Multi-Organ Expression Profiling Uncovers a Gene Module in Coronary Artery Disease Involving Transendothelial Migration of Leukocytes and LIM Domain Binding 2: The Stockholm Atherosclerosis Gene Expression (STAGE) Study

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version
doi://10.1371/journal.pgen.1000754

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:5360629

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Multi-Organ Expression Profiling Uncovers a Gene Module in Coronary Artery Disease Involving Transendothelial Migration of Leukocytes and LIM Domain Binding 2: The Stockholm Atherosclerosis Gene Expression (STAGE) Study

Sara Hägg1,2,3,9, Josef Skogsberg1,3,9, Jesper Lundström1,2,3,9, Peri Noori1,3, Roland Nilsson2,3, Hua Zhong4, Shohreh Maleki1, Ming-Mei Shang1,3, Björn Brinne2, Maria Bradshaw1,2,3, Vladimir B. Bajic5,6, Ann Samnegård7, Angela Silveira8, Lee M. Kaplan9, Bruna Gigante10, Karin Leander10, Ulf de Faire10, Stefan Rosfors11, Ulf Lockowandt12,13, Jan Liska12,13, Peter Konrad14, Rabbe Takolander14, Anders Franco-Cereceda12,13, Eric E. Schadt4, Torbjörn Ivert12,13, Anders Hamsten8, Jesper Tegnér1,2,3, Johan Börjkegren1,2,3,9

1 The Computational Medicine Group, Atherosclerosis Research Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden, 2 Department of Computational Biology, Linköping Institute of Technology, Linköping University, Linköping, Sweden, 3 Clinical Gene Networks AB, Karolinska Science Park, Stockholm, Sweden, 4 Rosetta Inpharmatics, Merck, Seattle, Washington, United States of America, 5 South African National Bioinformatics Institute (SANBI), University of the Western Cape, Cape Town, South Africa, 6 Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia, 7 Department of Clinical Sciences, Danderyd Hospital, Karolinska Institutet, Stockholm, Sweden, 8 Cardiovascular Genetics Group, Atherosclerosis Research Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden, 9 Massachusetts General Hospital (MGH) Weight Center and Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America, 10 Department of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, 11 Department of Clinical Physiology, Stockholm Söder Hospital, Karolinska Institutet, Stockholm, Sweden, 12 Department of Thoracic Surgery and Anesthesiology, Karolinska University Hospital, Stockholm, Sweden, 13 Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, 14 Department of Surgery, Stockholm Söder Hospital, Karolinska Institutet, Stockholm, Sweden

Abstract

Environmental exposures filtered through the genetic make-up of each individual alter the transcriptional repertoire in organs central to metabolic homeostasis, thereby affecting arterial lipid accumulation, inflammation, and the development of coronary artery disease (CAD). The primary aim of the Stockholm Atherosclerosis Gene Expression (STAGE) study was to determine whether there are functionally associated genes (rather than individual genes) important for CAD development. To this end, two-way clustering was used on 278 transcriptional profiles of liver, skeletal muscle, and visceral fat (n = 66/tissue) and atherosclerotic and unaffected arterial wall (n = 40/tissue) isolated from CAD patients during coronary artery bypass surgery. The first step, across all mRNA signals (n = 15,042/12,621 RefSeqs/genes) in each tissue, resulted in a total of 60 tissue clusters (n = 3958 genes). In the second step (performed within tissue clusters), one atherosclerotic lesion (n = 49/48) and one visceral fat (n = 59) cluster segregated the patients into two groups that differed in the extent of coronary stenosis (P = 0.008 and P = 0.00015). The associations of these clusters with coronary atherosclerosis were validated by analyzing carotid atherosclerosis expression profiles. Remarkably, in one cluster (n = 55/54) relating to carotid stenosis (P = 0.04), 27 genes in the two clusters relating to coronary stenosis were confirmed (n = 16/17, P = 10^{-22} and P = 10^{-3}). Genes in the transendothelial migration of leukocytes (TEMl) pathway were overrepresented in all three clusters, referred to as the atherosclerosis module (A-module). In a second validation step, using three independent cohorts, the A-module was found to be genetically enriched with CAD risk by 1.8-fold (P < 0.004). The transcription co-factor LIM domain binding 2 (LDB2) was identified as a potential high-hierarchy regulator of the A-module, a notion supported by subnetwork analysis, by cellular and lesion expression of LDB2, and by the expression of 13 TEMl genes in Ldb2-deficient arterial wall. Thus, the A-module appears to be important for atherosclerosis development and, together with LDB2, merits further attention in CAD research.


Editor: Kathleen Kerr, University of Washington, United States of America

Received May 26, 2009; Accepted November 4, 2009; Published December 4, 2009

Copyright: © 2009 Hägg et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the Swedish Research Council (JB, JT, JS), the Karolinska Institute (JB, AH), the Stockholm County Council (JB, AH), the Swedish Foundation for Strategic Research (JB, JT), the Swedish Heart-Lung Foundation (JB), the King Gustaf V and Queen Victoria Foundation (JB), the Swedish Society of Medicine (JB, JT), the Hans and Loo Osterman Foundation for Geriatric Research (JB, JS), the Professor Nanna Swartz Fund (JB), the Foundation for Old Servants (JB, JS), the Magnus Bergvalls Foundation (JB), Ake Wiberg Stiftelse (JB, JT), Wennergren Foundation (JT), Vinnova Sweden-Japan (JB, JT), Vinnova research grant (SM, JT, JB), Vinnova SAMPOST grant (IM-MS, JB), the PhD Programme in Medical Bioinformatics (JB, JT), Linköping University and Stockholm Bioinformatics Center (JT) and Carl Tryggers Foundation (JT), Swedish Match (unconditional research grant to JT), AstraZeneca (unconditional research grants to JB, AH), and Clinical Gene Networks (JB, JT). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Clinical Gene Networks AB with Johan Börjkegren and Jesper Tegnér as major shareholders has filed a PCT application for a screening method using genes in the atherosclerosis module including LDB2 (PCT/SE2007/00864).

