HHS Public Access

Author manuscript

Nat Cardiovasc Res. Author manuscript; available in PMC 2023 November 02.

Published in final edited form as:

Nat Cardiovasc Res. 2022 December ; 1(12): 1140–1155. doi:10.1038/s44161-022-00171-0.

Transcriptomic-based clustering of human atherosclerotic plaques identifies subgroups with different underlying biology and clinical presentation

Michal Mokry[1,2,*], **Arjan Boltjes**[1], **Lotte Slenders**[1], **Gemma Bel-Bordes**[1], **Kai Cui**[2], **Eli Brouwer**[1], **Joost M. Mekke**[3], **Marie A.C. Depuydt**[4], **Nathalie Timmerman**[3], **Farahnaz Waissi**[3], **Maarten C Verwer**[3], **Adam W. Turner**[5], **Mohammad Daud Khan**[5], **Chani J. Hodonsky**[5], **Ernest Diez Benavente**[2], **Robin J.G. Hartman**[2], **Noortje A M van den Dungen**[1], **Nico Lansu**[1], **Emilia Nagyova**[1], **Koen H.M. Prange**[6], **Jason C. Kovacic**[7,8], **Johan L.M. Björkegren**[9,10], **Eleftherios Pavlos**[11], **Evangelos Andreakos**[11], **Heribert Schunkert**[12,13], **Gary K. Owens**[14,15], **Claudia Monaco**[16], **Aloke V Finn**[17], **Renu Virmani**[17], **Nicholas J. Leeper**[18], **Menno P.J. de Winther**[6], **Johan Kuiper**[4], **Gert J. de Borst**[3], **Erik S.G. Stroes**[19], **Mete Civelek**[5], **Dominique P.V. de Kleijn**[3], **Hester M. den Ruijter**[2], **Folkert W. Asselbergs**[20,21], **Sander W. van der Laan**[1], **Clint L. Miller**[5,22,23,#], **Gerard Pasterkamp**[1,#,*]

[1.]Central Diagnostics Laboratory, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands.

[2.]Laboratory of Experimental Cardiology, Department of Cardiology, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands

[3.]Department of Vascular Surgery, University Medical Centre Utrecht, Utrecht, The Netherlands

[4.]Leiden Academic Centre for Drug Research, Division of Biotherapeutics, Leiden University, Leiden, The Netherlands

[5.]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

[*]Correspondence: Gerard Pasterkamp, MD PhD, Central Diagnostics Laboratory, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands, g.pasterkamp@umcutrecht.nl; Michal Mokry, MD PhD, Central Diagnostics Laboratory, Laboratory of Experimental Cardiology, Department of Cardiology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands, m.mokry@umcutrecht.nl.

[#]These authors jointly supervised this work

Author contributions

MM AB analyzed and integrated the data. SWv/dL and JM provided PRS calculations. AB, SWv/dL, performed patient selection, randomization and sample handling. GJdeB performed carotid endarterectomy procedures. NAMv/dD, NL and EM tested library preparation strategies and processed coronary samples for sequencing. MACD, KHMP, LS, MPJW, JP provided, analyzed and interpreted single cell sequencing data. KC, GBB performed deconvolution of bulk RNA-seq data. JM, NT, FW, MCV recruited the patients coordinated by DPVdeK. DPVdeK, ESGS provided and coordinated Olink measurements. LS and SWv/dL performed MAGMA analysis. AB, NT, FW, DPVdeK, HMdenR, FWA, JLMB, SWv/d Laan participated in conceptualization, data interpretation, and provided critical feedback on the article. EDB, RJGH, EP, EA, HS, GKO, CM, JCK, AVF, RV, NJL and MC participated in data interpretation, and provided critical feedback on the article. AWT, MDK, CJH, JCK, JLMB and CLM recruited, processed and analyzed coronary samples. EA, FWA, SWv/dL, CLM, MM, GP provided funding. MM, CLM and GP participated in the conceptualization and supervision of the project and finalization of the article. MM prepared the figures. MM and GP drafted the manuscript. All authors provided feedback on the research, analyses, and article.

Code availability

The core scripts used for the analysis can be found at https://github.com/CirculatoryHealth/PlaqueCluster

[6.]Amsterdam University Medical Centers – location AMC, University of Amsterdam, Experimental Vascular Biology, Department of Medical Biochemistry, Amsterdam Cardiovascular Sciences, Amsterdam Infection and Immunity, Amsterdam, The Netherlands

[7.]Cardiovascular Research Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[8.]Victor Chang Cardiac Research Institute, Darlinghurst, Australia; and St Vincent's Clinical School, University of New South Wales, Australia

[9.]Department of Medicine, Karolinska Institutet, Karolinska Universitetssjukhuset, Huddinge, Sweden

[10.]Department of Genetics & Genomic Sciences, Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, 10029-6574, USA

[11.]Laboratory of Immunobiology, Center for Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[12.]Department of Cardiology, German Heart Centre Munich, Technical University of Munich, Munich, Germany.

[13.]German Centre for Cardiovascular Research (DZHK e.V.), Partner Site Munich Heart Alliance, Munich, Germany.

[14.]Robert M. Berne Cardiovascular Research Center, University of Virginia, Charlottesville, VA, USA

[15.]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA, USA

[16.]Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford

[17.]CVPath Institute, Inc, Gaithersburg, MD, USA

[18.]Department of Surgery, Division of Vascular Surgery, Stanford University School of Medicine, Stanford, CA

[19.]Department of Vascular Medicine, Amsterdam University Medical Centers, Location AMC, Amsterdam, the Netherlands.

[20.]Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

[21.]Health Data Research UK and Institute of Health Informatics, University College London, London, United Kingdom

[22.]Department of Biomedical Engineering, Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, USA

[23.]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

Abstract

Histopathological studies have revealed key processes of atherosclerotic plaque thrombosis. However, the diversity and complexity of lesion types highlight the need for improved sub-phenotyping. Here we analyze the gene expression profiles of 654 advanced human carotid plaques. The unsupervised, transcriptome-driven clustering revealed five dominant plaque types. These plaque phenotypes were associated with clinical presentation and showed differences in cellular compositions. Validation in coronary segments showed that the molecular signature of these plaques was linked to coronary ischemia. One of the plaque types with the most severe clinical symptoms pointed to both inflammatory and fibrotic cell lineages. Further, we did a preliminary analysis of potential circulating biomarkers that mark the different plaques phenotypes. In conclusion, the definition of the plaque at risk for a thrombotic event can be fine-tuned by in-depth transcriptomic-based phenotyping. These differential plaque phenotypes prove clinically relevant for both carotid and coronary artery plaques and point to distinct underlying biology of symptomatic lesions.

Introduction

The classical concept of the ‘vulnerable plaque’ that depicts plaque rupture as the major pathological substrate for acute cardiovascular events originated in the 1980s from observations in patients who died of coronary syndromes[1,2]. This recognition spawned a generation of research that led to a greater understanding of how complicated atherosclerotic plaques form and precipitate into thromboembolic events secondary to plaque rupture. Current evidence suggests that a sole focus on plaque rupture of atheromatous lesions in clinical and basic research may have oversimplified the complex collection of atherosclerotic diseases and obscured other mechanisms that may mandate different management strategies.

In addition to histology, molecular phenotyping using the whole transcriptome provides more resolution and allows an in-depth understanding and discovery of processes active in the diseased vascular tissue[3–6]. These studies successfully utilized gene co-expression networks or compared cases and controls. However, they did not attempt to redefine the plaque type definitions based on gene expression signatures. Unsupervised clustering analysis, based on transcriptomic datasets, can be used to group patients with similar molecular characteristics of the diseased tissue and has the potential to unravel disease phenotypes that fine-tune the patho-histological evaluation[7–10]. This approach is often used in cancer research and has led to the identification of novel tumor subtypes. We, therefore, hypothesized that unbiased clustering based on gene expression of human advanced atherosclerotic plaques would unveil distinct phenotypes of late-stage human atherosclerotic plaques.

Using a multi-layered approach, we created gene expression maps within a large biobank of advanced carotid lesions (n=654) and studied histological characteristics and the clinical presentation. We report that transcriptome-based analysis of human atherosclerotic lesions identified five plaque clusters linked to the occurrence of clinical events and biological processes. We highlight a plaque type that is enriched with *ACTA2* as well as *CD14* expressing cells and with the highest expression of genes involved in neutrophil degranulation, mTOR, iron uptake, and other pathways linked to active inflammatory

response and increased expression of genes involved in glycolysis. We verified these findings in coronary artery segments where this plaque type is correlated with coronary ischemic events. Finally, we performed a pilot study in search of circulating biomarkers that reflect these transcriptomic clusters. Our data demonstrate that transcriptome-based plaque characterization may have significant added value in phenotyping advanced atherosclerotic lesions that lead to clinical symptoms.

Results

The Athero-Express study cohort of carotid segments

After quality control, based on measures of RNA library complexity, we included 654 patients for clustering analysis. From those patients, 632 had accessible clinical data. The baseline characteristics of patients from Athero-Express selected for this plaque study are provided in Supplementary Table 1. In total, 75.3% of included patients were males, 24.7% were females with a mean age of 68.4 years and a mean BMI of 26.6. 43.2% of included patients had a stroke, 24.2% had a transient ischemic attack, 17.3% of patients had ocular symptoms, and the remaining 15.3% were asymptomatic. In 3 years of follow-up, 13.1% of patients suffered major adverse cardiovascular events (MACE). Histologically, 30.0% of included plaques were classified as atheromatous, 32.3% as fibrous, and 37.7% as fibro-atheromatous.

Transcriptome-defined molecular plaque types

To identify groups of patients with similar molecular signatures of advanced atherosclerotic lesion characteristics, we utilized 654 individual transcriptomes from plaques (Fig. 1a) that passed the QC filters and identified five major molecular plaque types - referred to as #0, #1, #2, #3 and #4 (Fig. 1b, Extended Data figure 1a–b). All five clusters contained samples with similar numbers of detected protein-coding genes (Extended Data Fig. 1c). Based on the principal component analysis (PCA) projection and correlation to the most similar sample (Extended Data Fig. 1a–c), type #3 contained more heterogeneous plaques while the other clusters demonstrated higher correlations between the samples. Overall, we did not observe clear boundaries between five plaque types (Fig. 1b, Extended Data Fig. 1a–b). Different data integration and normalization methods[11] and correction for the technical variables (sequencing batch and hospital) yielded similar clusters (Extended Data Fig. 2a–f). Permutation analysis with 90% of samples showed that, on average, in 84.8% of cases, the sample was assigned to a cluster that matched the original classification. Samples more frequently classified in non-matching clusters were mostly found at the border zone between clusters in UMAP projection (Extended Data Fig. 2g–h). This altogether suggests that the clustering was robust while also suggesting the presence of intermediate types.

The five transcriptomics-based plaque types differed vastly in the expression of numerous candidate marker genes (Fig. 1c, Supplementary Table 2), which supports that the clustering is robust and reflects different underlying biology. For instance, type #0 plaques showed increased expression of *FGF13*, *LPIN1*, and *KYNU*. Types #1 and #3 showed increased expression of inflammatory molecules and leukocyte markers (*CXCL12*, *C1QA*, *CD14*, *CD73*, or *APOE*), while #3 also showed a modestly increased expression of classical

smooth muscle cell markers (*MYH11*, *MYH10*, and *ACTA2*). Type #2 plaques showed increased expression of *NOS1*, *SOD2*, *VDR*, *SLC35E3*, and *ATXN3*, while they mostly lacked expression of immune cells and SMC markers. Finally, type #4 had the highest expression of classical smooth muscle cell markers (*MYH11*, *MYH10*, and *ACTA2*) and reduced inflammatory and leukocyte marker gene expression.

Next, to understand which molecular processes and pathways underlie the five molecular plaque types, we performed pathway analysis on genes upregulated (based on the differential gene expression analysis) in the individual clusters. Clusters representing plaque types #1, #3, and #4 showed significant enrichment in numerous pathways (Supplementary Table 3). Increased expression of genes involved in neutrophil degranulation, mTOR, iron uptake, and other pathways linked to active inflammatory-response-related processes overlapped with increased expression of glycolysis genes, specifically in type #3 (Fig. 1d–e, Supplementary Table 4). The same plaque-type showed decreased activity of genes involved in fatty acid oxidation, while the activity of the citric acid cycle (TCA cycle) was not significantly different between clusters. Similarly, the processes involved in an extracellular matrix organization (synthesis, organization, and degradation) and cell plasticity (endothelial-mesenchymal transition - EMT) showed specific enrichment in individual plaque types.

The genes upregulated in plaque types #0 and #2 did not show significant enrichment in the pathway analysis. However, the projection of empirically selected pathways that involve some of the cluster-specific genes (for example, tryptophan catabolism, which involves L-kynureninase - *KYNU*) showed increased expression in these specific clusters (Fig. 1e).

Altogether, the transcriptome-defined plaque types differ in molecular signatures involving metabolism, the inflammatory response, and the processes involved in extracellular matrix (ECM) homeostasis.

Molecular plaque types differ in histological composition.

Since variation in the cellular composition of complex tissues is one of the main drivers of transcriptomic differences, we analyzed the histological features of the five molecular plaque clusters (Fig. 2a, Supplementary Table 5). Of interest, one specific plaque type (#3) was characterized by gene enrichment pointing to cell types and processes that are commonly used to differentiate between stable and unstable plaques (Fig. 2a). Type #3 showed high expression of genes specific for inflammatory cells (e.g., *C1QA* and *CD14*) and cells responsible for a fibrous phenotype (e.g. *ACTA2*). Clusters #0 and #4 were significantly enriched in fibrous plaques with lower fat content and increased expression for *ACTA2*. On the contrary, type #2 was more enriched in atheromatous plaques, high-fat content, and increased presence of *CD68* positive cells. No clusters exhibited a distinct pattern for the presence of calcification, media remnants, collagen content, and intraplaque hemorrhage (IPH).

