.D., Goudevenos, J.A., Tambaki, A.P., Michalis, L., Stroumbis, C.S.,
24 Tsoukatos, D.C., Elisaf, M. and Sideris, D.A. (1999) Platelet aggregatory response to platelet

- 1 activating factor (PAF), ex vivo, and PAF-acetylhydrolase activity in patients with unstable
2 angina: effect of c7E3 Fab (abciximab) therapy. *Cardiovascular research*, **43**, 183-191.
- 3 26 Wu, L., Xia, L., Chen, X., Ruan, M., Li, L. and Xia, R. (2021) Long non-coding RNA
4 LINC01003 suppresses the development of multiple myeloma by targeting miR-33a-5p/PIM1
5 axis. *Leukemia Research*, **106**, 106565.
- 6 27 KATAKAMI, N., KANETO, H., MIYATSUKA, T., KAWAMORI, D.,
7 GOROGAWA, S.-I., SAKAMOTO, K.Y., FUJITANI, Y., MATSUOKA, T.-A.,
8 Umayahara, Y. and Kajimoto, Y. (2004) A Novel Role of Proto-Oncogene Pim-1 in
9 Progression of Atherosclerosis. *Diabetes*, **53**.
- 10 28 Hopiavuori, B.R., Anderson, R.E. and Agbaga, M.-P. (2019) ELOVL4: Very long-
11 chain fatty acids serve an eclectic role in mammalian health and function. *Progress in retinal
12 and eye research*, **69**, 137-158.
- 13 29 Do, R., Stitzel, N.O., Won, H.-H., Jørgensen, A.B., Duga, S., Merlini, P.A., Kiezun,
14 A., Farrall, M., Goel, A. and Zuk, O. (2015) Multiple rare alleles at LDLR and APOA5 confer
15 risk for early-onset myocardial infarction. *Nature*, **518**, 102.
- 16 30 Schulz, C. and Massberg, S. (2012) Platelets in atherosclerosis and thrombosis.
17 *Antiplatelet Agents*, in press., 111-133.
- 18 31 Gorog, D.A., Fayad, Z.A. and Fuster, V. (2017) Arterial thrombus stability: does it
19 matter and can we detect it? *Journal of the American College of Cardiology*, **70**, 2036-2047.
- 20 32 Saraf, S., Christopoulos, C., Salha, I.B., Stott, D.J. and Gorog, D.A. (2010) Impaired
21 endogenous thrombolysis in acute coronary syndrome patients predicts cardiovascular death
22 and nonfatal myocardial infarction. *Journal of the American college of Cardiology*, **55**, 2107-
23 2115.
- 24 33 Okafor, O.N. and Gorog, D.A. (2015) Endogenous fibrinolysis: an important mediator

1 of thrombus formation and cardiovascular risk. *Journal of the American College of Cardiology*,
2 **65**, 1683-1699.

3 34 Farag, M., Spinhakis, N., Gue, Y.X., Srinivasan, M., Sullivan, K., Wellsted, D. and
4 Gorog, D.A. (2019) Impaired endogenous fibrinolysis in ST-segment elevation myocardial
5 infarction patients undergoing primary percutaneous coronary intervention is a predictor of
6 recurrent cardiovascular events: the RISK PPCI study. *European heart journal*, **40**, 295-305.

7 35 Khera, A.V., Chaffin, M., Zekavat, S.M., Collins, R.L., Roselli, C., Natarajan, P.,
8 Lichtman, J.H., D'Onofrio, G., Mattera, J. and Dreyer, R. (2019) Whole-genome sequencing
9 to characterize monogenic and polygenic contributions in patients hospitalized with early-onset
10 myocardial infarction. *Circulation*, **139**, 1593-1602.

11 36 Muyas, F., Bosio, M., Puig, A., Susak, H., Domènech, L., Escaramis, G., Zapata, L.,
12 Demidov, G., Estivill, X. and Rabionet, R. (2019) Allele balance bias identifies systematic
13 genotyping errors and false disease associations. *Human mutation*, **40**, 115-126.

14 37 Wickland, D.P., Ren, Y., Sinnwell, J.P., Reddy, J.S., Pottier, C., Sarangi, V.,
15 Carrasquillo, M.M., Ross, O.A., Younkin, S.G. and Ertekin-Taner, N. (2021) Impact of variant-
16 level batch effects on identification of genetic risk factors in large sequencing studies. *PLoS*
17 *one*, **16**, e0249305.

18 38 Jeon, S., Bhak, Y., Choi, Y., Jeon, Y., Kim, S., Jang, J., Jang, J., Blazyte, A., Kim, C.
19 and Kim, Y. (2020) Korean Genome Project: 1094 Korean personal genomes with clinical
20 information. *Science advances*, **6**, eaaz7835.

21 39 Jeon, Y., Jeon, S., Blazyte, A., Kim, Y.J., Lee, J.J., Bhak, Y., Cho, Y.S., Park, Y., Noh,
22 E.-K. and Manica, A. (2021) Welfare genome project: A participatory Korean personal genome
23 project with free health check-up and genetic report followed by counseling. *Frontiers in*
24 *genetics*, **12**.

1 40 Bhak, Y., Jeon, Y., Jeon, S., Yoon, C., Kim, M., Blazyte, A., Kim, Y., Kang, Y., Kim,
2 C. and Lee, S.Y. (2021) Polygenic risk score validation using Korean genomes of 265 early-
3 onset acute myocardial infarction patients and 636 healthy controls. *PloS one*, **16**, e0246538.