* E-mail: johan.bjorkkegren@ki.se
† These authors contributed equally to this work.
Introduction

The mapping of the human genome resulted in new technologies for studying complex diseases such as coronary artery disease (CAD) from a functional genomic perspective. By revealing comprehensive repertoires of molecular activities, these technologies combined with systems biology analyses will pave the way for a more detailed understanding of the complexity underlying common disorders—a prerequisite to advance molecular diagnostics for early identification of disease and to identify central disease pathways for therapies tailored to specific disease mechanisms [1–3].

The aim of the Stockholm Atherosclerosis Gene Expression (STAGE) study was to identify functionally associated genes important for CAD using whole-genome expression profiles from multiple organs. To this end, we used a modified version of a two-way clustering approach [4–6]. In the first step, the algorithm processed all mRNA signals within one organ to define a number of tissue clusters. The individual genes of the tissue clusters are defined by the level of associations between mRNA signals across all patients. In the second step, the patients are clustered according to the mRNA signals within each tissue cluster to identify signals related to clinical phenotypes. In this study, the clinical endpoint was the extent of coronary atherosclerotic lesions as judged from the degree of coronary stenosis, measured by quantitative coronary angiography (QCA). A secondary hypothesis was to reveal the extent to which any tissue cluster related to coronary stenosis acts in isolation in one organ or across several organs.

A multi-organ biopsy approach is primarily motivated by the nature of CAD development: atherosclerotic diseases are believed to start in adolescence and develop throughout life [7]. The pace of development depends on genetic and environmental risk factors. Of particular importance are metabolic disturbances (e.g., overweight, diabetes and dyslipidemias) that originate in organs central to energy metabolism, including liver, skeletal muscle, and fat deposits. Thus, molecular activities (mirrored by mRNA levels) distant from the actual site of CAD are likely to influence the progression and extent of coronary atherosclerosis.

The STAGE study comprises 114 carefully characterized patients, including a compendium of 278 global gene-expression profiles from five CAD-relevant tissues isolated during coronary artery bypass grafting (CABG). Using a two-way clustering approach, we analyzed this compendium to test our main hypothesis that there are groups of functionally associated genes (rather than individual genes) of importance for CAD and to determine whether those groups of genes act in isolation in each tissue or across several tissues.

Results

Exploratory Clustering of Gene-Expression Profiles in the STAGE Cohort

To test the main hypothesis of the study we explored the gene expression profiles of the STAGE cohort. Gene expression profiles could not be obtained from all tissues in all patients of the STAGE cohort (n = 114). Therefore, it was important to examine whether the two subgroups of patients in which gene expression profiles were obtained—66 patients with gene expression profile from visceral fat, liver, and skeletal muscle and 48 in whom expression profiles were also obtained from atherosclerotic and unaffected arterial wall—had similar clinical phenotypes. Indeed, this appeared to be the case (Table 1).

In the first step of the two-way clustering analysis, mRNA signals of 15,042 Reference Sequence transcripts (RefSeq) were examined in each tissue (Figure 1, Text S1, Figure S1). Importantly, the first step was performed without preconceptions about the extent of coronary atherosclerosis in the CABG patients. Instead, tissue-specific mRNA signals across the patients were analyzed solely to determine whether or not a given RefSeq belonged to a group of functionally associated genes in a tissue cluster. The first clustering step generated 60 tissue clusters representing 4007 RefSeqs/3958 genes (Table S1). Thus, 73% of the RefSeqs or 11,035 RefSeqs (3863 genes) were excluded from further analysis (i.e., the second clustering step). Of these 60 tissue clusters, 15 were identified from the liver gene expression profiles, 11 from skeletal muscle profiles, 20 from visceral fat profiles, and 14 from gene expression profiles of the atherosclerotic arterial wall (Table S1). To assess the repeatability and reliability of these clusters, resampling using Jackknife analysis was performed (Table S1).

In the second step of clustering, the mRNA signals within each of the 60 tissue clusters were used to cluster the patients. The extent of coronary stenosis, determined by QCA, was then compared in the resulting patient groups. Two of the 60 tissue clusters (n = 49 RefSeqs/48 genes, Table S2, (90% CI: 28–49) and n = 59 RefSeqs/genes, Table S3, (90% CI: 38–59), respectively) segregated the patients into groups according to the extent of coronary stenosis: one cluster in atherosclerotic arterial wall and one in visceral fat (P = 0.008 (Figure 2) and P = 0.00015 (Figure 3), respectively).

To determine whether the identified tissue clusters relating to coronary atherosclerosis are tissue-specific or present in several tissues, we assessed the gene overlap between the atherosclerosis-related clusters in atherosclerotic arterial wall and visceral fat. Seven genes (12%, 14% respectively) were present in both tissue clusters. Although this overlap may appear small, the statistical likelihood of observing an overlap of this size by chance was less than 10^{-10}. Thus, this overlap indicates atherosclerosis-related gene activity common to both visceral fat and atherosclerotic arterial wall.
Confirmatory Clustering of Gene-Expression Profiles of Carotid Lesions

The molecular underpinnings of atherosclerosis are believed to be very similar in all major arteries [7]. Accordingly, if the two atherosclerosis-related tissue clusters identified in the STAGE cohort are of general importance for atherosclerosis, they should be possible to confirm, at least in part, in another atherosclerotic tissue sample. To this end, total RNA samples from atherosclerotic carotid lesions were isolated from patients undergoing carotid stenosis surgery (Figure 1 and Table 1). Both the gene expression profiling and the subsequent two-way clustering analysis were performed exactly according to the protocol used for the STAGE cohort. A well-established surrogate measure of the extent of carotid atherosclerosis [8], the intima-media thickness (IMT), was determined preoperatively using ultrasound. The first clustering step generated a total of eight tissue clusters (Table S1).
representing 904 RefSeqs/894 genes. In the second clustering step, one of the eight tissue clusters (n = 55 RefSeqs/54 genes, Table S4, (90% CI: 32–55)) segregated the patients into two groups according to IMT score (P = 0.039, Figure 4). Remarkably, 16 of the 55 RefSeqs overlapped with genes in the visceral fat cluster (P = 10−227), and 17 with genes in the atherosclerotic arterial wall cluster (P = 10−230) (Figure 5A). Six RefSeqs (representing the genes encoding C-type lectin domain family-14, cadherin-5, chromosome 20 open reading frame-160, endothelial differentiation sphingolipid G-protein-coupled receptor-1, G protein-coupled receptor-116, and LIM domain binding 2 (LDB2)) were in all three clusters (P = 10−23); the union of the clusters contained 129 RefSeqs/128 genes (Figure 5A, Table S5).