Even though the plaque types #0 and # 4 showed both absolute and relative increase in *ACTA2* positive cells in the histological evaluation, only in type #4 was this followed by clear upregulation of classical SMC markers like *ACTA2* or *MYH11* in the transcriptomic data (Fig. 2b). Similarly, genes involved in ECM organization were downregulated in type

#0 compared to #4 and #3. While contractile SMC genes *ACTA2* or *MYH11* had the highest expression in plaque type #4, the molecular pathway term “smooth muscle contraction” (Reactome, R-HSA-445355) was the highest in plaque type #3. This term also contains genes involved in upstream signaling cascades and calcium handling. Specifically, genes involved in calcium transport or sensing (like annexins, dysferlin, or integrin *ITGB3*) were overexpressed in plaque type #3. Interestingly, this specific cluster was also enriched (though the trend is not statistically significant) in plaques with no evidence of calcification (Fig. 2a, Supplementary Table 5).

In line with the marker gene expression, genes overexpressed in type #0 did not show specific expression in *ACTA2*+ cells in single-cell transcriptomics data (Fig. 2c); instead, they seemed to be expressed in populations identified as “mixed cells” - a population without a clear cell type-defining expression profile. This mixed plaque type seemed to contain apoptotic myeloid and T cells[12] and express some of the foam cell driver genes (e.g., *LRP1B*[13]). Altogether, the histologically scored numbers of *ACTA2*+ cells in plaque type #0, which resides in a highly fibrotic environment, seem to be overestimated or represent a less transcriptionally active population.

Similarly, the histological evaluation of plaque type #2 was associated with increased *CD68*+ cells. However, the expression of macrophage markers like *CD14* and genes involved in inflammatory pathways was predominantly downregulated at the RNA level. Additionally, the expression of genes specific to the type #2 cluster project to a likely apoptotic “mixed cell” population. This suggests that the residing macrophages in type #2 plaques have decreased transcriptional activity, or their numbers in the histological evaluation are overestimated. Notably, plaque types do not seem to be strongly driven by the variation in NK-, B- or T-cell content (Fig. 2b–c).

These observations may also have implications for the interpretation and extrapolation of animal studies on mechanisms of atherosclerotic disease. After comparing the RNAseq data from *ApoE*−/− and *Ldlr*−/− mice and correlating them with the expression profiles of five human plaque types, cluster #3 is the closest match with murine models (Fig. 2d). This observation may have significant impact on the way we use and interpret atherosclerotic murine models, as they do not seem to cover all the phenotypic (transcriptomic) variability of human samples.

Transcriptome-defined plaque types reflect disease severity.

Next, to gain clinical insights into the distinct transcriptome-based plaque clusters, we explored the severity of clinical symptoms before the surgery. Clusters #2 and #3 had higher percentages of more severe symptoms – TIA or stroke (80.3%, 77.0%, respectively), compared to intermediate # 1 (68.3%) and less severe #0 and #4 (55.8% and 61.1% respectively) (p=0.001) (Table 1).

We have also analyzed the occurrence of MACE within 3 years from surgery among the clusters and found no significant difference in the frequency of MACE (p=0.629). This is somewhat surprising, considering that specific histological features correlate with MACE and the strong differences in clinical symptoms between clusters prior to sample collection.

Therefore to investigate whether gene expression data carry information that reflects the risk of MACE, we built a logistic regression model to predict adverse cardiovascular outcomes within a 3-year follow-up time. We achieved an AUC = 0.66 (95% CI: 0.58–0.72) by using only the histology data. By adding transcriptomic information (as PCs) the prediction model was significantly improved (AUC = 0.74, 95% CI: 0.67–0.80, p = 0.0111, NRI = 0.12) (Fig. 3a). These results suggest that the plaque transcriptomes encompass significant added value for the association with clinically relevant atherosclerotic disease.

Next, we compared the baseline clinical characteristics, risk factors and medication before the inclusion in different clusters. We observed a non-random distribution of age in years±SD (mean 67.2±8.88, 69.1±8.80, 70.4±9.23, 67.5±8.80 and 68.9±8.30 in clusters #0, #1, #2, #3 and #4 respectively, p=0.019) and total cholesterol levels in mmol/L±SD (mean 4.59±1.27, 4.48±1.24, 4.11±1.08, 4.58±1.32 and 4.44±1.11 in clusters #0, #1, #2, #3 and #4 respectively, p=0.050) (Table 1).

The other risk factors (including sex of patients, smoking status, diabetes, hypertension, body mass index, triglycerides, LDL, HDL, hypertension) and medication were not significantly different between the transcriptome-driven plaque types.

Genetic differences contribute to molecular plaque types.

Atherosclerosis is a complex disease with a significant genetic component. In studies published during or before 2021, GWASs have identified over 163 independent genetic loci associated with atherosclerotic disease[14]. We hypothesized that the transcriptome-defined clustering reflects the genetic component of atherosclerotic disease and compared polygenic risk scores (PRS) for coronary artery disease (CAD)[15]. We observed a non-random distribution (p = 0.003) with the highest PRS in plaque type #3 (Fig. 3b), suggesting that this plaque type is at least partially driven by known genetic associations with CAD. The expression of genes genetically associated with CAD using gene-based testing (MAGMA[16]) showed a similar trend, with the highest expression of CAD-associated genes in plaque type #3 (Fig 3c).

Clustering in coronaries is associated with ischemia.

The five transcriptomic clusters were identified in carotid arteries. Therefore, confirming a similar phenomenon in other vascular beds would strengthen the translational potential of this concept. To investigate whether plaques from other relevant anatomical locations form clusters based on transcriptional profiles, we repeated the clustering analysis on transcriptomics datasets derived from coronary arteries isolated from heart transplant recipient or donor hearts[17,18]. We applied the same algorithm as for plaques from carotid arteries. In coronary artery segments, we were able to identify four distinct clusters (Fig. 4a–b). Similar to carotids, the individual plaque types in coronary arteries showed specific gene expression profiles, molecular pathways, and processes (Fig. 4e). Finally, the four clusters identified in coronary arteries differed in percentages of samples from hearts with underlying ischemia (18.2%, 54.9%, 6.5 % and 0.0%, respectively; p < 0.001) (Fig. 4f). The cluster with the largest proportion of ischemic samples also showed increased expression of genes involved in ECM organization, neutrophil degranulation, TGF-beta signaling, and

glycolysis. Notably, these were the same pathways associated with clinical symptoms in bulk transcriptomic analyses of carotid plaques in plaque type #3 (Fig. 1d).

Since the coronary samples comprise the entire artery, including adventitial tissue, and differ in RNA quality - which can confound the comparisons, the four clusters identified in coronaries cannot be directly compared to the clusters from carotid arteries. Therefore, we performed integrative analysis to identify the most similar carotid sample for each sample from coronary arteries. We first employed the anchor-based data integration algorithm, which allows merging datasets with different confounders[19]. After the integration, we repeated the clustering and identified five clusters with similar gene signatures as in the carotid dataset (Extended Data Fig. 3ab). Next, we identified the best matching carotid sample to each coronary using Pearson correlation (Fig. 4c–d) and assigned the corresponding clusters from the original carotid data.

The samples assigned to the same clusters again showed the specific molecular pathway activity (like neutrophil degranulation or glycolysis, Fig. 4g) and correlated with clinical manifestation (Fig. 4h). Of note, plaque type #2 seems to be underrepresented in coronary artery data. The best matching coronary samples for the patients detected in the symptomatic carotid plaque type #3 were observed in coronary cluster #1 (Fig. 4d), which is the plaque type with the highest prevalence of ischemic cardiac disease. This suggests that plaques in different anatomical locations can be clustered in groups connected to the clinical manifestation. Some of those clusters also overlap between different anatomical localizations.

Circulating biomarkers correlate with plaque clusters.

The presented description and categorization of ‘vulnerable’ plaque types that goes beyond the scope of histopathological phenotyping is of great value to the scientific community in the field of vascular biology. However, the proposed type of patient and plaque stratification approach must also be translationally applicable – for example, throughout the identification of circulating biomarkers that can be measured in a clinical setting. Therefore, to explore these possibilities, we have utilized the measurements of circulating biomarkers, which were shown to associate with the disease severity (Supplementary Table 6), and correlated them to five plaque types. Out of 15 biomarkers, MRP8/MRP14 complex and osteopontin showed nominally significant uneven distribution between the clusters. Some others, like FABP4, CRP, IL8, or TNF alpha, showed interesting (not statistically significant) trends.

To approach this more systematically, we analyzed cardiovascular OLINK biomarker panels in available blood samples of 386 patients, searching for circulating biomarkers that associate with the different transcriptomic clusters. Out of 276 blood-derived biomarkers, we found that 8 were nominally associated ($p<0.05$) with patho-histological plaque type (atheromatous, fibroatheromatous, or atheromatous) and 11 with the presence of calcifications (Fig. 5a). However, 24 circulating biomarkers showed nominally significant, different levels among transcriptomic plaque clusters (Fig. 5a).

Only PAI (plasminogen activator inhibitor) remained significant after multiple testing corrections (FDR = 0.016) and was decreased in plaque type #4 patients. Studies have

shown that PAI inhibitors significantly reduce atherosclerosis formation in a murine model of obesity and metabolic syndrome, and pharmacological targeting of PAI inhibits macrophage accumulation and cell senescence in atherosclerotic plaques[20]. This altogether substantiates further studies to search for relevant circulating biomarkers specific to the transcriptome-defined plaque types.

Discussion

In the present study, using unsupervised clustering, we describe five transcriptomic clusters in carotid plaques that relate to biological pathways and clinical presentation (Fig. 6). Pathways that appeared overrepresented in a plaque type with more vulnerable patients pointed towards, among others, neutrophil degranulation, matrix organization, and glycolysis.

Pathology and transcriptome-based plaque characterization.

Our transcriptomic data indicate the existence of plaque types that partly overlap with the pathological characterization and are associated with clinical symptoms of ischemic stroke and myocardial ischemia. Our analyses also provide insight into cell-specific processes that can play a role in the destabilization of the plaque. Surprisingly, carotid plaques in plaque type #3, with the most severe symptoms and with the best match with symptomatic coronary plaques, showed enrichment of pathways and RNA-based cell types that are traditionally categorized in two distinct plaque categories. This plaque type revealed an abundance of the smooth muscle cell and macrophage-specific genes, cell types that in pathology and animal experimental categorizations are often considered as “stabilizing” and “destabilizing”, respectively. Our transcriptomic analyses confirm the existence of a complex landscape of atherosclerotic lesion phenotypes that associate with clinical symptoms. The inferences regarding the risk of clinical thrombotic events based purely on pathological findings therefore merit careful (re)consideration.

Plaque molecular types are associated with ischemic events.

In both carotid and coronary plaques, there was a strong distinction between clusters in the expression of genes annotated to different molecular pathways. Pathways indicating neutrophil activation, glycolysis, extracellular matrix organization, and iron uptake and transport were strongly over-represented in both arteries in the same cluster that was also associated with clinical events. This plaque type showed an overrepresentation of *CXCL12*, *CD14*, *C1QA*, *CD63*, *CD74*, and *APOE*; genes that are often observed in plaque-derived macrophages. Indeed, the overlap with single-cell sequencing demonstrated that a significant proportion of the transcripts upregulated in this plaque type pointed to monocyte/macrophage lineages (Fig. 2c).

The role of neutrophil activation in plaque stabilization has long been underestimated, but the discovery that neutrophil extracellular traps (NETs) may have a causal role in plaque destabilization confirms a potential role of genes indicating neutrophil activation[21–23]. Using in vivo and in vitro assays, Soehnlein and colleagues showed that activated SMCs in atherosclerotic plaques release chemotactic factors that attract neutrophils and trigger

the release of NETs containing histone H4, which has cytotoxic effects on SMCs[24]. In addition, it has been shown that neutrophil microvesicles accumulate at disease-prone regions of arteries exposed to disturbed flow patterns and promote vascular inflammation and atherosclerosis in a murine model[25].

The role of metabolic pathways such as glycolysis[26], fatty acid oxidation, and tryptophan catabolism may point to phenotype switching of vascular cells[27]. The differentiation of a contractile to a synthetic myofibroblast-like phenotype in smooth muscle cells is driven by switching from oxidative to glycolytic metabolism[28]. In endothelial cells, glycolysis is essential for ATP production and sprouting of vessels since the loss of the glycolytic activator *PFKFB3* in ECs impairs vessel formation[29]. In addition, enhanced glycolysis has been recognized as a critical role in initiating endothelial or epithelial to mesenchymal transition (EMT) progression[30]. Smooth muscle cell and endothelial cell differentiation are mediated by local inflammation. The increased activity of the glycolysis pathway may thus relate to an active state of cell types that differentiate and acclimate to an unstable (inflammatory) environment.

Surprisingly, carotid plaques in cluster #3, with the most severe symptoms and with the best match for coronary plaques in ischemic hearts, showed enrichment of pathways and RNA-based cell types that are traditionally categorized in two distinct plaque categories. This cluster revealed an abundance of smooth muscle cell and macrophage-specific genes, cell types that in pathology and experimental animal categorizations are often considered as “stabilizing” and “destabilizing” respectively. It could be of great interest if plaque transcriptomic analyses would allow the identification of plaque types that point to a high incidence of mesenchymal cell transitions in atherosclerotic lesions as a possible explanation for the expression of genes that are normally applied for the identification of different cell types such as *ACTA2* and *CD14* as has been observed in a number of cell lineage tracing studies in mice. However, bulk RNAseq does not provide the resolution of cell-specific RNA expression data that are required to draw inferences in the context of cell transitions.

The role of iron uptake in atherosclerosis has been debated and is considered multifaceted. Iron can cause oxidative damage and regulated cell death by lipid peroxidation (known as ferroptosis), and oxidized lipoprotein can then be taken up by the LDL receptor on macrophages leading to their development into foam cells (reviewed in[31]). On the other hand, *CD163*+ alternative macrophages engulfing the hemoglobin-haptoglobin complexes (HH) were shown to augment hyaluronan synthesis in vascular SMCs and prevent vascular calcification[32].

Genes overexpressed in the plaque transcriptomic type that predicted stroke and cardiac ischemia (like *CD14*) support the involvement of inflammation in this high-risk plaque. *CXCL12* is a well-studied anti-inflammatory gene that is also associated with coronary artery disease through GWAS[33]. *CD63* is a protein that is mainly associated with the membrane of intracellular vesicles and is also used as a marker for blood platelet activation[34] as well as circulating vesicles[35] and functionally plays a role in signal transduction. *C1QA* is considered to play a role in removing apoptotic cells via efferocytosis, and its presence reduces the atherosclerotic plaque size in animal models[36].