4 41 Van, R.G. and Drake, F. (2009) Python 3 reference manual. *Scotts Valley, CA:*
5 *CreateSpace*, **10**, 1593511.

6 42 Team, R.C. (2013) R: A language and environment for statistical computing. in press.

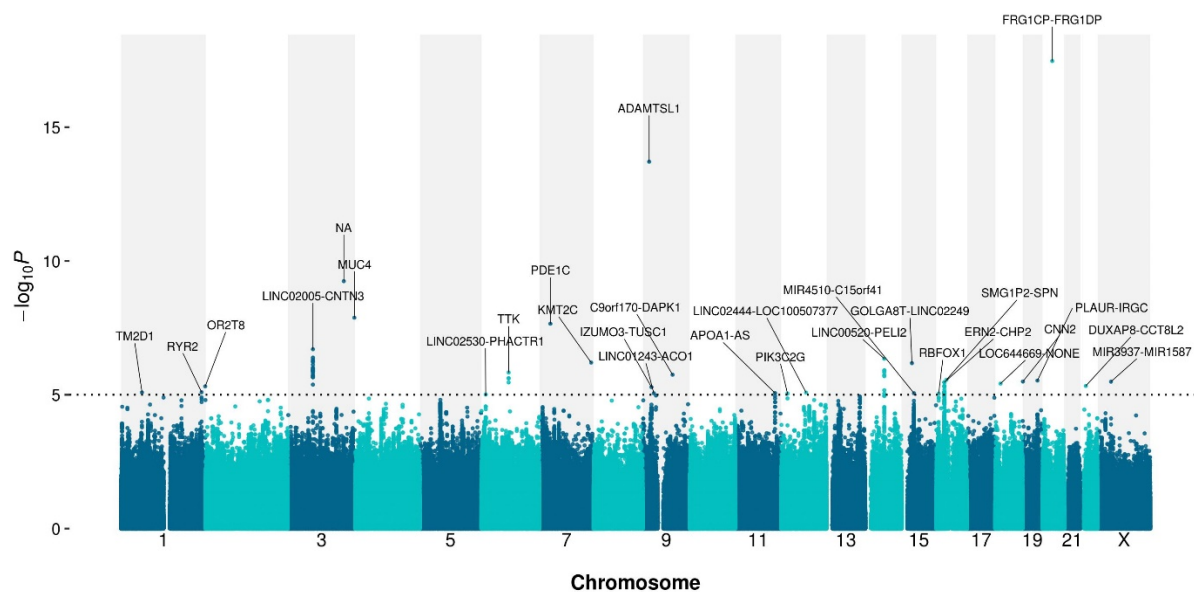
7 43 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller,
8 J., Sklar, P., De Bakker, P.I. and Daly, M.J. (2007) PLINK: a tool set for whole-genome
9 association and population-based linkage analyses. *The American journal of human genetics*,
10 **81**, 559-575.

11 44 Nica, A.C. and Dermitzakis, E.T. (2013) Expression quantitative trait loci: present and
12 future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**, 20120362.

13 45 Gaunt, T.R., Shihab, H.A., Hemani, G., Min, J.L., Woodward, G., Lyttleton, O., Zheng,
14 J., Duggirala, A., McArdle, W.L. and Ho, K. (2016) Systematic identification of genetic
15 influences on methylation across the human life course. *Genome biology*, **17**, 1-14.