Network and Bioinformatic Analyses of the Atherosclerosis Module

The highly significant overlaps between the three clusters in the atherosclerotic arterial wall, visceral fat and carotid stenosis suggest that the union of all genes may represent a module harboring biological activity important for human atherosclerosis (referred to as the A-module). To investigate interactions between genes in the A-module, gene expression profiles from these tissues were reused to infer a total of three gene networks (Text S1). In Figure 5B, a network supported by nodes and edges in at least two of the three networks is shown. The network of A-module genes consisted of 49 nodes (genes) interacting with a total of 55 edges, of which LDB2 had 19 edges and BCL6B had 14 edges.

To learn more about the functional representation of the A-module, bioinformatic analysis using Gene Ontology (GO) and KEGG pathway was performed (Table S6). Thirty-one of the 128 genes had previously been related to atherosclerosis (Table S9), 40 had no GO annotation, and six participated in regulatory activity (Text S1). Only 39 of the 128 genes had annotation in KEGG pathways. Twenty-three of these 39 genes (~60%) were associated with the transendothelial migration of leukocyte (TEML) pathway with a statistical significant enrichment score (P = 6.6 × 10−5, FDR = 0.01; Figure 5C).

Enrichment of Genetic Risk for CAD in the Atherosclerosis Module

If gene activity in the A-module is causally important for atherosclerosis development (and not merely reactive marker for the extent of atherosclerosis), functionally associated single nucleotide polymorphisms (SNPs) in the vicinity of the 128 A-module genes should be enriched for CAD risk. In addition, such enrichment would further strengthen our notion that the A-module genes as being important in atherogenesis. To investigate this, we first identified SNPs in the A-module that were significantly associated with gene expression (eSNPs, indicating a functional relation between the SNP allele distribution and gene expression (Text S1)) using two genetics of gene expression (GGE) studies [10]. Next, to test whether the identified eSNPs also were enriched for association with CAD, we assembled results from a recent genome-wide association study (GWAS), the Wellcome

![Figure 1. Analytical scheme of multi-organ clustering steps in the STAGE study.](https://www.plosgenetics.org/file/2009/12/01/Figure1.jpg)
Figure 2. Heat map of an atherosclerotic arterial wall cluster related to coronary stenosis. The cluster was defined by related mRNA levels (indicated by average probe signals on the arrays) and identified as one of fourteen atherosclerotic arterial wall clusters by the second step of coupled two-way clustering of mRNA profiles from STAGE patients (Text S1). Columns represent individual patients, and rows individual RefSeqs with corresponding gene symbols and mRNA ratios of the two patient groups. Above heat map: individual patient numbers, below heat map: bars indicating individual stenosis score together with means ± SD and average ratios in each group and P-values for comparing groups. EVA1 is represented by two RefSeqs.

doi:10.1371/journal.pgen.1000754.g002
Figure 3. Heat map of a visceral fat cluster related to coronary stenosis. The cluster was defined by related mRNA levels (indicated by average probe signals on the arrays) and identified as one of 20 visceral fat clusters by the second step of coupled two-way clustering of mRNA profiles from STAGE patients (Text S1). Columns represent individual patients, and rows individual RefSeqs with corresponding gene symbols and mRNA ratios of the two patient groups. Above heat map: individual patient numbers, below heat map: bars indicating individual stenosis score together with means ± SD and average ratios in each group and P-values for comparing groups. Red highlighting indicates genes also found in the cluster in Figure 2.

doi:10.1371/journal.pgen.1000754.g003
Comparing groups. Red highlighting indicates genes also identified in strong linkage disequilibrium (LD) (R and WTCCC studies used different SNP-microarray platforms, Trust Case Control Cohort (WTCCC) study [11]. Since the GGE means groups. Below heat map: bars indicating individual IMT together with corresponding gene symbols and mRNA ratios of the two patient represents individual patients, and rows individual RefSeqs with mRNA profiles from Carotid Stenosis patients (Text S1). Columns stenosis clusters by the second step of coupled two-way clustering of probe signals on the arrays) and identified as one of eight carotid The cluster was defined by related mRNA levels (indicated by average Figure 4. Heat map of a carotid stenosis cluster related to IMT. The cluster was defined by related mRNA levels (indicated by average probe signals on the arrays) and identified as one of eight carotid stenosis clusters by the second step of coupled two-way clustering of mRNA profiles from Carotid Stenosis patients (Text S1). Columns represent individual patients, and rows individual RefSeqs with corresponding gene symbols and mRNA ratios of the two patient groups. Below heat map: bars indicating individual IMT together with means ± SD and average ratios in each group and P-values for comparing groups. Red highlighting indicates genes also identified in the clusters in Figure 2 and Figure 3. EVA1 is represented by two RefSeqs. doi:10.1371/journal.pgen.1000754.g004