Clinical value of transcriptomic-based plaque characterization.—The pathology-based description of the ‘vulnerable plaque’ has proven crucial value for understanding the progression and complications of atherosclerotic disease in experimental research[37]. In addition, the clinical utility of numerous vascular imaging modalities is based on the well-known definition of thin cap fibroatheroma. The discovery and evaluation of pharmaceutical treatments rely on animal models that apply the pathology-based surrogate measures of destabilizing atherosclerotic disease. Our observations provide evidence that transcriptomic profiling and subsequent clustering of human plaques may have substantial added value in searching for suitable drug targets. These observations that significantly fine-tune the phenotype of the destabilizing plaque can be translated to animal models and cell culture systems used in drug-related research.

There is an unmet need for biomarkers associated with the presence of plaques at risk for a thrombotic ischemic event. Proteins can today be measured with high accuracy using multiplex methods[38]. We feel that the observed enrichment of biomarkers associated with different transcriptomic plaque clusters is promising. Our sample size with combined plaque transcriptomic and clinical data does not allow strong inferences and surely requires verification. Our data do show that non-stratified biomarker analyses in large pooled patient samples may mask subgroups in whom biomarker profiles could be predictive. For example, clusters #0 and #4 both encompassed patients with mild symptoms but had different associations with biomarker profiles. Future efforts will be focused on more complex analysis[39] and on building prediction models[40], which, based on measurements of circulating biomarkers, can predict the plaque types present in patients.

Limitations

Our results could be biased by the type of source material for the bulk RNAseq analyses. Different regions in a plaque can reveal different cell types that could explain the differential clustering among patients. To avoid this bias, the plaque samples selected for the RNAseq were those with available segments closest to a 0.5 cm length culprit lesion (which is used for pathology). In addition, the associations observed between clusters, clinical symptoms, and polygenic risk score and the substantial overlap with observations in coronary samples do support our view that this bias cannot fully explain the observed outcomes.

Similarly, the quality of RNA obtained from atherosclerotic lesions can be affected by the lesional morphology, such as the presence of necrotic debris. This may have biased our results. However, we were able to verify the presence of similar clusters, driven by identical biological processes in the coronary samples that also encompassed less advanced lesions which support the view that results cannot be fully explained by lesion type with differential grades of RNA degradation.

In summary, our study shows that deciphering the transcriptomic profile of atherosclerotic lesions results in an updated description of the ‘vulnerable plaque’ with clinical relevance. Our results also demonstrate that transcriptomic analyses can contribute to assessing lesion phenotypes that are predisposed to a thrombotic event and potentially reveal underlying pathogenetic mechanisms.

Methods

Carotid plaque samples

The Athero-Express Biobank (AE) includes patients undergoing carotid endarterectomy, of which the study design has been published before[41–43]. The AE study is an ongoing biobank, and extensive baseline characteristics, blood samples, and atherosclerotic plaque specimens are collected. Clinical data were obtained from patient files and through standardized questionnaires. The indication for CEA was based on the recommendations from the Asymptomatic Carotid Surgery Trial (ACST) for asymptomatic patients and the European Carotid Surgery Trial (ECST) and North American Symptomatic Carotid Endarterectomy Trial for symptomatic patients (NASCET). Indications for CEA were evaluated by a multidisciplinary vascular team. The removal of atherosclerotic plaques was performed by a team of experienced surgeons, and standardized treatment protocols were applied. All patients were examined by a neurologist for assessment of their preoperative neurologic status. For this study, subsequent patients were included who underwent carotid endarterectomy between 2002 and 2016 and of whom genotyping data were available. The performed study is in line with the Declaration of Helsinki and informed consent was provided by all study participants after the approval for this study by medical ethical committees of the different hospitals (the University Medical Center Utrecht NL and St. Antonius Hospital Nieuwegein NL) was obtained.

Baseline characteristics

Baseline data were obtained by chart review and from extensive questionnaires completed by the participating patients that included questions on the history of cardiovascular disease, cardiovascular risk factors (smoking, hypertension, diabetes), and use of medication. Presenting symptoms and duplex stenosis were retrieved from patient charts. Symptom categories were “asymptomatic”, defined as not having ipsilateral cerebrovascular symptoms in the previous 6 months; “ocular” - amaurosis fugax, defined as ipsilateral mono-ocular blindness of acute onset lasting <24 hours; cerebral “transient ischemic attack” (TIA), defined as the ipsilateral focal neurologic deficit of acute onset lasting <24 hours; and ipsilateral “stroke”. Lipid spectra were determined in blood specimens drawn at baseline.

Follow up

All patients answered a questionnaire 1, 2, and 3 years after the carotid endarterectomy. In case an adverse event was reported or suspected, the referring hospital or general practitioner was approached for additional medical information. The primary outcome was defined as a composite of endpoints including, any death of vascular origin (fatal stroke, fatal myocardial infarction, sudden death, and other vascular death), non-fatal stroke (either ischemic or hemorrhagic), or non-fatal myocardial infarction, and any arterial vascular intervention that had not already been planned at the time of inclusion (e.g. carotid surgery or angioplasty, coronary artery bypass, percutaneous coronary artery intervention, peripheral vascular surgery, or angioplasty).

Sample handling

The atherosclerotic plaques were transported to the laboratory and processed immediately after the surgical removal. An experienced technician identified the culprit lesion, which is defined as the segment with the smallest lumen. In case of doubt, the segment with the largest plaque diameter was selected. According to a standardized protocol, the plaque was divided into segments of 5 mm thickness along the longitudinal axis. The segment with the culprit lesion was then prepared and stored in 4% formaldehyde, decalcified, and embedded in paraffin for histological analysis. The rest of the plaque was snap-frozen using liquid nitrogen and stored at −80 degrees Celsius.

Histological assessment

The assessment was performed according to a previously validated protocol, which was described in detail before[44]. Cross-sections of the culprit lesion are stained and quantified for each patient at 40× magnification. A hematoxylin-eosin (HE) staining was performed for the assessment of calcification and picrosirius red for collagen. Collagen staining and calcification were assessed and categorized as (1) no, (2) minor, (3) moderate, and (4) heavy independent of the localization of the staining. CD68 (antibody: Roche cat# 790–2931, used undiluted) was stained to identify macrophages. The criteria for classification were defined as follows: macrophages: (1) no or just a few scattered cells (2) minor CD68 staining with clusters with less than 10 cells present; (3) moderate: cell clusters with >10 cells present or (4) heavy, the abundance of positive cells. Alpha-smooth muscle actin (Biosite BSH-7459–1, clone BS66, 1:20000 dilution) was stained to determine the presence of smooth muscle cells. The criteria for classification were defined as follows: (1) no or few scattered cells (2 minor alpha-actin staining over the entire circumference with absent staining at parts of the circumference of the arterial wall; (3) moderate: positive cells along the circumference of the luminal border, with locally at least few scattering cells (4) heavy: SMC dominant plaque with cells within the entire cap and also large clusters deep in the lesion. The location of SMCs was evaluated as “mainly basal”, “homogeneous,” or ”mainly luminal”. The location of macrophages was evaluated as “basal”, “homogeneous”, ”luminal”, or “no macrophages”. In the present study, some of these categories were binned into no/minor staining and moderate/heavy staining. Immunohistochemical staining for *CD34* (Ventana, CONFIRM anti-CD34 (QBEnd/10) Primary Antibody, 790–2927, used undiluted) was performed to assess vessel density. Plaque microvessels were quantified in three hotspots and subsequently, the average number of vessels per hotspot was calculated. Picrosirius red in combination with elastic-Van Gieson, HE, and polarized light was used to visualize the lipid core. The lipid content of the plaque was estimated as a percentage of the total plaque area, with a cutoff at 10% and 40% for carotid plaques. Plaques with<10%, 10–40%, and >40% fat were categorized as fibrous, fibro-atheromatous, and atheromatous, respectively. The presence of plaque hemorrhage was determined using HE staining, fibrin staining, and glycophorin staining. All plaque characteristics were scored and quantified with good intra- and interobserver reproducibility by two independent observers[54].

RNA isolation and library preparation

A total of 700 segments were selected from patients who were included in the study between 2002 and 2016. The atherosclerotic plaque was grounded while frozen with liquid nitrogen, and after that Tripure (Roche, cat# 11667165001) was added, and the plaque pieces and further disrupted by the Precellys 25 homogenizer (Bertin Instruments). The sample was incubated at room temperature for 5 minutes and centrifuged at 20.000g for 1 minute at 4°C. The supernatant was mixed with chloroform and incubated at RT for 15 minutes. The sample was centrifuged at 12.000xg for 5 minutes at 4°C, and the upper phase was used for RNA isolation. Then isopropanol and GlycoBlue (Invitrogen, cat# 10301575) were added to the aqueous phase to precipitate the RNA and centrifuged at 12.000xg for 10 minutes at 4°C. The pellet was washed with 75% ethanol and resuspended in RNase-free water.

The RNA isolated from the archived advanced atherosclerotic lesion is fragmented (Extended Data Fig. 4a). We have, therefore, tested four different library preparation strategies (Extended Data Fig. 4bc): CEL-seq2[45], QIAseq (QIAseq Stranded Total RNA Lib Kit, Qiagen, cat# 180743), NEXTflex (NEXTflex Rapid Directional RNA-Seq Kit, Bioo Scientific, cat# NOVA-5138–08), and SMARTer (SMARTer® Stranded RNA-Seq Kit, Takara, cat# 634862) using the manufacturer's or author's recommendations. We have ultimately employed the CEL-seq2 method. CEL-seq2 yielded the highest mappability reads to the annotated genes compared to other library preparation protocols (Extended Data Fig. 4bc). The methodology captures the 3'-end of polyadenylated RNA species and includes unique molecular identifiers (UMIs), which allow direct counting of unique RNA molecules in each sample. 50ng of total RNA was precipitated using isopropanol and washed with 75% ethanol. After removing ethanol and air-drying the pellet, a primer mix containing 5ng primer per reaction was added, initiating primer annealing at 65°C for 5min. Subsequent RT reaction was performed; first-strand reaction for 1h at 42°C, heat-inactivated for 10m at 70°C, second strand reaction for 2h at 16°C, and then put on ice until proceeding to sample pooling. The primer used for this initial reverse-transcription (RT) reaction was designed as follows: an anchored polyT, a unique 6bp barcode, a unique molecular identifier (UMI) of 6bp, the 5' Illumina adapter, and a T7 promoter, as described. Each sample now contained its own unique barcode due to the primer used in the RNA amplification, making it possible to pool together cDNA samples at 7 samples per pool. Complementary DNA (cDNA) was cleaned using AMPure XP beads (Beckman Coulter, cat# A63882), washed with 80% ethanol, and resuspended in water before proceeding to the *in vitro* transcription (IVT) reaction (AM1334; Thermo-Fisher) incubated at 37°C for 13 hours. Next, primers were removed by treating with Exo-SAP (Affymetrix, Thermo-Fisher, cat# 78201.1.ML), and amplified RNA (aRNA) was fragmented and then cleaned with RNAClean XP (Beckman-Coulter, cat# A63987), washed with 70% ethanol, air-dried, and resuspended in water. After removing the beads using a magnetic stand, RNA yield and quality in the suspension were checked by Bioanalyzer (Agilent).

cDNA library construction was then initiated by performing an RT reaction using SuperScript II reverse transcriptase (Invitrogen/Thermo-Fisher, cat# 18064022) according to the manufacturer's protocol, adding randomhexRT primer as a random primer. Next, PCR amplification was done with Phusion High-Fidelity PCR Master Mix with HF buffer (NEB,

MA, USA, cat# F531L) and a unique indexed RNA PCR primer (Illumina) per reaction, for a total of 11–15 cycles, depending on aRNA concentration, with 30 seconds elongation time. PCR products were cleaned twice with AMPure XP beads (Beckman Coulter, cat# A63882). Library cDNA yield and quality were checked by Qubit fluorometric quantification (Thermo-Fisher, cat# Q32851) and Bioanalyzer (Agilent), respectively. Libraries were sequenced on the Illumina Nextseq500 platform; paired-end, 2 × 75bp.

Sequencing read mapping and quality filtering

Libraries were sequenced on the Illumina Nextseq500 platform; a high output paired-end run of 2 × 75 bp was performed (Utrecht Sequencing Facility). The reads were demultiplexed and aligned to human cDNA reference (Ensembl 84) using the BWA (0.7.13) by calling 'bwa aln' with settings -B 6 -q 0 -n 0.00 -k 2 -l 200 -t 6 for R1 and -B 0 -q 0 -n 0.04 -k 2 -l 200 -t 6 for R2, 'bwa sampe' with settings -n 100 -N 100. Multiple reads mapping to the same gene with the same unique molecular identifier (UMI, 6bp long) were counted as a single read. The raw read counts were corrected for UMI sampling (corrected_count=−4096*(ln(1-(raw_count/4096)))), normalized for sequencing depth and quantile normalized. We have detected a median of 19.501 (SD = 5.874) genes per sample with at least one unique read (Extended Data Fig. 4d) and discarded samples (n=46) with less than 9000 detected genes from further analysis (Extended Data Fig. 4e and Fig. 1a). For all the subsequent analyses, we have excluded all the ribosomal genes and used only the protein-coding genes with annotated HGCN names.

Clustering of transcriptomics datasets

Clustering of datasets from carotid arteries was based on the first 12 principal components (PCs) calculated using 5000 most variable genes from the normalized gene expression data. We used the shared nearest neighbor (SNN) modularity optimization-based clustering algorithm[46] implemented in the Seurat package[47] (core scripts can be found at https://github.com/CirculatoryHealth/PlaqueCluster). 162 transcriptomics data sets from coronary arteries were clustered separately in the same way.

Pathway analysis

Pathway analysis was performed using the "ReactomePA" R/Bioconductor package[48]. Module score were calculated using the Seurat's AddModuleScore() function.

Baseline characteristics tables

Baseline characteristics tables were produced using R's "tableone" package with default settings. Statistical significance of differences between groups was tested using the chi-square test for categorical variables and one-way analysis of variance (ANOVA) for continuous variables (with equal variance assumption, i.e., regular ANOVA).