16

1 FIGURES



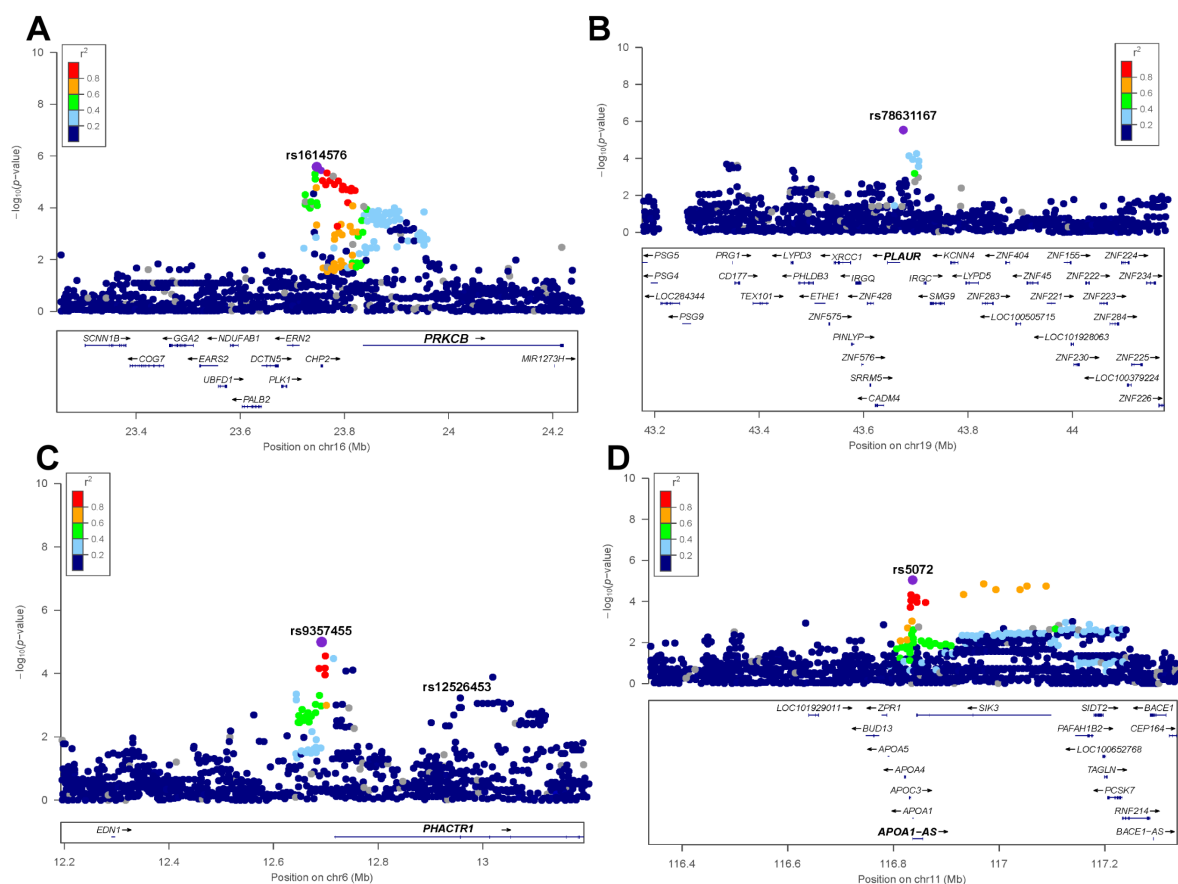
2

3 **Figure 1. Manhattan plot of the associated genomic loci for early-onset AMI via a GWAS**

4 X-axis indicates the genomic positions (hg38). Y-axis indicates $-\log_{10}P$ -value from the GWAS.

5 The dotted line indicates the genome-wide suggestive threshold (1×10^{-5}). The suggestive

6 genomic loci are presented as gene symbols.

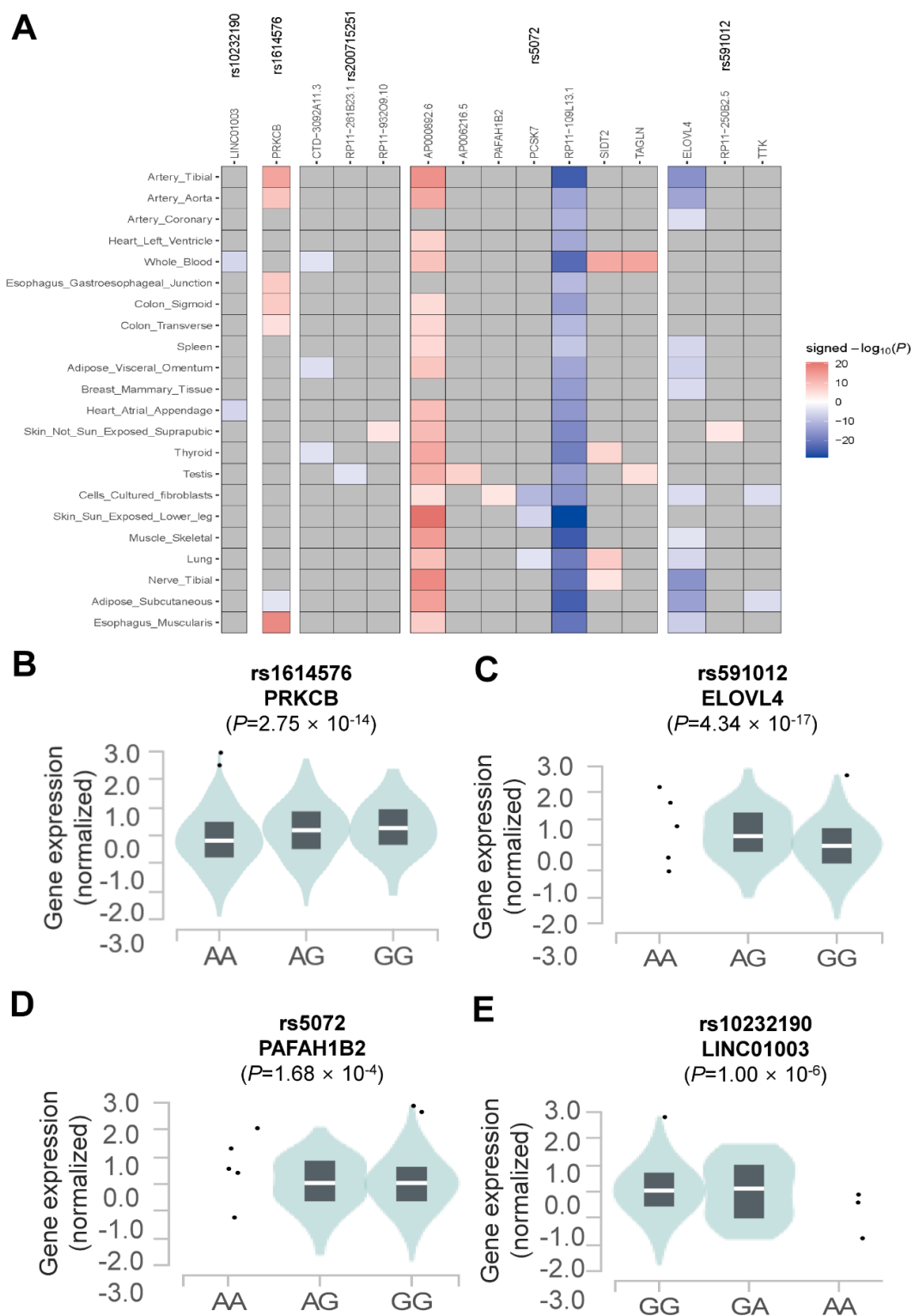


1

2 **Figure 2. Regional association plot of genomic loci associated with early-onset AMI**
 3 **related to thrombosis, fibrinolysis, inflammation, and lipid metabolism**

4 (A-D) Regional association plot at (A) rs1614576 (B) rs78631167 (C) rs9357455 (D) rs5072.
 5 For each plot, the x-axis indicates the genomic positions (hg38). Y-axis indicates $-\log_{10} P$ -value
 6 from the GWAS.