Identifying a Putative Regulator of the Atherosclerosis Module

Of the six genes in the intersection of all three clusters making up the A-module (Figure 5A), LDB2 was the only transcriptional regulator. The re-occurrence of this transcriptional co-factor in three separate genome-wide analyses suggested a regulatory role of the A-module genes. A notion supported by the interconnectivity of LDB2 in the network analysis (Figure 5B). To investigate this possibility further, we first identified seven transcription factors (TFs) (ISL-1alpha, Lmo2, Lhx3a, Lhx3b, LHX2, LHX4, and BRCa1) having LIM-binding domains [12] or otherwise previously been shown to interact with LDB2 [13]. We then performed in silico sequence matching for 161 promoters (Ensembl) found in 122 of the 128 A-module genes using TRANSFAC (v11.2) [14]. Of these 161 promoters (target promoters), 81% had binding site(s) for at least one of the seven TFs, suggesting that LDB2 could regulate the A-module via these TFs. In relation to a background of 10,255 human promoters covering a [-600,-1] region relative to transcription start sites, binding to the target promoters was enriched 1.2- to 5-fold (Text S1, Table S10). The enrichment for the entire family of 7 TFs was statistically significant (P = 0.011).

Functional Validation of LDB2 in Atherosclerosis

Next, we investigated the possible role of LDB2 in atherosclerosis in vivo in three major atherosclerosis cell types as well as in vitro in atherosclerosis-free arterial wall and in early and late atherosclerotic lesions in atherosclerosis-prone Ldb2−/− ApoB100/100 mice [15]. The presence of LDB2 in the arterial endothelium was first assessed by co-localization of LDB2 with the endothelial marker von Willebrand factor (VWF). LDB2 expression was most obvious in the endothelium before an atherosclerotic lesion had developed and generally co-localized with VWF (Figure 6A, 40×). In late and early lesions, LDB2 endothelial expression was patchy and subtler, and the co-localization with VWF was less clear except in the endothelium of lesion-free areas (e.g., cusps; Figure 6A). LDB2 expression in endothelial cells was confirmed by RT-PCR analyses in a human endothelial cell line (EAHY926) and in human umbilical vein endothelial cells (HUVECs) (Figure 6B). In accordance with the immunohistochemical results, the mRNA levels were higher in noninduced than in induced EAHY926 cells (Figure 6B).

To investigate LDB2 protein expression in other atherosclerosis cell types, CD68 was used as a marker of lesion macrophage/foam cells and SM22 (transgelin) as a marker of lesion smooth muscle cells (SMCs). In early lesions, LDB2 staining was subtle (but clearly present compared to control) and appeared to co-localize with both CD68 and SM22 (Figure 6C). In late lesions, LDB2 staining was marked, and in all locations of LDB2 staining there was also CD68 staining. In this sense, there was co-localization of LDB2 and CD68. However, the CD68 staining was generally stronger, and some areas with CD68 staining had little or no LDB2 staining. LDB2 also co-localized with SM22, but some areas with marked LDB2 staining had no SM22 staining (Figure 6B, ovals). LDB2 was also expressed in macrophages/foam cells in human carotid lesions (Figure S2).
The immunohistochemical results were largely confirmed by RT-PCR analyses of primary SMCs and macrophages and a human monocytic cell line (THP-1) (Figure 6D). Consistent with the higher protein expression in late lesions than in early lesions, LDB2 mRNA levels increased with differentiation of THP-1 monocytes to macrophages and foam cells (panel 1). The expression of LDB2 in THP-1 was also confirmed in primary macrophages (panel 2). In primary SMCs isolated from human pulmonary artery, there was also clear expression of LDB2, which in comparison with the immunohistochemical results was surprisingly high (panel 3).

In summary, LDB2 was expressed by all three major atherosclerosis cell types; before lesion formation and in early lesions primarily in the endothelium and in late lesions, mainly in macrophages/foam cells but also in SMCs. The generally higher LDB2 expression in late lesions was confirmed by RT-PCR of total RNA from early and late lesions isolated from mouse aortic arch samples (Figure 6E).

Last, we examined mRNA levels of 20 genes central to TEML in the arterial wall of 6-week-old Ldb2−/− mice. Our goal was to investigate a possible role of LDB2 as a regulator of TEML genes in general and specifically as a regulator of A-module genes. All 20 genes had higher levels of expression in Ldb2−/− than in wild-type mice whereof 13 was significantly higher (Table 2). Eight of these 13 genes were specific to the A-module, and five were not. Of note, five of the investigated genes have previously been targeted in mouse models of atherosclerosis and found to be affecting lesion development [16–20].
Taken together, the functional validation supports a role for LDB2 in TEML and atherosclerosis development. Particularly, since endothelial LDB2 seems to regulate TEML already before microscopic evidence of lesion formation.

**Discussion**

In the STAGE study, we profiled five CAD-relevant tissues to identify functionally associated genes with potential importance in
coronary atherosclerosis. This analysis revealed 128 genes that were strongly associated with atherosclerosis severity [A-module]. The A-module was found to be enriched with genetic risk for CAD and involve the TEML pathway. Parts of the A-module were active in both atherosclerotic arterial wall and visceral fat. The latter may be a local source of inflammation contributing to coronary atherosclerosis. We also identified a putative high-hierarchy regulator of the A-module, LDB2, which was robustly expressed in all major lesion cell types both in lesion-free and in late atherosclerosis lesions. Interestingly, key genes in the TEML pathway were differentially regulated in the arterial wall of Ldb2-deficient mice. Our findings suggest that the A-module, including LDB2, is important in the regulation of TEML in and atherosclerosis development.

TEML is an established pathway in atherosclerosis and other inflammatory diseases [21]. Transendothelial migration of monocytes is essential for foam-cell formation and for early phases of atherogenesis, and transendothelial migration of T-cells may be central in later phases [22]. Indeed, leukocyte migration has been suggested as a therapeutic target [23]. The identified module was enriched in genes involved in TEML and thus may be causally involved in the development of clinically significant atherosclerotic lesions (as indicated by the extent of coronary stenosis and IMT). However, most of the identified A-module genes lack pathway annotations but may in future studies be proven important to leukocyte migration or its regulation.