Logistic regression model

The logistic regression models were built using the principal components (PCs) calculated from 5000 most variable genes from the normalized gene expression data. Raw counts were corrected for UMI sampling bias (corrected.genecounts = (−4096*(ln(1-(raw.genecounts/

4096))))) and sequencing depth and quantile normalized. We have used (Model: “RNA based model”) the first 25 PCs from RNA sequencing data, (Model: “Histology”) histological parameters - plaque phenotype, the mean number of macrophages, the mean number of smooth muscle cells, fat core presence (atheromatous, fibro-atheromatous, fibrous), calcification score, collagen presence score, SMC presence score, SMC location, macrophages presence, macrophages location, SMC to macrophages ratio, media presence and intraplaque hemorrhage and (Model: “RNA + Histology”) combined histological and transcriptomics parameters. All models included sex and age of the patients. Next, we used a backward stepwise algorithm based on the Akaike information criterion to reduce the number of parameters in the final models.

Correlation with mouse dataset

We have accessed three transcriptomics datasets Apoe#1 (Affymetrix HT-MG-430 PM microarray, whole aorta from the arch to the mid-abdomen): GSE66569[49], Apoe#2 (RNA-seq, vascular tissue): GSE186252[50] and LDLR (RNA-seq, aortic arch): GSE163206[51]. First, gene IDs and microarray probes were matched to Ensembl IDs, quantile normalized, and log2 transformed after adding 1 to each expression value. Next, the corresponding human orthologue for each mouse Ensembl ID is found and the correlation between each human and mouse dataset was calculated with a Pearson’s correlation.

Human coronary artery tissue procurement

Human coronary artery tissue biospecimens were obtained at Stanford University from diseased heart transplant donors consenting to research studies under Institutional Review Board protocols (#4237 and #11925). Hearts were arrested in cardioplegic solution and transported on ice prior to dissecting proximal coronary artery segments from main branches of left anterior descending, circumflex or right coronary arteries. Epicardial and perivascular adipose was trimmed on ice, rinsed in cold phosphate-buffered saline, rapidly frozen in liquid nitrogen, and stored at −80C. Normal human coronary artery tissue biospecimens were also obtained at Stanford University from non-diseased donor hearts rejected for orthotopic heart transplantation processed following the same protocol as hearts for transplant. Tissues were de-identified and clinical and histopathology information was used to classify ischemic and non-ischemic arteries. All normal arteries originated from hearts with a left ventricular ejection fraction (LVEF) greater than 50%. Frozen tissues were transferred to the University of Virginia through a material transfer agreement and Institutional Review Board-approved protocol (#20008).

RNA sequencing of coronary samples

Total RNA was extracted from frozen coronary artery segments (n = 162, 57 females, 105 males, Age (mean) = 50.1years (standard deviation = 15.2years) using the Qiagen miRNeasy Mini RNA Extraction kit (catalog #217004). Approximately 50 mg of frozen tissue was pulverized using a mortar and pestle under liquid nitrogen. Tissue powder was then further homogenized in Qiazol lysis buffer using stainless steel beads in a Bullet Blender (Next Advance) homogenizer, followed by column-based purification. RNA concentration was determined using Qubit 3.0 and RNA quality was determined using Agilent 4200 TapeStation. Samples with RNA Integrity Number (RIN) greater than 5.5

and Illumina DV_{200} values greater than 75 were included for library construction. Total RNA libraries were constructed using the Illumina TruSeq Stranded Total RNA Gold kit (catalog #20020599) and barcoded using Illumina TruSeq RNA unique dual indexes (catalog #20022371). After re-evaluating library quality using TapeStation, individually barcoded libraries were sent to Novogene for next-generation sequencing. After passing additional QC, libraries were multiplexed and subjected to paired-end 150 bp read sequencing on an Illumina NovaSeq S4 Flowcell to a median depth of 100 million total reads (>30 G) per library.

RNA-seq processing and analysis of coronary samples

The raw passed filter sequencing reads obtained from Novogene were demultiplexed using the bcl2fastq script. The quality of the reads was assessed using FASTQC and the adapter sequences were trimmed using trimgalore. Trimmed reads were aligned to the hg38 human reference genome using STAR v2.7.3a according to the GATK Best Practices for RNA-seq. To increase mapping efficiency and sensitivity, novel splice junctions discovered in a first alignment pass with high stringency were used as annotation in a second pass to permit lower stringency alignment and therefore increase sensitivity. PCR duplicates were marked using Picard and WASP was used to filter reads prone to mapping bias. Total read counts and RPKM were calculated with RNA-SeQC v1.1.8 using default parameters and additional flags “-n 1000 -noDoC -strictMode” and GENCODE v30 reference annotation.

Projection of coronary artery datasets with carotid clusters

To integrate two heterogeneous datasets, we have first used the anchor-based data integration algorithm[19]. Next, we created a pairwise correlation matrix between individual samples using the Pearson correlation. Then for each coronary sample, we assigned the cluster identity of the closest (best positively correlated) carotid sample.

Measurement of circulating biomarkers

In 386 selected patients from the Athero-Express cohort, we used a commercially available multiplex proximity extension assay[52] from the Olink proteomics AB platform (Uppsala, Sweden) to measure 276 proteins using the Olink® Cardiovascular II (n=386), Olink® Cardiovascular III (n=386), and Olink® Cardiometabolic (n=208) panels. These panels were selected for their known associations with CV disease. Proteins are expressed on a log2-scale as normalized protein expression (NPX) values. Patients were randomly distributed across plates.

Genotyping and weighted polygenic scores calculation

DNA isolation and genotyping—We genotyped the AE in three separate but consecutive experiments[53]. The DNA was extracted from EDTA blood or (when no blood was available) plaque samples using in-house validated protocols and genotyped in 3 batches (Athero-Express Genomics Studies). For the Athero-Express Genomics Study 1 (AEGS1), included between 2002 and 2007, were genotyped (440,763 markers) using the Affymetrix Genome-Wide Human SNP Array 5.0 (SNP5) chip (Affymetrix Inc., Santa Clara, CA, USA) at Eurofins Genomics, https://www.eurofinsgenomics.eu/). For the Athero-Express

Genomics Study 2 (AEGS2), included between 2002 and 2013, were genotyped (587,351 markers) using the Affymetrix Axiom® GW CEU 1 Array (AxM) at the Helmholtz Genome Analysis Center (https://www.helmholtz-muenchen.de/no_cache/gac/index.html). For the Athero-Express Genomics Study 3 (AEGS3), included between 2002 and 2016, were genotyped (693,931 markers) using the Illumina GSA MD v1 BeadArray (GSA) at Human Genomics Facility, HUGE-F (http://glimdna.org/index.html). All experiments were carried out according to OECD standards and as advised by the respective manufacturer. We used the genotyping calling algorithms as advised by Affymetrix (BRLMM-P for AEGS1 and AxiomGT1 for AEGS2) and Illumina (Illumina GenomeStudio For AEGS3).

Quality control after genotyping—After genotype calling, we adhered to community standard quality control and assurance (QCA) procedures of the genotype data from AEGS1, AEGS2, and AEGS3[54]. Samples with low average genotype calling and sex discrepancies (compared to the clinical data available) were excluded. The data was further filtered on: 1) individual (sample) call rate > 97%, 2) SNP call rate > 97%, 3) minor allele frequencies (MAF) > 3%, 4) average heterozygosity rate ± 3.0 cs.d., 5) relatedness (pi-hat > 0.20), 6) Hardy–Weinberg Equilibrium (HWE $p < 1.0\times10^{-3}$, 7) Monomorphic SNPs (MAF< 1.0×10^{-6}), and 8) deviation in the principal component analysis plot using 1000G phase 3 as reference (6 iterations ± 3s.d.).

Imputation—Before phasing using SHAPEIT2[55], data was lifted to genome build b37 using the liftOver tool from UCSC (https://genome.ucsc.edu/cgi-bin/hgLiftOver). Finally, data were imputed with 1000G phase 3, version 5 and HRC release 1.1 as a reference using the Michigan Imputation Server (https://imputationserver.sph.umich.edu/). These results were further integrated using QCTOOL v2 (https://www.well.ox.ac.uk/~gav/qctool_v2/), where HRC imputed variants are given precedence over 1000G phase 3 imputed variants. After imputation, we compared the quality of the three AEGS datasets based on sample type (EDTA blood or plaque) and genotyping chip used. We checked identity-by-descent (IBD) within and between datasets to aid in sample mixups, duplicate sample use, and relatedness.

Weighted polygenic score calculation—We estimated the weighted polygenic cardiovascular disease susceptibility using the previously published polygenic risk score for coronary artery disease (MetaGRS) described before[56]. Briefly, the MetaGRS comprises 1,745,179 genetic variants with a minor allele frequency (MAF) > 0.1% associated with CAD and was constructed through meta-analysis of three genomic risk scores: GRS46K (comprising 46,000 cardiometabolic genetic variants), FDR202 (including 202 genetic variants associated with CAD at false discovery rate $p < 0.05$ in the recent GWAS CARDIoGRAMplusC4D), and the 1000Genomes genetic score also created with CARDIoGRAMplusC4D. The MetaGRS was internally and externally validated for the primary risk of prevalent and incident CAD in the UK Biobank[63]. We matched the 1.7 million variants from the MetaGRS to 1,742,593 variants in our data (2586 variants were not present in our data). Given that the median imputation quality was high (INFO = 0.978 [IQR 0.945–0.991]), and the variants included in the MetaGRS have MAF >0.1%, we did not further filter on imputation quality. Moreover, since we used the imputed genotype probabilities to calculate the MetaGRS, rather than the hard-coded genotypes, bias arising

from imputation error, i.e., low imputation quality, will only reduce predictive accuracy. Thus, we calculated the weighted polygenic score (PGS) for each included patient in this study using PRSice-2[57] as follows. To account for the imputation quality, we used the allelic dosages (D) estimated by IMPUTE2 based on the posterior genotype probabilities (P_g) for the B-allele (B) for the i^{th} variant. Thus:

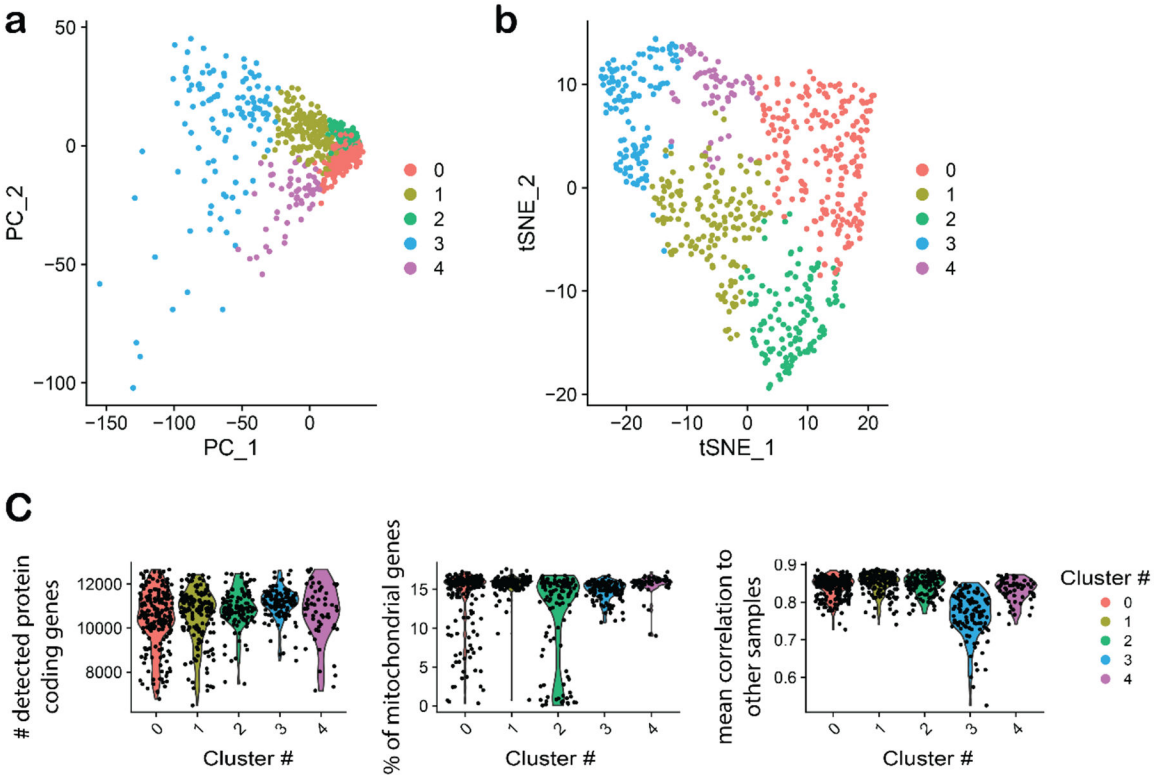
$$D_i = 2 \times P_g(\mathrm{BB}) + 1 \times P_g(\mathrm{AB}) + 0 \times P_g(\mathrm{AA})$$

For each individual, we calculated the aggregate polygenic scores using the effect D of all modeled variants weighted by the effect size (β) of the i^{th} variant as given in the MetaGRS. Thus, for each individual n, the weighted PGS is the sum of the β of the i^{th} variant multiplied by the dosage D of that respective variant:

$$polygenic\ score = \sum_{i}^{n} \beta_i \times D_i$$

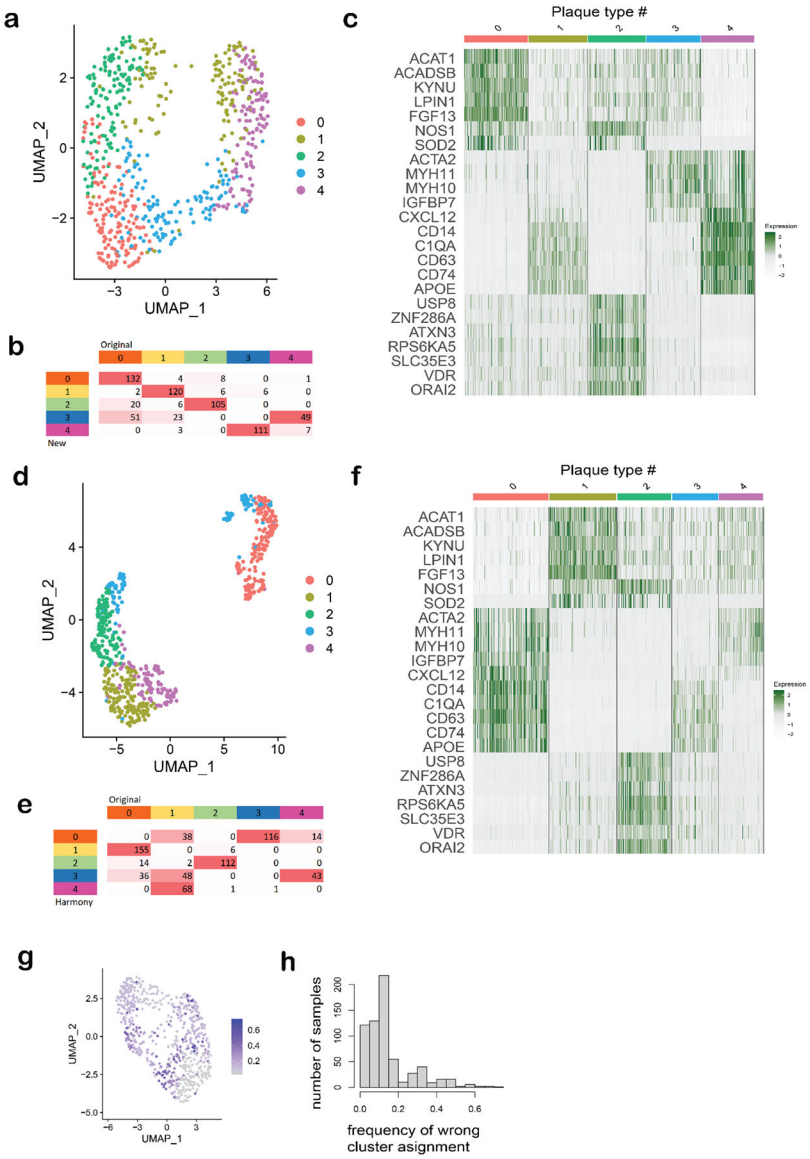
We standardized the PGS to mean-zero and unit-variance for each genotyping batch separately.