7



1

2 **Figure 3. eQTL for the novel loci associated with early-onset AMI**

3 (A) Heatmap of eQTL mapping to the novel loci for early-onset AMI. Rows show the tissue.

1 Columns show the novel loci and the genes affected by the loci. Colors represent the signed -
2 \log_{10} P -value from the eQTL results. (B-E) Violin plot of eQTL at (B) rs1614576 (C) rs591012
3 (D) rs5072 (E) rs10232190. The lower and upper border of the box correspond to the first and
4 third quartiles, respectively, the central line indicates the median.

5

6

1 **TABLES**

2 **Table 1. Baseline characteristics**

	Early-onset AMI (N=596)	Control (N=643)
Male, n (%)	486 (81.5)	328 (51.0)
Age, median (Q1 - Q3)	46 (42.0-48.0)	44 (29.5-57.0)
Body mass index, mean \pm SD	25.5 \pm 3.9	24.0 \pm 3.4
Hypertension, n (%)	188 (31.5)	90 (14.0)
Diabetes mellitus, n (%)	118 (19.8)	33 (5.1)
Current smoking, n (%)	360 (60.4)	82 (12.8)
Lipid levels, mg/dL		
Total cholesterol, mean \pm SD	202.7 \pm 49.2	179.6 \pm 34.3
LDL cholesterol, mean \pm SD	129.1 \pm 43.4	115.8 \pm 33.0
HDL cholesterol, mean \pm SD	43.9 \pm 13.0	57.4 \pm 13.9
Triglycerides, mean \pm SD	209.1 \pm 185.1	115.8 \pm 77.0

3 AMI, acute myocardial infarction; family history of CAD, 1st degree family history of coronary
4 artery disease; LDL, low-density lipoprotein; HDL, high-density lipoprotein. Values are mean
5 \pm SD, median (interquartile range, 25th - 75th), or n (%).

6

7

8

1 **Table 2. Novel genomic loci for early-onset AMI compared to European and Japanese MI**
 2 **studies**

Gene	Chromosome	Position	rs ID	REF	ALT	Effect allele	Korean Genome Project (early-onset AMI in Korean)		BBJ (MI in Japanese)		UK biobank (MI in European)		
							Odds ratio	P	BE TA	P	BE TA	P	
FRG1CP-FRG1DP	20	28,772,995	rs1277393322	G	A	A	5.713	3.36 E-18	NA	NA	NA	NA	
ADAMTSL1	9	18,523,845	rs10810978	C	A	C	0.451	1.94 E-14	NA	NA	NA	NA	
NA	3	164,704,630	rs1351282285	TTAAATG	T	T	0.239	5.67 E-10	NA	NA	NA	NA	
MUC4	3	195,788,126	rs1560389462	GGGTGGTGTGACCTGTGGATAC TGAGGAAGTGTCCGGTACAGG AAGAGAGGTGGCGTGACCTGT GGATGCTGAGGAAGTGTCCGGT ACAGGAAGAGGGTGGTGTGA CCTGTGGATACTGAGGAAGTGT CGGTGACAGGAAGAGA	G	G	G	0.256	1.31 E-08	NA	NA	NA	NA
PDE1C	7	32,082,944	rs10235820	A	G	A	0.497	2.21 E-08	NA	NA	NA	NA	
LINC02005-CNTN3	3	73,926,364	rs2323404	A	G	A	1.606	1.98 E-07	0.012 243	3.68 E-01	- 0.005 94	2.32 E-01	
LINC00520-PELI2	14	55,900,262	rs8009869	A	G	G	1.653	4.49 E-07	NA	NA	NA	NA	
KMT2C	7	152,231,944	rs10232190	G	A	A	1.980	6.24 E-07	- 0.019 416	3.96 E-01	NA	NA	
GOLGA8T-LINC02249	15	30,175,826	rs200715251	A	G	G	1.840	6.56 E-07	NA	NA	NA	NA	
TTK	6	80,038,900	rs591012	G	A	G	0.583	1.47 E-06	0.013 298	4.28 E-01	0.001 87	5.73 E-01	
C9orf170-DAPK1	9	87,414,203	rs200514416	T	C	C	1.948	1.78 E-06	NA	NA	NA	NA	
SMG1P2-SPN	16	29,655,837	rs372341820	G	A	A	0.180	2.75 E-06	NA	NA	NA	NA	
PLAUR-IRGC	19	43,676,115	rs78631167	T	C	C	0.473	2.93 E-06	NA	NA	0.003 27	8.02 E-01	
CNN2	19	1,029,328	rs527249040	G	C	C	0.292	3.20 E-06	NA	NA	0.011 11	7.53 E-01	
MIR3937-MIR1587	X	39,788,017	rs184610791	C	G	G	0.234	3.23 E-06	NA	NA	NA	NA	
ERN2-CHP2	16	23,747,804	rs1614576	A	G	G	1.646	3.41 E-06	0.003 445	8.44 E-01	0.000 26	9.44 E-01	
LOC644669-NONE	18	15,377,667	rs7232479	G	A	A	1.763	3.82 E-06	NA	NA	NA	NA	
DUXAP8-CCT8L2	22	16,400,343	rs131570	T	C	C	0.544	4.66 E-06	NA	NA	NA	NA	
OR2T8	1	247,921,552	rs4584426	A	G	G	1.538	4.83 E-06	NA	NA	NA	NA	