The STAGE study was designed as a “top-down” systems biological approach to identify gene networks or groups of otherwise functionally associated genes [modules] of importance for disease severity [3]. The term “top-down” refers to our belief that these modules must first be identified in clinical studies as the most disease relevant and then be consecutively detailed by studies in animal and cellular models to reveal high-resolution networks [24]. In contrast, “bottom-up” systems biology approaches first identify full biological networks in prokaryotic or yeast cells and then examine their roles in more disease-relevant systems. Systems biological approaches have advantages over traditional gene-expression profiling studies, which usually focus on identifying individual genes differentially expressed as a result of disease. Such gene-by-gene analyses generate many false positives due to a vast “multiple testing” problem. In contrast, the two-way clustering approach first focuses on identifying functionally associated genes (which in the current study reduced the number of genes from 12,621 to 3958 represented in 60 tissue clusters) and then investigate whether the generated clusters (not individual genes) are related to a given disease phenotype.

Using a multi-organ approach [3], we hypothesized the liver, skeletal muscle, or fat deposits would harbour functionally related genes (e.g., clusters, modules, networks) reflecting molecular processes in those organs affecting the levels of inflammatory mediators, blood lipids, glucose or unknown blood constituents that contribute to coronary atherosclerosis development. There were no clusters relating to the extent of coronary atherosclerosis

### Table 2. mRNA levels measured by real-time PCR from the aortic arch of 6-week-old mice deficient in Ldb2 (Ldb2<sup>−/−</sup>) and littermate wild-type controls (Ldb2<sup>wt/wt</sup>.

<table>
<thead>
<tr>
<th>Category</th>
<th>Gene Symbol</th>
<th>Ldb2&lt;sup&gt;wt/wt&lt;/sup&gt;</th>
<th>Ldb2&lt;sup&gt;−/−&lt;/sup&gt;</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-module genes associated to TEML</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claudin 5</td>
<td>Cldn5</td>
<td>307 ± 108</td>
<td>397 ± 271</td>
<td>0.47</td>
</tr>
<tr>
<td>Phospholipase C gamma 2</td>
<td>Pldc2</td>
<td>461 ± 65</td>
<td>726 ± 219</td>
<td>0.019</td>
</tr>
<tr>
<td>Cadherin 5</td>
<td>Cdh5</td>
<td>352 ± 114</td>
<td>603 ± 179</td>
<td>0.011</td>
</tr>
<tr>
<td>Chemokine (C-X-C motif) ligand 12</td>
<td>Cxcl12</td>
<td>498 ± 103</td>
<td>715 ± 168</td>
<td>0.015</td>
</tr>
<tr>
<td>Platelet/endothelial cell adhesion molecule</td>
<td>Pecam1</td>
<td>345 ± 122</td>
<td>564 ± 157</td>
<td>0.016</td>
</tr>
<tr>
<td>Angiotensin II receptor-like 1</td>
<td>Aplnr</td>
<td>435 ± 253</td>
<td>846 ± 404</td>
<td>0.069</td>
</tr>
<tr>
<td>Kinase insert domain receptor</td>
<td>Kdr</td>
<td>386 ± 224</td>
<td>964 ± 555</td>
<td>0.043</td>
</tr>
<tr>
<td>Protocadherin 12</td>
<td>Pcdh12</td>
<td>491 ± 188</td>
<td>785 ± 339</td>
<td>0.10</td>
</tr>
<tr>
<td>Protein Kinase N3</td>
<td>Pkn3</td>
<td>410 ± 193</td>
<td>1076 ± 697</td>
<td>0.050</td>
</tr>
<tr>
<td>Protein kinase C eta</td>
<td>Prkch</td>
<td>547 ± 199</td>
<td>1045 ± 369</td>
<td>0.019</td>
</tr>
<tr>
<td>Protein tyrosine phosphatase receptor type B</td>
<td>Ptprb</td>
<td>486 ± 167</td>
<td>1115 ± 575</td>
<td>0.030</td>
</tr>
<tr>
<td>Tek tyrosine kinase (endothelial)</td>
<td>Tek</td>
<td>430 ± 122</td>
<td>1068 ± 551</td>
<td>0.021</td>
</tr>
<tr>
<td>Tyrosine kinase with immunoglobulin-like and EGF-like domains 1</td>
<td>Tie1</td>
<td>524 ± 170</td>
<td>895 ± 374</td>
<td>0.056</td>
</tr>
<tr>
<td>Other TEML genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercellular adhesion molecule 1</td>
<td>Icam1</td>
<td>405 ± 54</td>
<td>533 ± 73</td>
<td>0.0042</td>
</tr>
<tr>
<td>F11 receptor</td>
<td>F11r</td>
<td>388 ± 59</td>
<td>614 ± 151</td>
<td>0.0037</td>
</tr>
<tr>
<td>Junction adhesion molecule 2</td>
<td>Jam2</td>
<td>452 ± 70</td>
<td>616 ± 137</td>
<td>0.018</td>
</tr>
<tr>
<td>Junction adhesion molecule 3</td>
<td>Jam3</td>
<td>567 ± 53</td>
<td>741 ± 163</td>
<td>0.022</td>
</tr>
<tr>
<td>Vascular cell adhesion molecule 1</td>
<td>Vcam1</td>
<td>492 ± 83</td>
<td>730 ± 134</td>
<td>0.0025</td>
</tr>
<tr>
<td>Thymus cell antigen 1</td>
<td>Thy1</td>
<td>556 ± 158</td>
<td>707 ± 264</td>
<td>0.23</td>
</tr>
<tr>
<td>CDC42 effector protein (Rho GTPase binding) 5</td>
<td>Cdc42ep5</td>
<td>540 ± 127</td>
<td>622 ± 119</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Values are mean ± SD. p-Values are calculated with unpaired t-test.
Values are normalized to acidic ribosomal phosphoprotein P0 and TATA box binding protein.
Ldb2<sup>−/−</sup>, n = 5–6; Ldb2<sup>wt/wt</sup>, n = 6–7.
doi:10.1371/journal.pgen.1000754.t002
in the liver and skeletal muscle. This was surprising given the importance of these organs for CAD risk factors, such as plasma cholesterol and diabetes. However, therapies to reduce plasma lipid and glucose levels (Table 1) might have normalized disease-promoting activities in CAD-modules in these organs. In contrast, we identified one part of the A-module in visceral fat that segregated patients according to the degree of coronary stenosis. Although the relation of visceral fat to CAD risk factors in blood is less clear, a high waist-hip ratio—an indicator of increased visceral fat mass in the abdomen—is a strong predictor of CAD [25]. An interesting aspect of the visceral fat in the mediastinum is its anatomic location and the possibility that it is a source of local macrophages releasing inflammatory mediators [26]. Another possible cellular source for the presence of the TEML-enriched atherosclerosis module in visceral fat may be endothelial cells, which are relatively enriched in this tissue. Although our study does not directly address the subcellular origin of the A-module in visceral fat or how it contributes to atherosclerosis, it might be a local source of inflammatory mediators that increase the rate of atherosclerosis progression [27].