Extended Data



Extended Data Figure 1. Unsupervised clustering of plaques based on transcriptomics data.

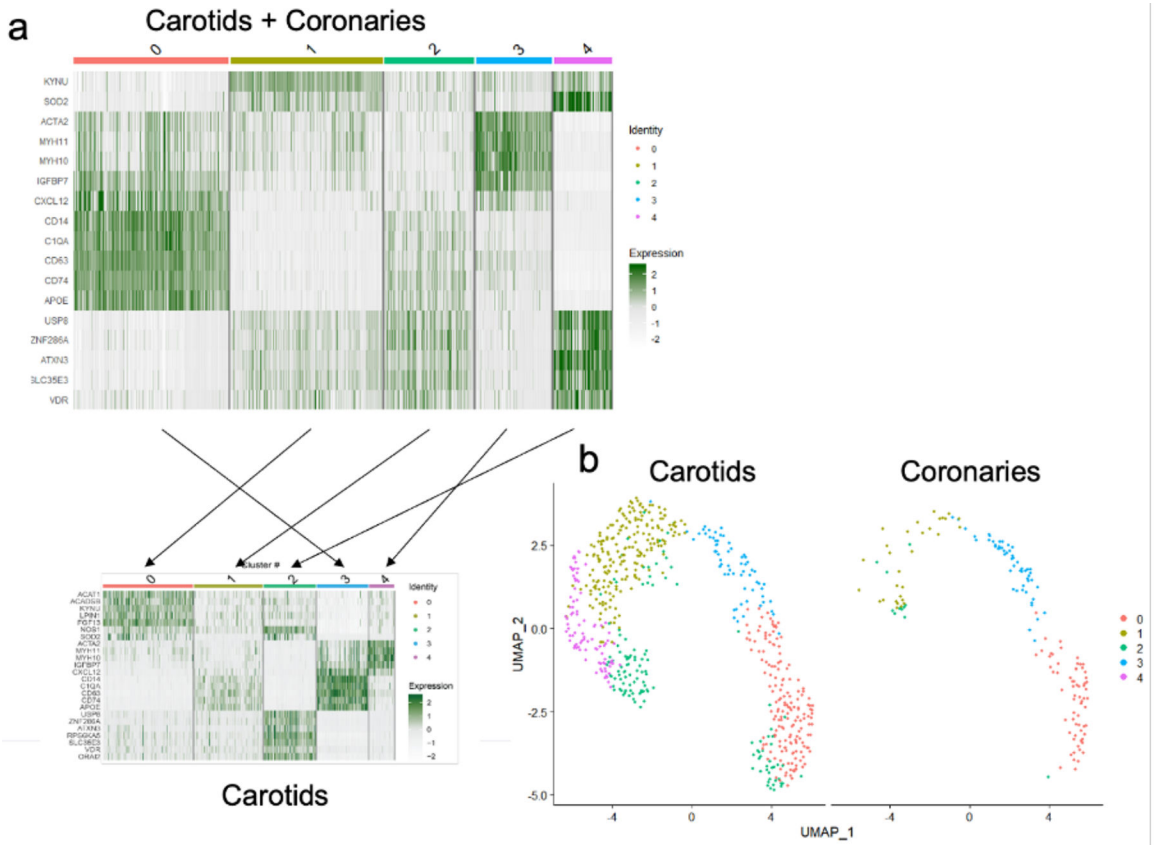
a) PCA plot and **b)** tSNE projection of the 654 plaque samples based on RNA-seq dataset. The color indicates the cluster corresponding to the plaque type cluster from the SNN modularity optimization based clustering algorithm. **c)** Distribution of non-ribosomal protein-coding genes with annotated HGNC name; reads mapping to mitochondrial genes and mean Pearson correlation of samples per cluster



Extended Data Figure 2. Robustness of the clustering.

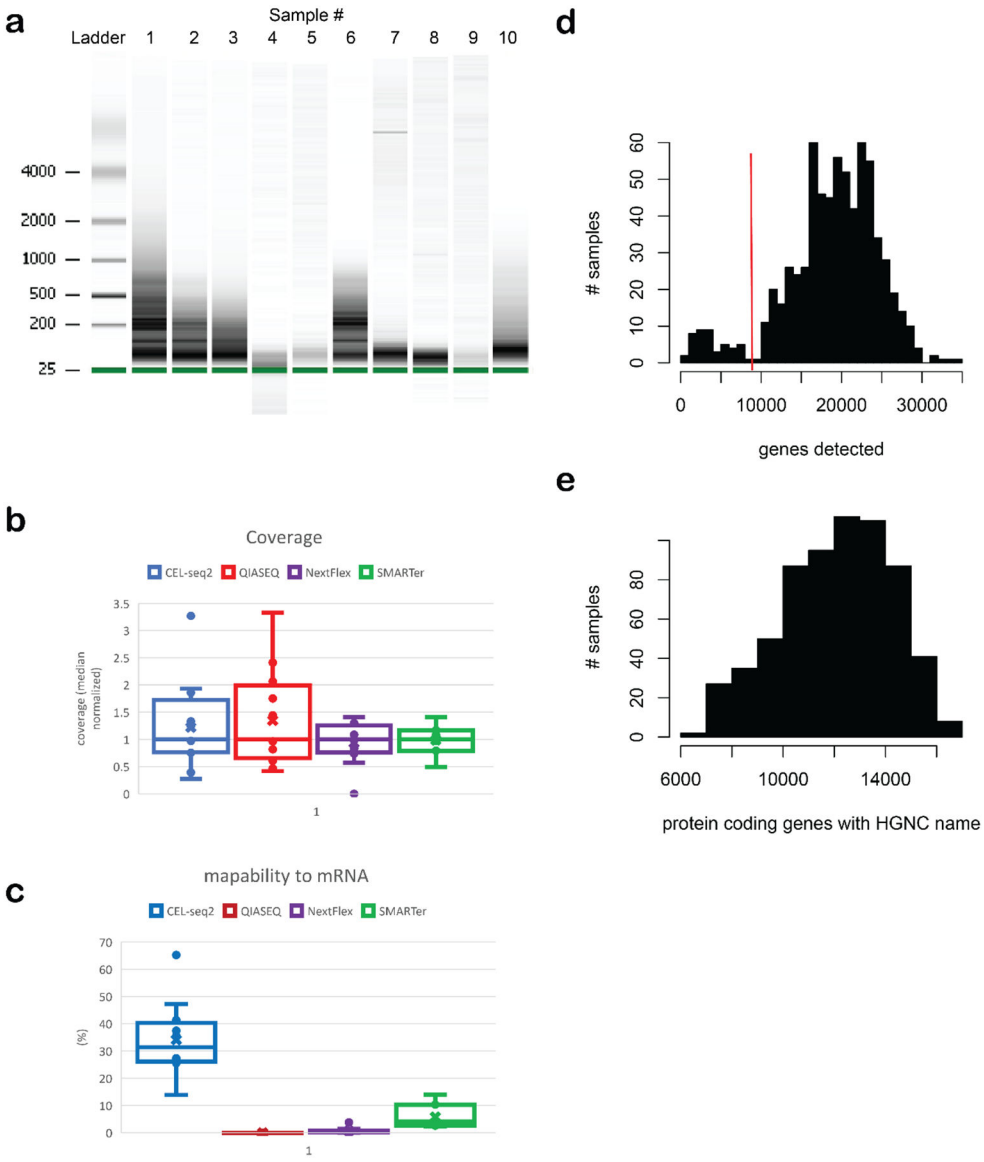
a) UMAP projection of the 654 plaque samples based on RNA-seq dataset. The color indicates the cluster corresponding to the plaque type cluster from the SNN modularity optimization-based clustering algorithm after batch and hospital correction. b) Heatmap depicts relative gene expression levels of selected plaque type enriched genes in individual samples and plaque clusters. c) Correspondence of clusters derived from batch corrected and original dataset. Numbers indicate sample counts in the intersection of the corresponding

clusters. d) UMAP projection of the 654 plaque samples based on RNA-seq dataset. The color indicates the cluster corresponding to the plaque type cluster from the SNN modularity optimization-based clustering algorithm after batch and hospital correction using Harmony integration. e) Heatmap depicts relative gene expression levels of selected plaque type enriched genes in individual samples and plaque clusters. f) Correspondence of clusters derived from batch corrected (using Harmony) and original dataset. Numbers indicate sample counts in the intersection of the corresponding clusters. g) UMAP projection of the 654 plaque samples based on RNA-seq dataset. The color indicates the frequency of the sample being assigned in other than the original cluster in permutation analysis. h) Distribution of the frequencies of the sample being assigned in other than the original cluster in permutation analysis.



Extended Data Figure 3. Integrative analysis of coronary and carotid dataset

a) Heatmaps depicting relative gene expression levels of selected plaque type enriched genes in individual samples and plaque clusters (upper panel - combined coronary and carotid dataset, lower panel – original carotid dataset). b) UMAP projection of the coronary and carotid sample based on RNA-seq data. The color indicates the cluster corresponding to the plaque type cluster from the SNN modularity optimization based clustering algorithm



Extended Data Figure 4. Plaque transcriptomics.

a) representative bioanalyzer profiles of total RNA isolated from 10 samples of advanced atherosclerotic lesions. **b)** Distribution of sequencing reads between samples using four different library preparation strategies. **c)** Percentage of sequenced reads mapped to annotated genes using four different library preparation strategies. **d)** Number of annotated genes identified per sample with at least one mapped read. Samples with less than 9000 genes were excluded from the analysis. **e)** Number of non-ribosomal protein-coding genes with annotated HGNC name per sample used in the analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by The Dutch Heart Foundation (CVON2017-20: Generating the best evidence-based pharmaceutical targets and drugs for atherosclerosis [GENIUS II] to G. Pasterkamp, S.W. van der Laan, M.P.J. de Winther; Targeting Macrophages in Atherosclerotic Disease (T-MAD) to M.P.J. de Winther); Fondation Leducq (Transatlantic Network Grant PlaqOmics to N. J. Leeper, M. Civelek, G. K. Owens, A. V. Finn, J.L.M. Bjorkegren, C. L. Miller and G. Pasterkamp; Transatlantic Network LEAN to M.P.J. de Winther); EU 755320 Taxinomisis grant (E. Pavlos, E. Andreakos, G.J. de Borst, A. Boltjes, G. Pasterkamp); Dutch Research Council (NWO) VENI grant (VI.VENI.212.196) to K.H.M. Prange. We acknowledge the European Research Area Network on Cardiovascular Diseases (ERA-CVD, grant number 01KL1802 to F. W. Asselbergs, S.W. van der Laan); the ERA-Endless consortium (Dutch Heart Foundation, grant number 2017/T099 to H.M. den Ruijter and G. Pasterkamp), European Research Council (ERC) consolidator grant (grant number 866478 UCARE to H.M. den Ruijter), and National Institutes of Health (NIH) grant (R01HL148239 to C.L. Miller). Folkert Asselbergs is supported by UCL Hospitals NIHR Biomedical Research Centre. The authors would like to thank the Utrecht Sequencing Facility for continuous support and patience.

Competing Interests Statement

CLM has received funding support from AstraZeneca for work unrelated to this study. GP received funding support from Roche to partly cover the generation of biomarker data. The remaining authors declare no competing interests.

Data availability

The raw RNA-seq data from the Athero-Express cohort are not publicly available due to research participant privacy/consent. Data can be accessed via DataverseNL at this address: https://doi.org/10.34894/D1MDKL. There are restrictions on use by commercial parties, and on sharing openly based on (inter)national laws and regulations and the written informed consent. Therefore these data (and additional clinical data) are only available upon discussion and signing a Data Sharing Agreement (see Terms of Access in DataverseNL) and within a specially designed UMC Utrecht provided environment.

The individual genes from the dataset can be queried and accessed using the PlaqView portal (https://www.plaqview.com/).

The raw and processed bulk RNA-seq data from the coronary artery tissues will be made available on GEO as well as on plaqview.com upon publication of the primary manuscript describing these data. In the meantime, requests to access these data can be addressed to Clint L. Miller clintm@virginia.edu.