IZUMO3-TUSC1	9	25,189,022	rs147169123	A	AT	A	0.523	5.14 E-06	0.011 244	5.95 E-01	- 0.000 03	9.95 E-01
RYR2	1	237,214,855	rs2152884	G	A	A	1.530	8.02 E-06	- 0.015 869	2.55 E-01	0.0007 28	4.58 E-02
TM2D1	1	61,717,206	rs141073026	A	C	C	1.673	8.13 E-06	NA	NA	NA	NA
LINC02444 - LOC100507377	12	73,222,956	rs200261514	G	GTA AT	GT AA T	2.644	8.25 E-06	- 0.011 721	7.76 E-01	NA	NA
APOA1-AS	11	116,836,867	rs5072	A	G	A	1.545	8.61 E-06	0.011 949	3.82 E-01	- 0.000 46	9.41 E-01
MIR4510-C15orf41	15	36,119,067	rs12438548	G	A	G	0.609	8.72 E-06	0.004 081	8.14 E-01	- 0.0003 43	3.30 E-01
LINC01243-ACO1	9	32,154,607	rs199830494	G	A	G	0.631	8.77 E-06	NA	NA	NA	NA
PIK3C2G	12	18,556,791	rs117684941	T	C	C	0.249	8.95 E-06	- 0.008 694	8.35 E-01	NA	NA
RBFOX1	16	7,002,773	rs12921822	C	T	C	1.601	9.05 E-06	- 0.018 229	2.43 E-01	- 0.000 56	8.63 E-01
LINC02530-PHACTR1	6	12,693,654	rs9357455	C	T	C	0.666	9.73 E-06	0.012 942	3.34 E-01	0.0004 79	3.12 E-01