In all, 60 tissue clusters were identified, two of which—one in athero sclerotic lesion and one in visceral fat—related to the extent of coronary atherosclerosis. This might appear to be a small fraction (2/60, ~3%). However, since the first clustering step takes no phenotypic data into consideration but is entirely based on the mRNA signals in each tissue, these 60 clusters may relate to tissue physiology or subtraits of CABG patients (Table 1). Examining the latter possibility, we found that as many as 41 of the tissue clusters (besides the two related to extent of coronary atherosclerosis) segregated the patients into groups with significant difference in the levels of subtraits (not shown).

The gene expression clustering was done with the absolute value of Spearman rank correlation as distance measure. Thus, we also included inverse correlated genes which could be implicated in the same pathway and functionally related. Moreover, Spearman rank correlation is a non-parametric measure stable against outliers and in this sense a better distance measure than commonly used Euclidean and Manhattan distances, where the magnitude in expression levels are important. Of note, a clustering algorithm could produce different clusters depending on the distance measure used and the A-module could therefore have been different or even lost by other metric clustering choices.

We used atherosclerotic aortic wall/internal mammary artery (IMA) ratios to highlight atherosclerosis gene expression in the aortic wall because both aortic wall and IMA samples contain normal wall gene expression. Unlike the aortic wall, however, the IMA has no atherosclerosis [28]. This notion was supported by macro- and microscopic examinations of randomly chosen sets of aortic wall and IMA samples. Moreover, two-way clustering of mRNA signals from the aortic wall samples alone did not generate any cluster that segregated patients by stenosis scores (not shown), which may be due to a relative large portion of normal vascular wall gene expression in this tissue. However, we cannot entirely exclude the possibility that using the aortic wall/IMA ratios resulted in some false-positive genes (nonatherosclerosis genes related to normal vascular wall gene expression) that should have been excluded from the A-module or false-negative genes that otherwise should have been included.

We decided to use two different atherosclerosis cohorts—coronary for the exploration and carotid for the confirmatory step. In doing so, we added more credibility to the confirmatory step that would have been lost if we instead had used identical cohort for exploration and confirmation. The validation in the carotid cohort indicates a general importance of the A-module in atherosclerosis and at the same time rules out the possible risk that any of the tissue clusters identified in the STAGE cohort was a result of the exploratory study design (e.g. choice of sample locations and/or using ratios instead of straight expression) rather than related to atherosclerosis. The extents of coronary and carotid atherosclerosis (as judge from the surrogate measurements of stenosis score and IMT [8,29]) have repeatedly been shown to be highly correlated [30]. This observation is not entirely surprising since atherosclerosis development and the principal molecular processes underlying this development have been found to be very similar in all major arteries, regardless of location [7].

Currently, GWAS are given much attention in leading scientific journals. However, such studies have some limitations, since they are primarily designed to identify the relatively few DNA variants that influence the risk of developing complex diseases, like CAD, independently of other risk factors [31]. In the current study, we used a recently published GWAS [11] to further validate the A-module genes by calculating the relative enrichment of genetic CAD risk in the module. Unlike today’s GWAS, which link DNA variation directly to clinical phenotypes, future studies that also include intermediate expression phenotypes have the potential to extract much more disease-relevant information on DNA variation that contributes to the development of complex diseases. For now, this information remains hidden in the data generated by GWAS.

Genes encoding LIM domain-binding factors such as LDB2 were initially isolated in a screen for proteins that physically interact with the LIM domains of nuclear proteins. These proteins bind to a variety of TFs and are likely to function as enhancers, bringing together diverse TFs to form higher-order activation complexes [32–33]. Our screen of LDB2-associated TFs identified ISL-1alpha, Lmo2, Lhx3a, Lhx3b, LHX2, LHX4, and BRCA1. ISL-1alpha enhances HNF4 activity and thus insulin signaling [34–35]. Lmo2 is involved in angiogenesis [36–37]. Lhx3 and Lhx4 regulate proliferation and differentiation of pituitary-specific cell lineages [38] and are expressed in subsets of lymphocytes [39] and thymocyte tumor cell lines [40]. BRCA1 is associated with a selective deficiency in spontaneous and LPS-induced production of tumor necrosis factor (TNF-α) and of TNF-alpha-induced expression of intercellular adhesion molecule-1 (ICAM1) on peripheral blood monocytes [41] and in controlling the life cycle of T-lymphocytes [42]. LDB2 has not previously been related to CAD or atherosclerosis. Because of its high-hierarchy regulatory role and involvement in diverse biological processes, LDB2 is an interesting target for further evaluation in complex diseases.