References

1. Falk E Plaque rupture with severe pre-existing stenosis precipitating coronary thrombosis. Characteristics of coronary atherosclerotic plaques underlying fatal occlusive thrombi. Br. Heart J 50, 127–134 (1983). [PubMed: 6882602]
2. Davies MJ & Thomas AC Plaque fissuring--the cause of acute myocardial infarction, sudden ischaemic death, and crescendo angina. Br. Heart J 53, 363–373 (1985). [PubMed: 3885978]
3. Franzén O et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. Science 353, 827–830 (2016). [PubMed: 27540175]
4. Hartman RJG et al. Sex-Stratified Gene Regulatory Networks Reveal Female Key Driver Genes of Atherosclerosis Involved in Smooth Muscle Cell Phenotype Switching. Circulation 143, 713–726 (2021). [PubMed: 33499648]
5. Perisic L et al. Gene expression signatures, pathways and networks in carotid atherosclerosis. J. Intern. Med 279, 293–308 (2016). [PubMed: 26620734]

6. Shakhtshneider EV et al. RNA-SEQ transcriptome analysis of stable and unstable atherosclerotic plaques. Atherosclerosis 315, e133 (2020).
7. Aure MR et al. Integrative clustering reveals a novel split in the luminal A subtype of breast cancer with impact on outcome. Breast Cancer Res. 19, 44 (2017). [PubMed: 28356166]
8. Zhao L, Zhao H & Yan H Gene expression profiling of 1200 pancreatic ductal adenocarcinoma reveals novel subtypes. BMC Cancer 18, 603 (2018). [PubMed: 29843660]
9. Lapointe J et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc. Natl. Acad. Sci. U. S. A 101, 811–816 (2004). [PubMed: 14711987]
10. Bertucci F et al. Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. Cancer Res. 65, 2170–2178 (2005). [PubMed: 15781628]
11. Korsunsky I et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296 (2019). [PubMed: 31740819]
12. Depuydt MAC et al. Microanatomy of the Human Atherosclerotic Plaque by Single-Cell Transcriptomics. Circ. Res 127, 1437–1455 (2020). [PubMed: 32981416]
13. Costales P et al. K Domain CR9 of Low Density Lipoprotein (LDL) Receptor-related Protein 1 (LRP1) Is Critical for Aggregated LDL-induced Foam Cell Formation from Human Vascular Smooth Muscle Cells. J. Biol. Chem 290, 14852–14865 (2015). [PubMed: 25918169]
14. Erdmann J, Kessler T, Munoz Venegas L & Schunkert H A decade of genome-wide association studies for coronary artery disease: the challenges ahead. Cardiovasc. Res 114, 1241–1257 (2018). [PubMed: 29617720]
15. Timmerman N et al. Pre-Operative Plasma Extracellular Vesicle Proteins are Associated with a High Risk of Long Term Secondary Major Cardiovascular Events in Patients Undergoing Carotid Endarterectomy. Eur. J. Vasc. Endovasc. Surg 62, 705–715 (2021). [PubMed: 34511318]
16. de Leeuw CA, Mooij JM, Heskes T & Posthuma D MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol 11, e1004219 (2015). [PubMed: 25885710]
17. Hartiala JA et al. Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. Eur. Heart J 42, 919–933 (2021). [PubMed: 33532862]
18. Turner AW et al. Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. Nat. Genet 54, 804–816 (2022). [PubMed: 35590109]
19. Stuart T et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21 (2019). [PubMed: 31178118]
20. Khoukaz HB et al. Drug Targeting of Plasminogen Activator Inhibitor-1 Inhibits Metabolic Dysfunction and Atherosclerosis in a Murine Model of Metabolic Syndrome. Arterioscler. Thromb. Vasc. Biol 40, 1479–1490 (2020). [PubMed: 32268785]
21. Josefs T et al. Neutrophil extracellular traps promote macrophage inflammation and impair atherosclerosis resolution in diabetic mice. JCI Insight 5, (2020).
22. Schumski A et al. Endotoxinemia Accelerates Atherosclerosis Through Electrostatic Charge-Mediated Monocyte Adhesion. Circulation 143, 254–266 (2021). [PubMed: 33167684]
23. Döring Y, Libby P & Soehnlein O Neutrophil Extracellular Traps Participate in Cardiovascular Diseases: Recent Experimental and Clinical Insights. Circ. Res 126, 1228–1241 (2020). [PubMed: 32324499]
24. Silvestre-Roig C et al. Externalized histone H4 orchestrates chronic inflammation by inducing lytic cell death. Nature 569, 236–240 (2019). [PubMed: 31043745]
25. Gomez I et al. Neutrophil microvesicles drive atherosclerosis by delivering miR-155 to atheroprone endothelium. Nat. Commun 11, 214 (2020). [PubMed: 31924781]
26. Tomas L et al. Altered metabolism distinguishes high-risk from stable carotid atherosclerotic plaques. Eur. Heart J 39, 2301–2310 (2018). [PubMed: 29562241]
27. Newman AAC et al. Multiple cell types contribute to the atherosclerotic lesion fibrous cap by PDGFRβ and bioenergetic mechanisms. Nat Metab 3, 166–181 (2021). [PubMed: 33619382]
28. Shi J, Yang Y, Cheng A, Xu G & He F Metabolism of vascular smooth muscle cells in vascular diseases. Am. J. Physiol. Heart Circ. Physiol 319, H613–H631 (2020). [PubMed: 32762559]
29. De Bock K et al. Role of PFKFB3-driven glycolysis in vessel sprouting. Cell 154, 651–663 (2013). [PubMed: 23911327]

30. Kang H, Kim H, Lee S, Youn H & Youn B Role of Metabolic Reprogramming in Epithelial-Mesenchymal Transition (EMT). Int. J. Mol. Sci 20, (2019).
31. Cornelissen A, Guo L, Sakamoto A, Virmani R & Finn AV New insights into the role of iron in inflammation and atherosclerosis. EBioMedicine 47, 598–606 (2019). [PubMed: 31416722]
32. Sakamoto A et al. Abstract 13413: CD163+ Alternative Macrophage Inhibits Atherosclerotic Calcification via Enhancement of Hyaluronan Production in Vascular Smooth Muscle Cells. Circulation 140, A13413–A13413 (2019).
33. Mega JL et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet 385, 2264–2271 (2015). [PubMed: 25748612]
34. Fateh-Moghadam S et al. Platelet degranulation is associated with progression of intima-media thickness of the common carotid artery in patients with diabetes mellitus type 2. Arterioscler. Thromb. Vasc. Biol 25, 1299–1303 (2005). [PubMed: 15817881]
35. Li X et al. The Regulation of Exosome-Derived miRNA on Heterogeneity of Macrophages in Atherosclerotic Plaques. Front. Immunol 11, 2175 (2020). [PubMed: 33013913]
36. Bhatia VK et al. Complement C1q reduces early atherosclerosis in low-density lipoprotein receptor-deficient mice. Am. J. Pathol 170, 416–426 (2007). [PubMed: 17200212]
37. Tomaniak M et al. Vulnerable plaques and patients: state-of-the-art. Eur. Heart J 41, 2997–3004 (2020). [PubMed: 32402086]
38. Lind L et al. Plasma Protein Profile of Carotid Artery Atherosclerosis and Atherosclerotic Outcomes. Arterioscler. Thromb. Vasc. Biol 41, 1777–1788 (2021). [PubMed: 33657885]
39. Hao Y et al. Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29 (2021). [PubMed: 34062119]
40. Nurmohamed NS et al. Targeted proteomics improves cardiovascular risk prediction in secondary prevention. Eur. Heart J 43, 1569–1577 (2022). [PubMed: 35139537]
41. Verhoeven BAN et al. Athero-express: differential atherosclerotic plaque expression of mRNA and protein in relation to cardiovascular events and patient characteristics. Rationale and design. Eur. J. Epidemiol 19, 1127–1133 (2004). [PubMed: 15678794]
42. Hellings WE et al. Composition of Carotid Atherosclerotic Plaque Is Associated With Cardiovascular Outcome. Circulation 121, 1941–1950 (2010). [PubMed: 20404256]
43. Hellings WE et al. Atherosclerotic plaque composition and occurrence of restenosis after carotid endarterectomy. JAMA 299, 547–554 (2008). [PubMed: 18252885]
44. van Lammeren GW et al. Time-dependent changes in atherosclerotic plaque composition in patients undergoing carotid surgery. Circulation 129, 2269–2276 (2014). [PubMed: 24637558]
45. Hashimshony T et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. 17, 77 (2016). [PubMed: 27121950]
46. Waltman L & van Eck NJ A smart local moving algorithm for large-scale modularity-based community detection. Eur. Phys. J. B 86, 471 (2013).
47. satijalab/seurat: R toolkit for single cell genomics - GitHub. https://github.com›satijalab›seurathttps://github.com›satijalab›seurat.
48. Yu G & He Q-Y ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol. Biosyst 12, 477–479 (2016). [PubMed: 26661513]
49. Bennett BJ et al. Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. PLoS Genet. 11, e1005711 (2015). [PubMed: 26694027]
50. Dai X et al. Betaine Supplementation Attenuates S-Adenosylhomocysteine Hydrolase-Deficiency-Accelerated Atherosclerosis in Apolipoprotein E-Deficient Mice. Nutrients 14, (2022).
51. Ikeda J et al. Radiation Impacts Early Atherosclerosis by Suppressing Intimal LDL Accumulation. Circ. Res 128, 530–543 (2021). [PubMed: 33397122]
52. Gullberg M et al. Cytokine detection by antibody-based proximity ligation. Proc. Natl. Acad. Sci. U. S. A 101, 8420–8424 (2004). [PubMed: 15155907]
53. van der Laan SW et al. Genetic susceptibility loci for cardiovascular disease and their impact on atherosclerotic plaques. Circ. Genom. Precis. Med 11, e002115 (2018). [PubMed: 30354329]

54. Laurie CC et al. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet. Epidemiol 34, 591–602 (2010). [PubMed: 20718045]

55. Delaneau O, Marchini J & Zagury J-F A linear complexity phasing method for thousands of genomes. Nat. Methods 9, 179–181 (2011). [PubMed: 22138821]

56. Inouye M et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J. Am. Coll. Cardiol 72, 1883–1893 (2018). [PubMed: 30309464]

57. Choi SW & O'Reilly PF PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience 8, (2019).

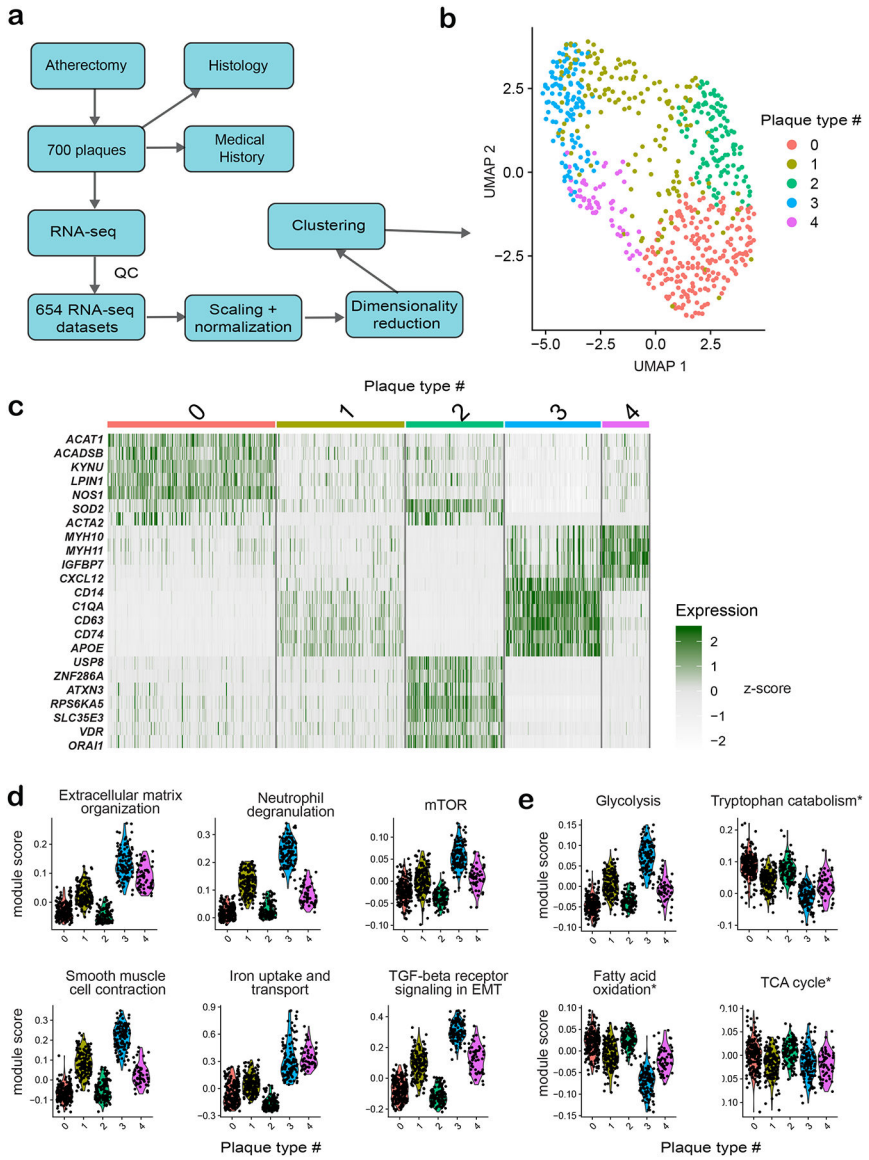


Figure 1. Unsupervised clustering of plaques based on transcriptomics data.

a) A schematic workflow of the clustering analysis. **b)** UMAP projection of the 654 plaque samples based on RNA-seq dataset. The color indicates the cluster corresponding to the plaque type from the SNN modularity optimization-based clustering algorithm. **c)** Heatmap depicts relative gene expression levels of selected plaque type enriched genes in individual samples and plaque clusters. **d)** Module scores of genes annotated to selected molecular pathways in different plaque clusters. **e)** Module scores of genes annotated to selected metabolic pathways in different plaque clusters. * depicts empirically selected pathways that were not significantly enriched in the pathway analysis.