1 NA represents an unreported value.

2

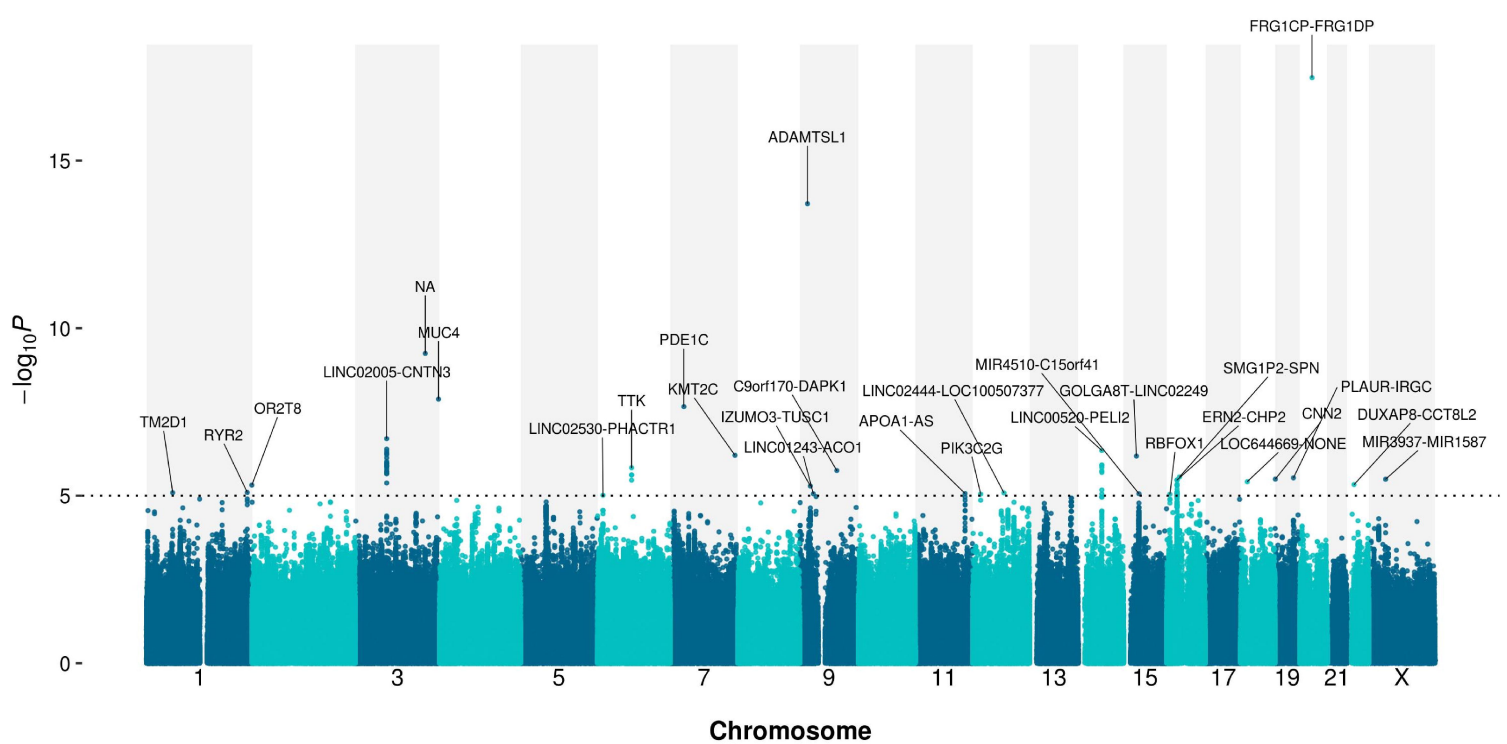
1 **ABBREVIATIONS**

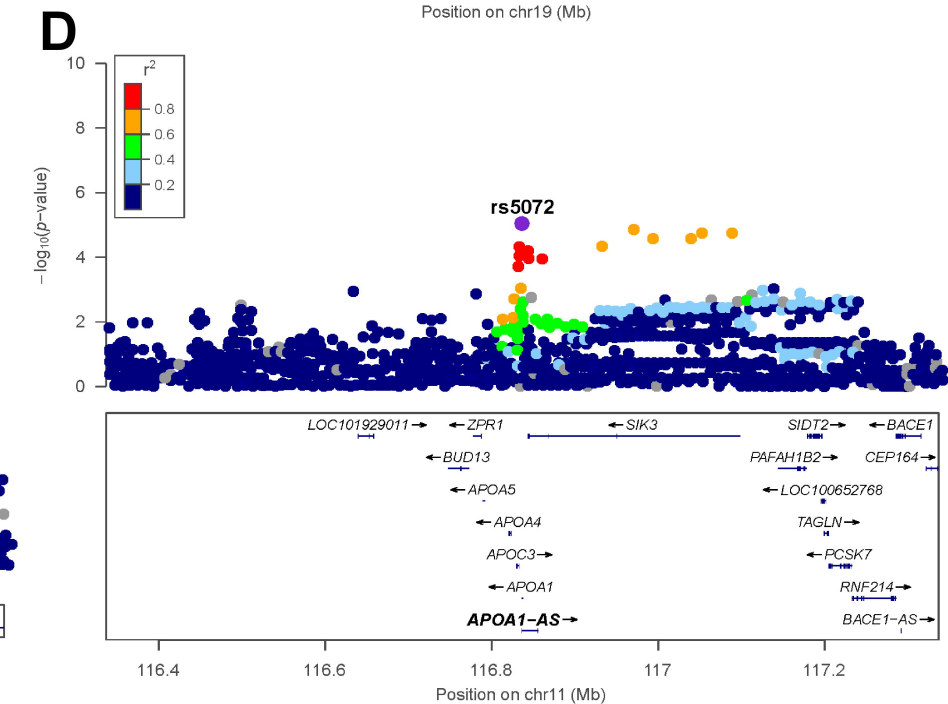
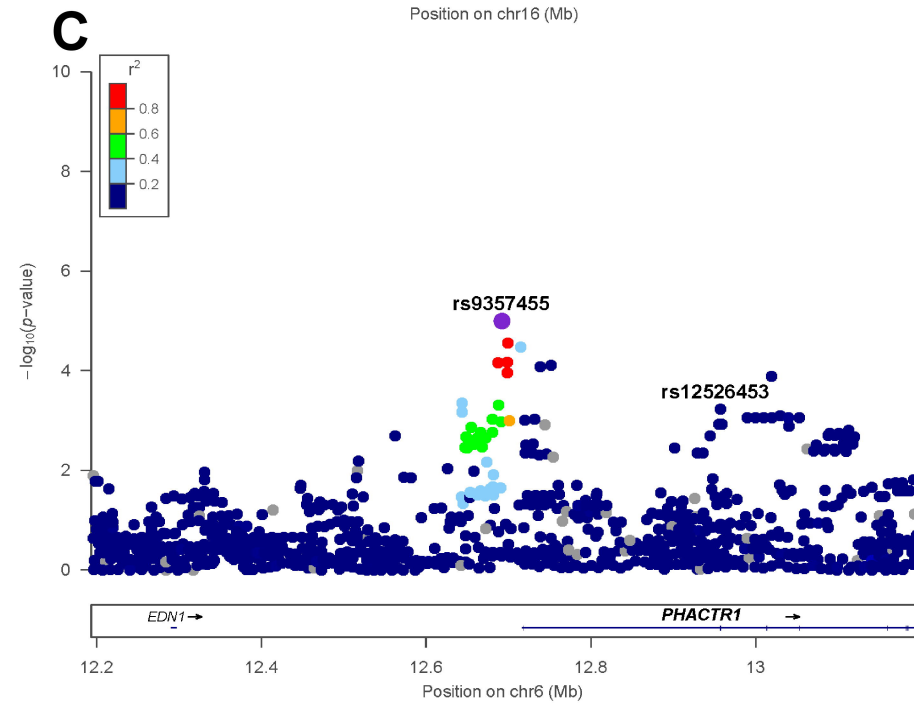
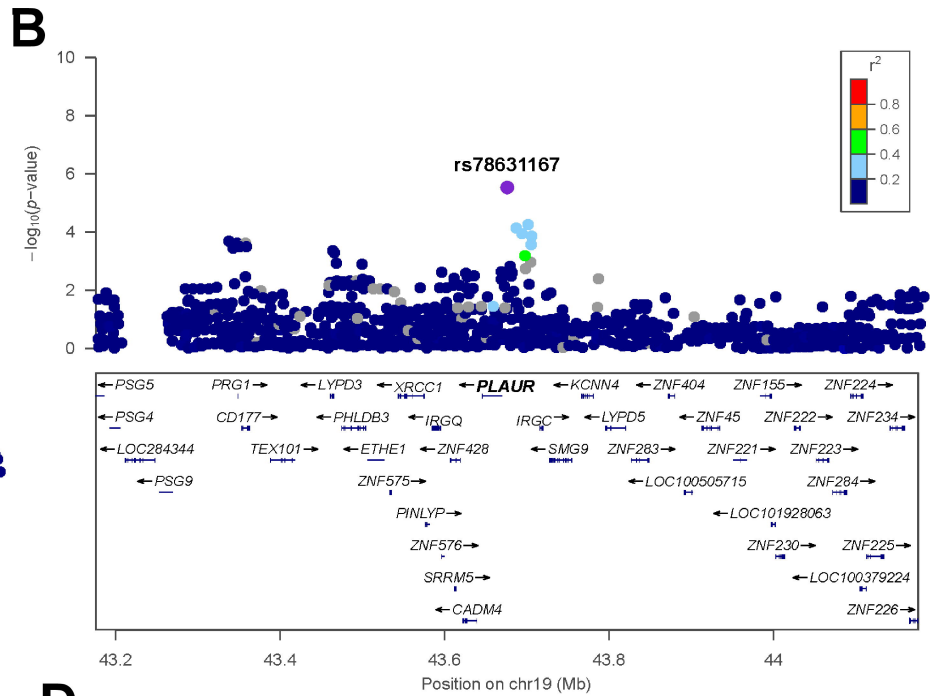
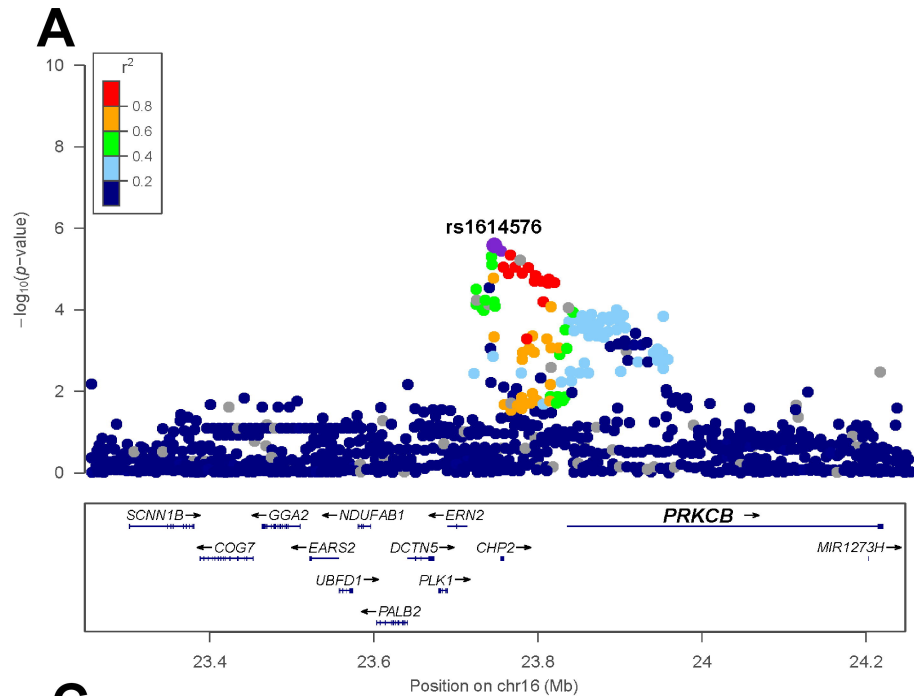
2 AMI, acute myocardial infarction; PCI, percutaneous coronary intervention; GWAS, genome-
3 wide association study; WGS, whole-genome sequencing, SNV, single nucleotide variant;
4 KGP, Korean Genome Project; STEMI, ST-segment elevation myocardial infarction; NSTEMI,
5 non-ST-segment elevation myocardial infarction; QTL, quantitative trait loci; eQTL,
6 expression QTL; sQTL, splicing QTL; mQTL, methylation QTL; PRKCB; protein kinase C
7 beta; PLAUR, plasminogen activator, urokinase receptor; PAFAH1B2, platelet-activating
8 factor acetylhydrolase 1b catalytic subunit 2; PHACTR1, phosphatase and actin regulator 1;
9 APOA1-AS, apolipoprotein A1 antisense RNA; PPI, protein-protein interaction