Being the only transcriptional regulator among the six genes relating to severity of atherosclerosis present in all three tissue clusters (Figure 6A), LDB2 was chosen for functional validation in atherosclerosis. However, despite the fact that none of the other five genes were transcriptional regulators, they might still be of functional importance for atherosclerosis development, which remains to be determined. In nonatherosclerotic arterial wall and in early lesions, LDB2 was mainly expressed by the endothelium. In late lesions, LDB2 expression was more intense and mainly seen in macrophages/foam cells but also in SMCs. The TEML pathway has been implicated in both early and late atherosclerosis [23]. This pathway is also active in lesion SMCs accompanying endothelial cells in recruiting monocytes from the blood to the atherosclerotic plaque [43–44]. The pattern of LDB2 expression seen in early and late lesions has been observed for other key TEML genes (Vcam1, Icam1, Ccl2, -14, and -16, and Cdc20) [45]. The notion that LDB2 is an important regulator of TEML is further supported by the fact that 13 key genes in TEML were differentially expressed in the arterial wall of *Ldb2*−/− mice already at 6 weeks of age. Five of those genes have previously been
shown to affect atherosclerosis in mouse model studies [16–20]. In addition, a very recent study demonstrated that LDB2 regulates cell migration both in vitro and in vivo [46]. However, the final verdict on LDB2 as an important regulator of atherosclerosis development remains to be determined.

Although it cannot be excluded that the A-module also will be of importance for early stage of atherosclerosis (e.g., by promoting early lesion development through activating TELM in the atherosclerosis-free endothelium), the current study mainly supports a role of the A-module in late stages of coronary atherosclerosis. If the activity of this cassette of genes is mirrored, at least in part, by gene expression in blood (i.e., in leukocytes) or by plasma protein levels, the A-module may be helpful as a complement to semi-invasive investigations (e.g., angiography) as markers of degree of coronary and carotid stenosis.

In conclusion, by adopting a new strategy for functional analysis of expression profiles isolated from multiple CAD-relevant organs, we identified a module that is genetically enriched with CAD risk and important for TELM and atherosclerosis development. The clinical usefulness, and exact role in CAD of this module and its high-hierarchy regulator [32–33] LDB2, merit further investigation.

Methods

Study Patients, Biopsy Collection, and Follow-Up

The STAGE study enrolled 124 patients undergoing CABG at Karolinska University Hospital, Solna. Forty-two patients undergoing carotid surgery at Stockholm Söder Hospital were recruited as a confirmatory cohort. The studies were approved by the Ethics Committee of Karolinska University Hospital. All patients gave written informed consent.

Tissue samples from the distal IMA, wall of the ascending aorta (aortic root) at the site of proximal vein anastomosis, anterior hepatic edge (liver), skeletal muscle, and visceral fat in the mediastinum were preserved in RNA later (Qiagen) and frozen at −80°C. Lesions in aortic wall samples [47–48] and the absence of lesions in the IMA [28] were confirmed by macroscopic and microscopic examinations (not shown). Carotid plaques were embedded in OCT (Histolab Products), frozen in liquid isopentane and dry ice, and stored at −80°C.

One hundred fourteen CABG and 39 carotid stenosis patients came to a 3-month follow-up visit. Using a standard questionnaire, a research nurse obtained a medical history and lifestyle information (e.g., smoking, alcohol consumption, and physical activity). A physical examination was performed including venous blood sampling (Text S1).

Coronary and Carotid Atherosclerosis Measurements

All CABG patients underwent preoperative biplane coronary angiography (Judkins technique). Angiograms were evaluated with QCA techniques (Medis). The left and right coronary arteries and their branches were divided into segments [49]. Each segment was measured during end-diastole. A stenosis score was calculated from all major lesions in the coronary arteries (1 point, 20–50% luminal obstruction; 2 points >50% obstruction). In some patients, right coronary artery occlusion prohibited QCA evaluation. Before surgery, carotid arteries were examined with B-mode ultrasound. The far wall of the common carotid artery was used to measure IMT from the endarterectomy side [50].

RNA Isolation and Expression Profiling

We performed gene expression profiling on three tissues (liver, skeletal muscle, visceral fat) in 66 of 114 STAGE patients, and also in 40 of these 66 patients, on atherosclerotic arterial wall and IMA. In the validation cohort, 23 carotid lesions from 39 patients were randomly selected for RNA isolation and gene expression profiling. Aortic arches (third rib to aortic root) were isolated in RNA later (Ambion) from 6-week-old mice deficient in Ldb2 (Ldb2−/−; Mutant Mouse Regional Resource Center, University of California, Davis), heterozygous and wildtype littersmates, and 20- and 40-week-old atherosclerosis-prone mice deficient in the low density lipoprotein receptor and expressing exclusively apolipoprotein B100 (Ldb2−/−ApoB100/100 mice). Total RNA was isolated from all biopsies with Trizol (BRL-Life Technologies) and FastPrep (MP Biomedicals) and purified with RNaseasy Mini kit using DNase1 treatment (Qiagen). Sample quality was assessed with an Agilent Bioanalyzer 2100. cRNA yield was assessed with a spectrophotometer (ND-1000, NanoDrop Technologies) before hybridization to HG-U133 Plus 2.0 arrays (Affymetrix). The arrays were processed with a Fluidics Station 450, scanned with a GeneArray Scanner 3000, and analyzed with GeneChip Operational Software 2.0.

Immunohistochemistry

Mouse aortic roots (aortic valve level) and human carotid lesions were isolated and frozen in liquid nitrogen, embedded in OCT compound (Histolab Products), cryosectioned (5 μm), and fixed in acetone. Endogenous peroxidase activity was quenched with 0.3% hydrogen peroxide/0.01% NaN3 in water for 10 minutes, and sections were incubated with 5% blocking serum. Consecutive sections were incubated with goat anti-LDB2 (Santa Cruz Biotechnology) [51], rat monoclonal anti-mouse CD68 (Serotec), mouse monoclonal anti-human CD68 (Novocastra Laboratories), rabbit polyclonal anti-mouse SM22 alpha (transgelin, Abcam), or rabbit polyclonal anti-human VWF (DakoCytomation). Avidin-biotin peroxidase complexes (Vector Laboratories) were added followed by visualization with DAB (Vector Laboratories). All sections were counterstained with Gill hematoxylin (Histolab Products).