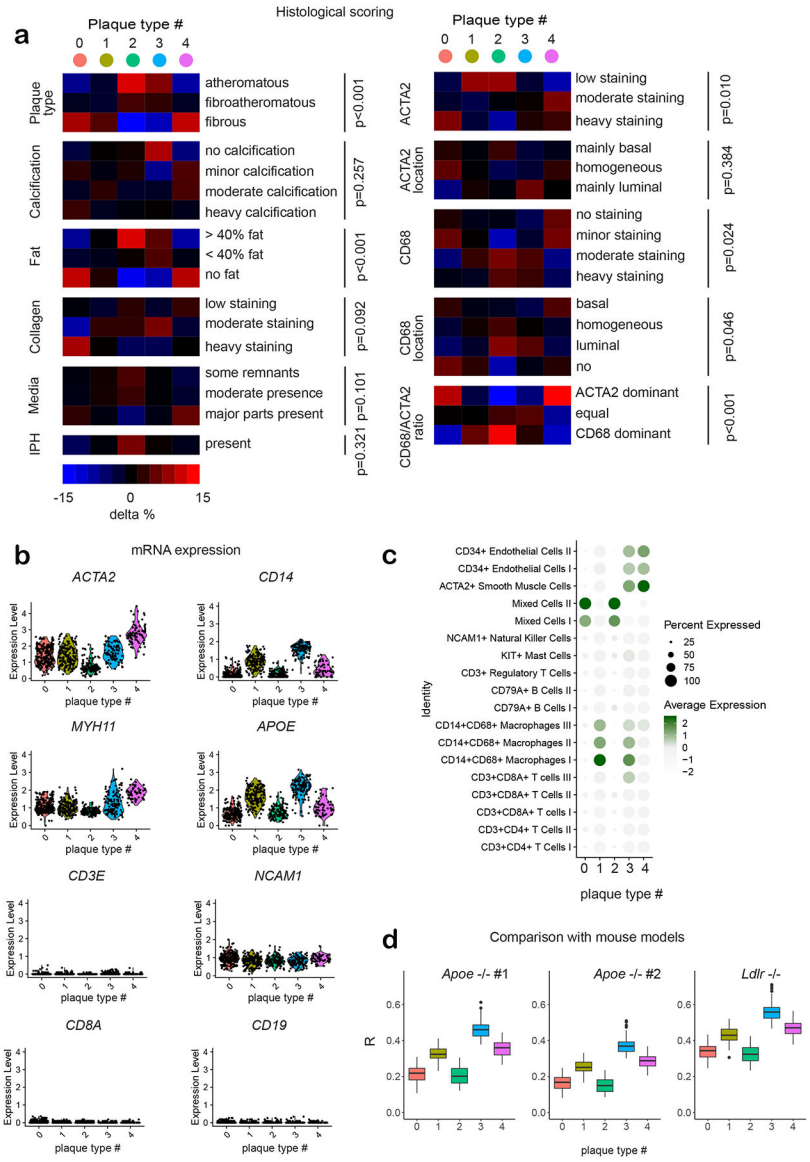


Figure 2. Unsupervised clustering of plaques based and cell composition.

a) Distribution of histological features in five transcriptome-based clusters. The percentage represents the difference over equal distribution. (IPH - intraplaque hemorrhage). P-values were calculated using the chi-square test for categorical variables and one-way analysis of variance (ANOVA) for continuous variables. Categories with less than 15 samples were omitted. Exact proportions can be found in Supplementary Table 4 **b)** Expression of selected cell marker genes in transcriptome-based clusters. **c)** Expression of cluster-specific gene sets in single-cell transcriptomics datasets from atherosclerotic plaques. **d)** Correlation of transcriptomic profiles between human plaques and bulk RNAseq data of plaques from the different mice models: *Apoe* −/− (n=20 animals and n=9 animals) and *Ldlr* −/− (n=3 animals). Boxplot's top, middle, and bottom lines represent values at 25th, 50th, and 75th percentile. Whiskers extend up to 1.5 times the interquartile range from the top (bottom) of the box to the furthest data point within that distance.

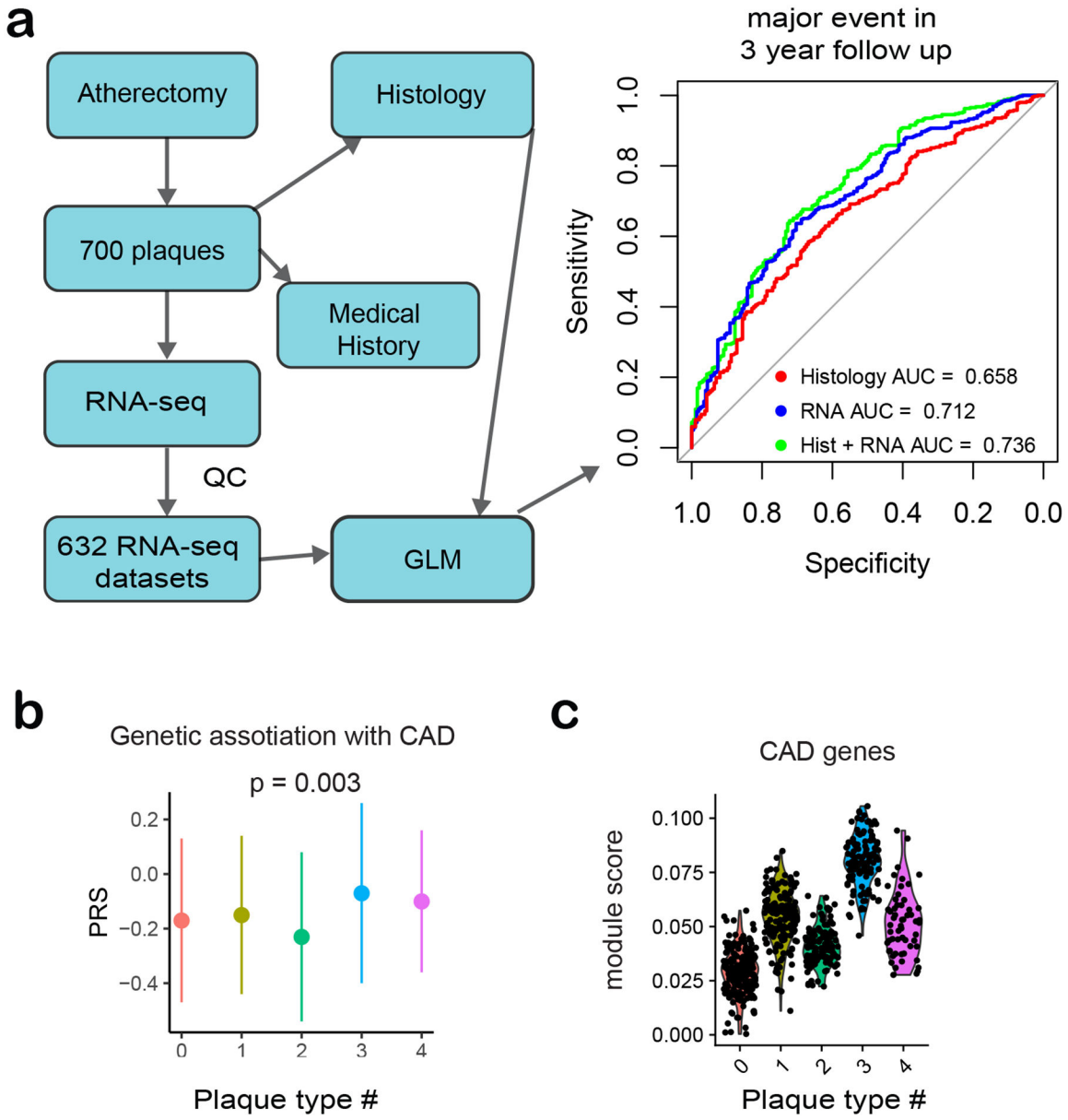
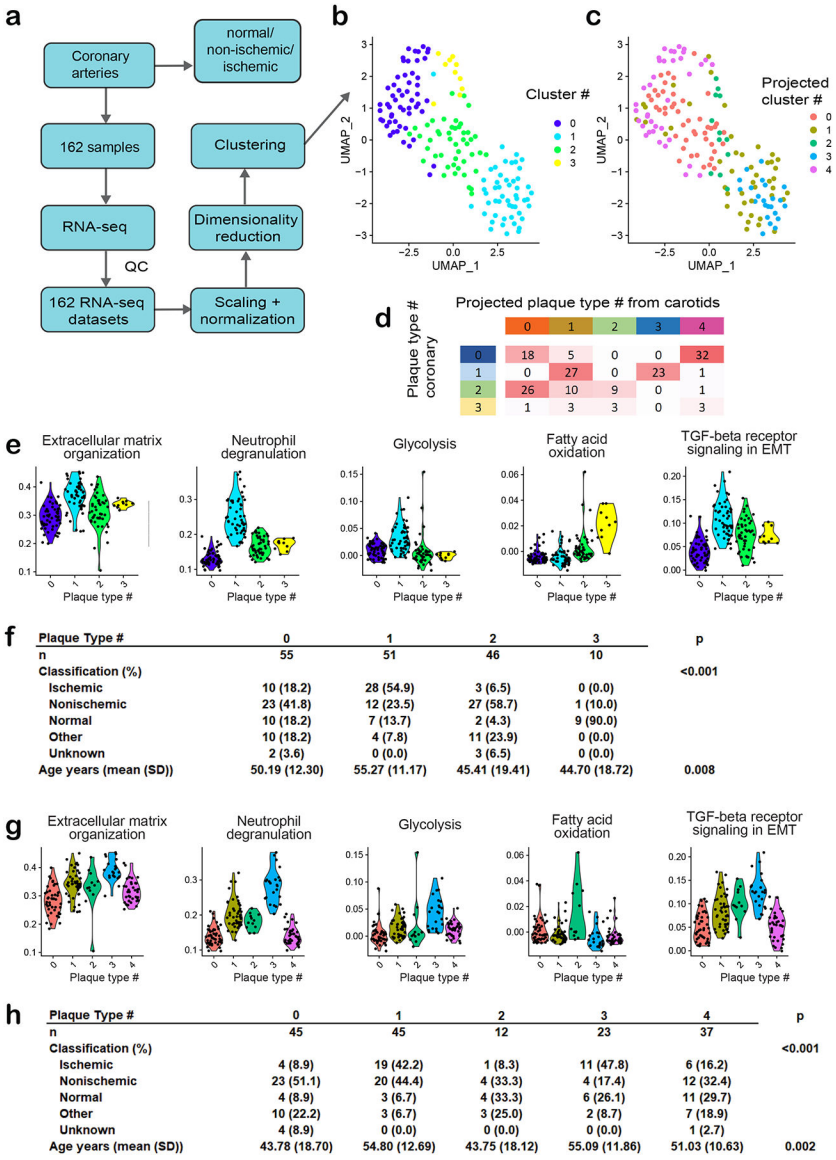


Figure 3. Association of molecular plaque types with secondary risk and genetic risk

a) Workflow of the generalized linear model (GLM) analysis. Receiver operating characteristic (ROC) curve depicting the ability of GLM to discriminate between the occurrence of the primary outcome within a 3-year follow-up time using histological data (red), transcriptomics data (blue), or combined histological and transcriptomics data (green). **b)** Distribution of polygenic risk scores (PRS) for coronary artery disease (CAD) among patients (n=632) from different clusters. Error bars represent the standard deviation and are centered around the mean value. P-value is calculated using one-way ANOVA **c)** Module scores of genes genetically associated with CAD in different plaque clusters



Plaque Type #	0	1	2	3	p
n	55	51	46	10	
Classification (%)					<0.001
Ischemic	10 (18.2)	28 (54.9)	3 (6.5)	0 (0.0)	
Nonischemic	23 (41.8)	12 (23.5)	27 (58.7)	1 (10.0)	
Normal	10 (18.2)	7 (13.7)	2 (4.3)	9 (90.0)	
Other	10 (18.2)	4 (7.8)	11 (23.9)	0 (0.0)	
Unknown	2 (3.6)	0 (0.0)	3 (6.5)	0 (0.0)	
Age years (mean (SD))	50.19 (12.30)	55.27 (11.17)	45.41 (19.41)	44.70 (18.72)	0.008

Plaque Type #	0	1	2	3	4	p
n	45	45	12	23	37	
Classification (%)						<0.001
Ischemic	4 (8.9)	19 (42.2)	1 (8.3)	11 (47.8)	6 (16.2)	
Nonischemic	23 (51.1)	20 (44.4)	4 (33.3)	4 (17.4)	12 (32.4)	
Normal	4 (8.9)	3 (6.7)	4 (33.3)	6 (26.1)	11 (29.7)	
Other	10 (22.2)	3 (6.7)	3 (25.0)	2 (8.7)	7 (18.9)	
Unknown	4 (8.9)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.7)	
Age years (mean (SD))	43.78 (18.70)	54.80 (12.69)	43.75 (18.12)	55.09 (11.86)	51.03 (10.63)	0.002

Figure 4. Unsupervised clustering of coronary artery samples based on transcriptomics data.

a) A schematic workflow of the clustering analysis. **b)** UMAP projection of the 162 coronary samples based on RNA-seq dataset. The color indicates the cluster corresponding to the plaque type from the SNN modularity optimization-based clustering algorithm. **c)** UMAP projection of the 162 coronary samples based on RNA-seq dataset. The color indicates the projected plaque type identity from the carotid dataset. **d)** Correspondence of clusters derived from coronary data and projected clusters from carotids. Numbers in the table indicate sample counts in the intersection of the corresponding clusters. **e)** Module scores of genes annotated to selected molecular pathways in individual plaque clusters derived from coronary arteries. **f)** Distribution of clinical status and age of sample donors in coronary clusters. **g)** Module scores of genes annotated to selected molecular pathways in individual plaque clusters projected from carotid arteries. **h)** Distribution of clinical status and age of coronary artery sample donors in projected carotid clusters. P-values in f) and h)

were calculated using the chi-square test for categorical variables and one-way analysis of variance (ANOVA) for continuous variables.