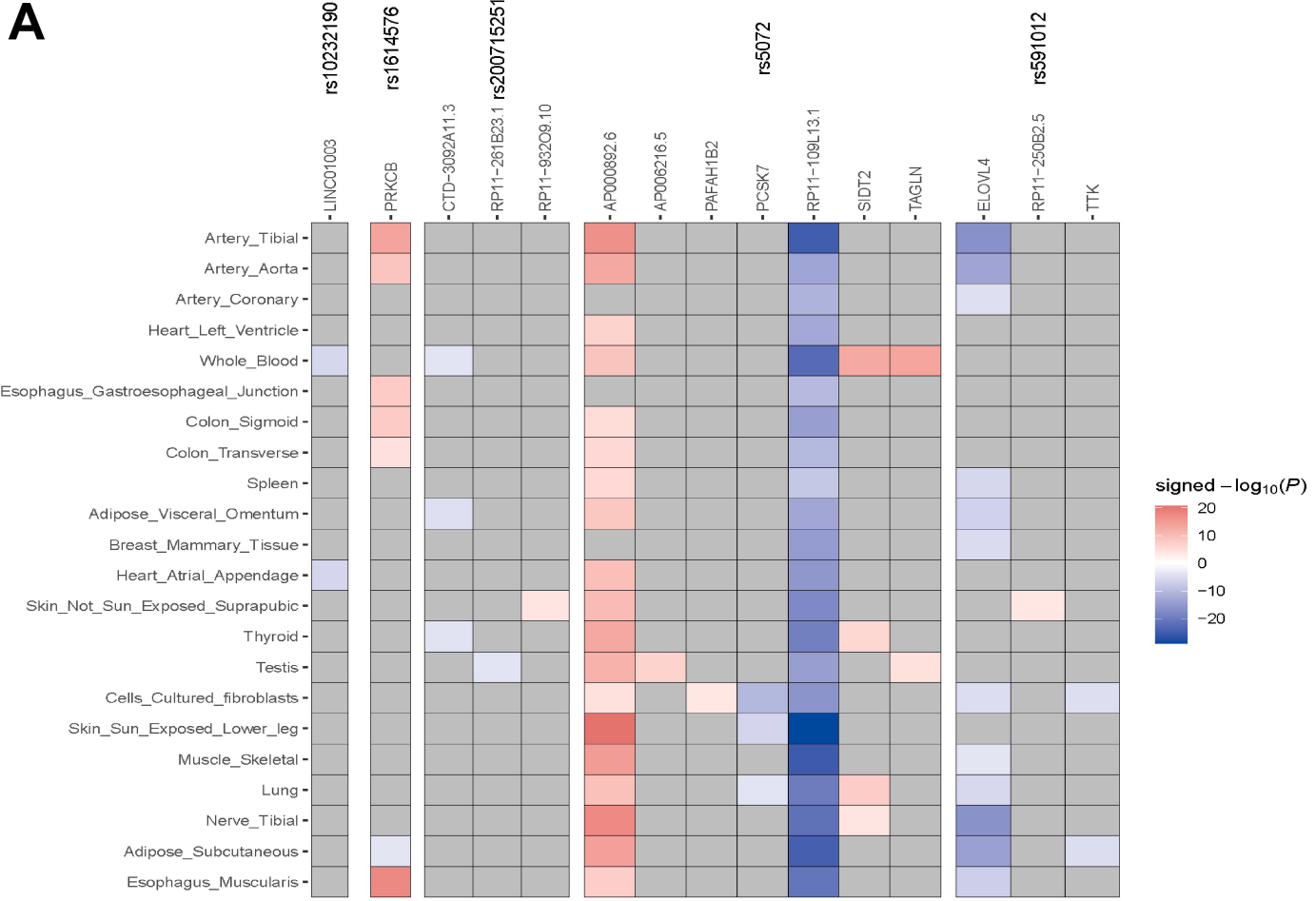
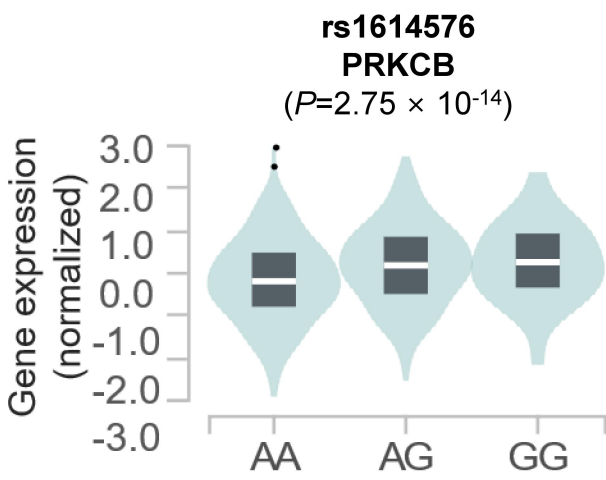
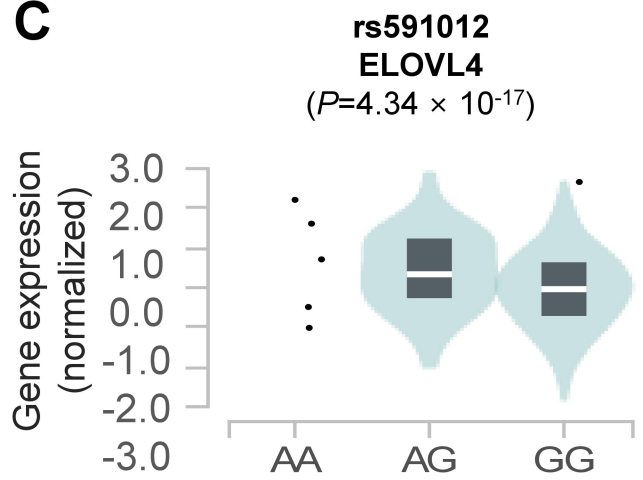
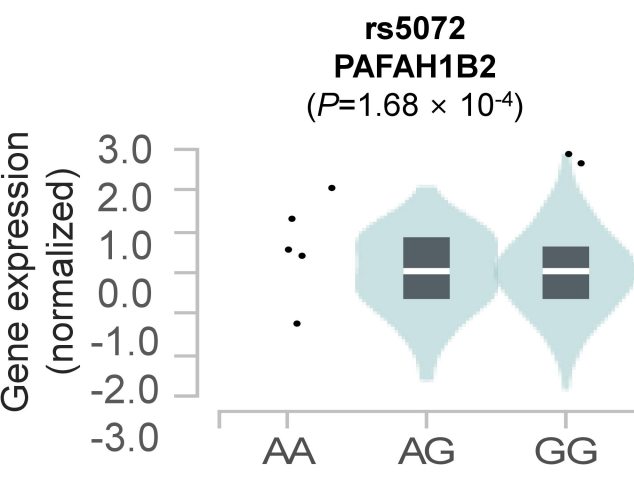
10

11

12





A**B****C****D****E**