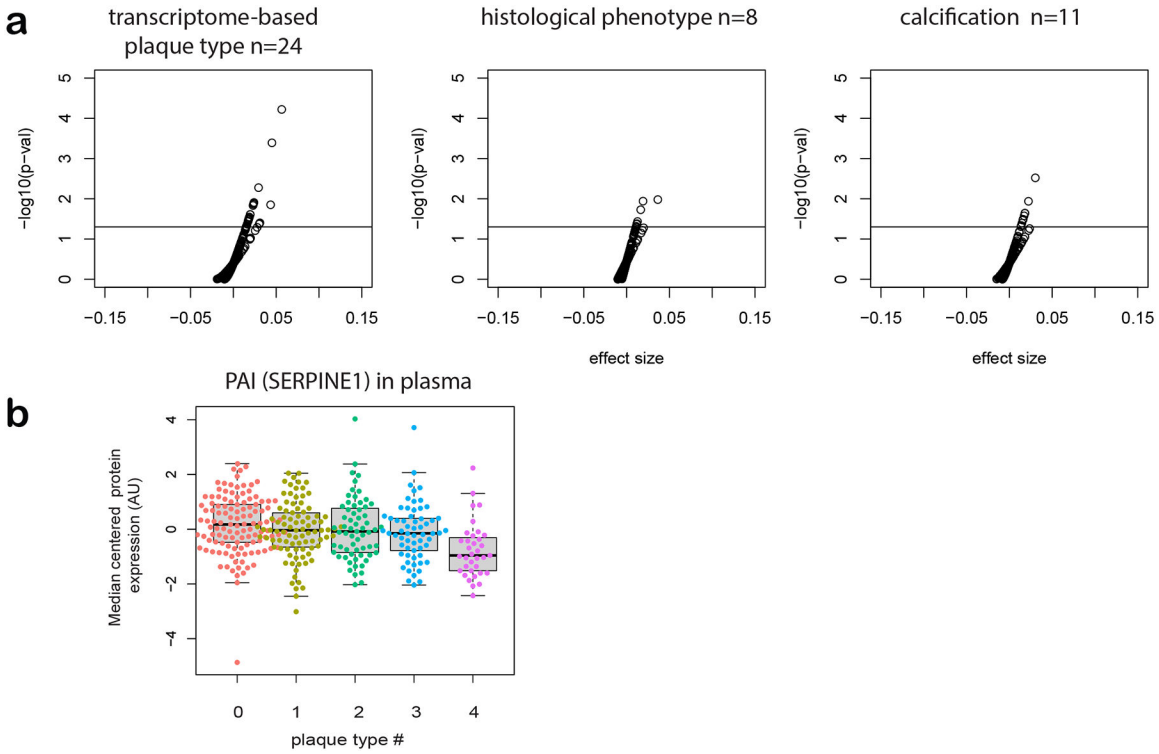


Figure 5. Circulating biomarkers differ between transcriptome-based clusters.

a) Volcano plot showing the statistical significance (using Kruskal-Wallis test) versus effect size (chi-squared statistic) differential distribution of 276 biomarkers between transcriptome clusters or pathological plaque features. “n” represents the number of significant biomarkers. **b)** Plasma levels of plasminogen activator inhibitor (PAI). n = 386, Boxplot’s top, middle, and bottom lines represent values at 25th, 50th, and 75th percentile. Whiskers extend up to 1.5 times the interquartile range from the top (bottom) of the box to the furthest data point within that distance.

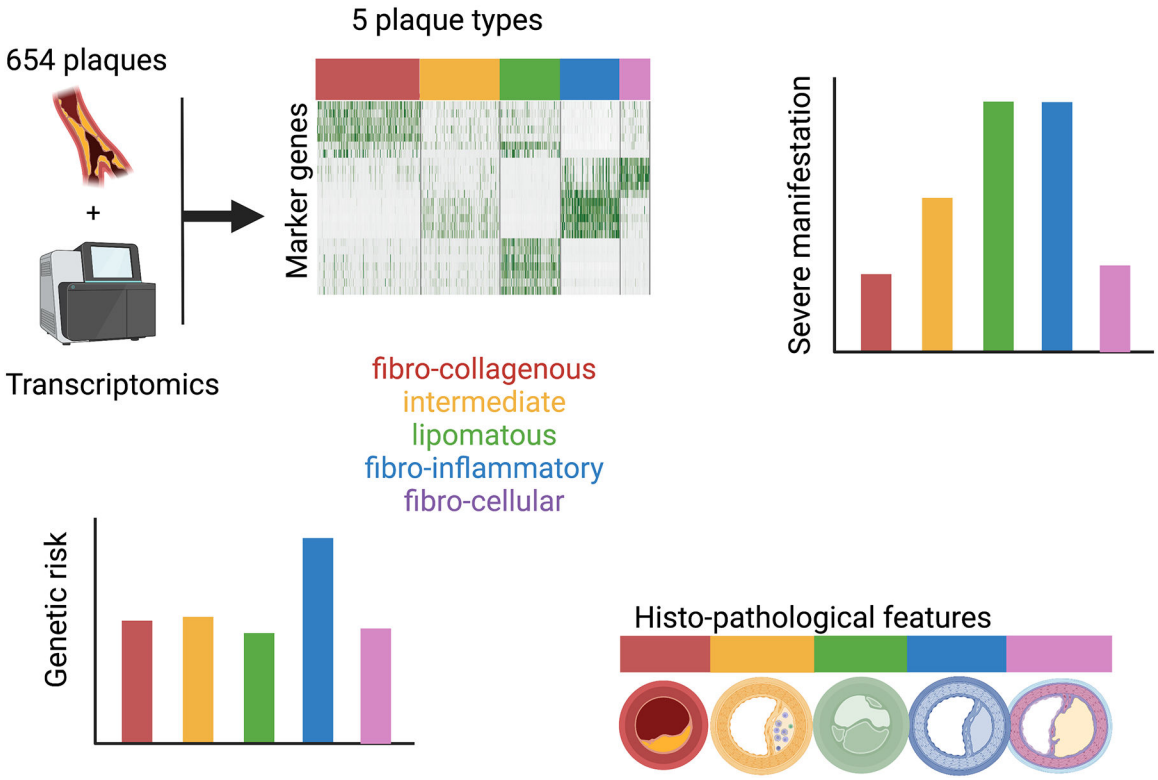


Figure 6. Graphical summary of the study.

Schematic overview representing our main observations. In this transcriptomic study, we were able to differentiate 5 clusters that were associated with clinical manifestations (stroke, transient ischemic attack - TIA, coronary ischemia), genetic risk score, and histo-pathological characteristics.

Table 1

Distribution of clinical parameters in five transcriptome-based molecular plaque types.

Cluster #	0	1	2	3	4	p-value
n	200	147	116	114	55	
Male (%)	144 (72.0)	111 (75.5)	92 (79.3)	89 (78.1)	40 (72.7)	0.584
Number of plaque segments stored (0.5 cm each) (mean (SD))	6.5 (2.5)	6.2 (2.3)	6.6 (2.6)	6.2 (2.1)	6.2 (2.3)	0.424
Smoker current (%)	80 (40.0)	46 (31.9)	36 (31.6)	41 (36.3)	23 (42.6)	0.356
DM (%)	54 (27.0)	28 (19.0)	20 (17.2)	23 (20.2)	11 (20.0)	0.236
CAD history = No history CAD (%)	135 (67.5)	86 (58.5)	83 (71.6)	79 (69.3)	38 (69.1)	0.184
Stroke history = No history of stroke (%)	146 (73.0)	102 (69.4)	73 (62.9)	78 (68.4)	42 (76.4)	0.308
Peripheral arterial occlusive disease = yes (%)	50 (25.0)	40 (27.2)	21 (18.1)	20 (17.5)	8 (14.5)	0.112
Stenosis (%)						0.642
0–50%	0 (0.0)	0 (0.0)	1 (0.9)	1 (0.9)	0 (0.0)	
50–70%	13 (6.7)	7 (5.0)	10 (9.0)	5 (4.5)	4 (7.4)	
70–99%	180 (93.3)	134 (95.0)	100 (90.1)	104 (94.5)	50 (92.6)	
Stenosis contralateral= 50–100% (%)	84 (46.7)	60 (48.0)	46 (46.0)	37 (35.6)	16 (31.4)	0.102
Time between event and surgery [days] (mean (SD))	81.3 (86.7)	62.6 (65.6)	53.3 (65.8)	73.6 (79.9)	64.7 (68.2)	0.057
eGFR [mL/min/1.73m2] (mean (SD))	72.6 (21.4)	72.0 (18.7)	70.2 (22.5)	76.2 (19.3)	75.8 (18.2)	0.175
Hypertension (%)	144 (74.2)	106 (73.1)	82 (71.9)	81 (73.0)	34 (61.8)	0.486
Systolic tension [mmHg] (mean (SD))	153 (26)	153 (26)	159 (25)	155 (23)	152 (22)	0.393
Diastolic tension [mmHg] (mean (SD))	82 (14)	81 (12)	84 (15)	84 (13)	82 (13)	0.457
Symptoms (%)						0.001
asymptomatic	42 (21.1)	17 (12.0)	9 (8.0)	17 (15.0)	10 (18.5)	
ocular	46 (23.1)	28 (19.7)	13 (11.6)	9 (8.0)	11 (20.4)	
stroke	38 (19.1)	36 (25.4)	37 (33.0)	29 (25.7)	10 (18.5)	
TIA	73 (36.7)	61 (43.0)	53 (47.3)	58 (51.3)	23 (42.6)	
Restenosis (%)	3 (1.5)	4 (2.8)	0 (0.0)	4 (3.6)	0 (0.0)	0.207
Emergency (%)	9 (4.6)	4 (2.9)	6 (5.5)	4 (3.5)	0 (0.0)	0.683
Age (mean (SD))	67.2 (8.9)	69.1 (8.8)	70.4 (9.2)	67.5 (8.8)	68.9 (8.3)	0.019
BMI (mean (SD))	27.2 (3.6)	26.4 (4.1)	26.6 (3.2)	26.3 (4.1)	25.8 (3.9)	0.095
Hemoglobin [mmol/L] (mean (SD))	8.72 (1.02)	8.65 (0.88)	8.69 (0.94)	8.81 (0.90)	8.74 (0.91)	0.734
Hematocrit (fraction of erythrocytes in the blood) (mean (SD))	0.42 (0.05)	0.41 (0.04)	0.41 (0.05)	0.42 (0.05)	0.41 (0.04)	0.745
Creatinin [umol/L] (mean (SD))	97.7 (36.4)	96.0 (28.2)	101.1 (31.9)	92.1 (28.3)	89.9 (21.2)	0.123
Homocysteine [umol/L] (mean (SD))	15.3 (11.1)	13.6 (4.5)	18.8 (21.0)	12.9 (4.0)	14.0 (6.1)	0.345
Glucose [mmol/L] (mean (SD))	6.63 (2.41)	6.57 (2.06)	6.51 (1.44)	6.16 (1.91)	6.09 (1.29)	0.386
Cholesterol [mmol/L] (mean (SD))	4.59 (1.27)	4.48 (1.24)	4.11 (1.08)	4.58 (1.32)	4.44 (1.11)	0.050
Triglycerides [mmol/L] (mean (SD))	1.75 (0.95)	1.67 (0.83)	1.70 (1.01)	1.76 (1.17)	1.52 (0.82)	0.688
LDL-cholesterol [mmol/L] (mean (SD))	2.62 (1.01)	2.50 (0.90)	2.39 (0.92)	2.61 (1.10)	2.49 (0.91)	0.464
HDL-cholesterol [mmol/L] (mean (SD))	1.14 (0.38)	1.10 (0.33)	1.03 (0.32)	1.11 (0.28)	1.12 (0.39)	0.162
Plaquephenotype (%)						<0.001
atheromatous	44 (22.2)	40 (27.6)	49 (43.0)	43 (38.1)	11 (20.8)	

Cluster #	0	1	2	3	4	p-value
fibroatheromatous	71 (35.9)	51 (35.2)	48 (42.1)	46 (40.7)	19 (35.8)	
fibrous	83 (41.9)	54 (37.2)	17 (14.9)	24 (21.2)	23 (43.4)	
MACE free (3 years) (%)	176 (88.9)	126 (86.3)	103 (88.8)	95 (83.3)	49 (89.1)	0.629
Acetylsalicylic acid = yes (%)	85 (42.5)	49 (33.6)	33 (28.4)	37 (32.5)	20 (36.4)	0.111
Carbasalate = yes (%)	96 (48.0)	82 (56.2)	68 (58.6)	68 (59.6)	33 (60.0)	0.184
Dipyridamole = yes (%)	98 (49.0)	76 (52.1)	67 (57.8)	59 (51.8)	28 (50.9)	0.682
Clopidogrel (or other ADP inhibitor) = yes (%)	22 (11.0)	20 (13.7)	10 (8.6)	12 (10.5)	8 (14.5)	0.680
Anti platelet drug use = yes (%)	178 (89.0)	130 (89.0)	105 (90.5)	104 (91.2)	47 (85.5)	0.825
RAAS medication use = yes (%)	108 (54.0)	71 (48.6)	58 (50.0)	53 (46.5)	23 (41.8)	0.496
Anti-arrythmic drugs used = yes (%)	9 (4.5)	8 (5.5)	6 (5.2)	6 (5.3)	3 (5.5)	0.995
Statins = yes (%)	143 (71.5)	108 (74.0)	91 (78.4)	87 (76.3)	41 (74.5)	0.714
Diuretic use = yes (%)	75 (37.5)	46 (31.5)	48 (41.4)	38 (33.3)	15 (27.3)	0.290
Beta-blocker use = yes (%)	86 (43.0)	74 (50.7)	50 (43.1)	55 (48.2)	22 (40.0)	0.497
Calcium channel blockers use = yes (%)	58 (29.0)	40 (27.4)	27 (23.3)	35 (30.7)	7 (12.7)	0.103
ACE-inhibitor use = yes (%)	71 (35.5)	45 (30.8)	35 (30.2)	38 (33.3)	15 (27.3)	0.733
Angiotensin II receptor antagonist use = yes (%)	39 (19.5)	29 (19.9)	24 (20.7)	18 (15.8)	10 (18.2)	0.893
Use of one or more antihypertensive drugs = yes (%)	161 (80.5)	113 (77.4)	83 (71.6)	90 (78.9)	41 (74.5)	0.438

P-values were calculated using the chi-square test for categorical variables and one-way analysis of variance (ANOVA) for continuous variables. DM – diabetes mellitus, CAD – coronary artery disease, eGFR - estimated glomerular filtration rate, TIA - transient ischemic attack, MBI - body mass index, LDL - low-density lipoprotein, HDL high-density lipoprotein, MACE - major adverse cardiovascular event, ACE - angiotensin-converting enzyme, RAAS - renin-angiotensin-aldosterone-system