Cell Cultures

THP-1 monocytes were plated in 10% fetal calf serum/RPMI-1640 with L-glutamine (2 mM) and HEPES buffer (25 mM) (Gibco-Invitrogen) supplemented with penicillin (100 U/ml) and streptomycin (100 μg/ml) and differentiated into macrophages with phorbol 12-myristate 13-acetate (50 ng/ml) (Sigma) for 72 hours. To generate foam cells, macrophages were incubated with acetylated low density lipoproteins (50 μg/ml) for 48 hours. Human monocytes were isolated from blood with Ficoll/Hypaque as described [52], placed in six-well dishes, and allowed to adhere overnight in RPMI-1640 supplemented with penicillin (100 U/ml), streptomycin (100 μg/ml), and 10% pooled human AB serum. After washing, fresh serum-containing medium was added, and cells were cultured for 6 days and harvested. EAHY926 cells were cultured in DMEM containing high glucose, penicillin (100 U/ml), streptomycin (100 μg/ml), 10% fetal calf serum, hypoxanthine (100 μmol/l), aminopterin (0.4 mmol/l), and thymidine (16 mmol/l). HUVECs were obtained by collagenase treatment, cultivated, and identified as described [53]. SMCs from human pulmonary artery (Clonetics) were cultured in SmGm2 medium containing growth factors (Clonetics) as described [54].
Real-Time PCR

Total RNA (0.15 µg) was reverse transcribed with Superscript III (Invitrogen). After threefold dilution, cDNA (3 µl) was amplified by real-time PCR with 1X TaqMan universal PCR master mix (Applied Biosystems) on an ABI Prism 7000 (PE Biosystems) using Assay-On-Demand kits containing corresponding primers and probes (Applied Biosystems). mRNA levels were normalized to acidic ribosomal phosphoprotein P0 and TATA-box binding protein. Samples were analyzed in duplicate.

Pre-Processing of Gene Expression Data

Gene-expression values were pre-processed with the robust multichip average [55] procedure in three steps (background adjustment, quantile normalization, summarization). Of 604,258 perfect-match Affymetrix probe signals, 423,636 were mapped to transcripts using RefSeq numbers as identifiers [56], generating 13,042 RefSeq transcripts corresponding to 12,621 genes. Straight expression values (i.e., mRNA signals obtained from one microarray) were used for data analyses of all tissue biopsies (including the carotid lesion biopsy in the confirmatory cohort) except for the atherosclerotic arterial wall and IMA. The latter two biopises were combined in atherosclerotic arterial wall/IMA mRNA ratios before data analysis. mRNA signals in the atherosclerotic arterial wall biopsy reflect gene activity in the atherosclerotic lesion and in normal arterial wall, whereas mRNA signals in the IMA mainly reflect normal arterial wall gene activity (the IMA is almost entirely devoid of atherosclerotic lesions) [28]. Thus, the use of atherosclerotic arterial wall/IMA ratios highlights gene activity related to atherosclerotic lesions in arterial wall and excludes that relating to normal arterial wall.

Two-Way Clustering

Coupled two-way clustering [4–6] was performed to identify small and stable clusters of related signals of importance for CAD. In the first step, clusters were defined using superparamagnetic clustering [4], with the absolute value of Spearman rank correlation as a distance measure between genes. Spearman rank is a non-parametric measure which is robust to outliers and by using absolute values we also put together anti-correlated genes. The analysis was done without using any predefined conceptions (i.e., phenotypes of the patients). Genes that did not belong to a cluster were excluded. Then, in the second step, the identified clusters were related to coronary atherosclerosis by hierarchical clustering [57] of the patients, using Manhattan distance and average linkage as distance measures, based on the mRNA signals in each of the clusters defined in the first step (Text S1).

To assess the repeatability and reliability of these clusters, resampling using Jackknife analysis was performed [58] (Text S1).

Genetic Enrichment Analysis

A-module genes were mapped to eSNPs (Text S1) using two GGE studies [10] and tested for enrichment of association with CAD using the results from the WTCCC study [11]. Different SNP panels were used in the GGE and WTCCC studies, therefore we included eSNPs and all SNPs in strong LD (R²>0.84) with the eSNPs. In the 128 A-module genes, there were 97 eSNPs and 387 LD SNPs of the eSNPs, resulting in an expanded set of 484 eSNPs. Random sampling strategy was used to assess whether the expanded eSNP set was more likely to associate with CAD than randomly selected sets of SNPs of equal number. In each random sample, 97 SNPs located within 1 megabase of human gene regions were selected to ensure the location of the random SNP sets matched that of the eSNP set in the A-module. The randomly selected SNP sets were then expanded by including SNPs in strong LD (R²>0.84) with any of the randomly identified SNPs. We required the final size of the expanded random set of SNPs to be within ±10% of the expanded set of eSNPs in the A-module.

The process was repeated 100,000 times. For each random SNP set, we counted the percentage of SNPs with association P-value to CAD<0.05, and constructed the null distribution. The enrichment P-value was calculated as the number of times that the percentage exceeds 10.3% from random sampling divided by 100,000.

Statistical Analysis

Clinical and metabolic characteristics are given as continuous variables with means ± SD and as categorical variables with percentages and numbers of subjects. P-values were calculated with unpaired t tests; skewed values were log-transformed. Statistical significances in Venn diagrams were computed using hypergeometric distributions (Text S1). GO and pathway analyses were performed with DAVID (Database for Annotation, Visualization and Integration Discovery) software [9]. Mathematica 5.2 or StatView 5.0.1 was used for all other calculations. Text mining was used to define transcripts previously related to CAD and atherosclerosis (Text S1, Table S9). For promoter analysis, TRANSFAC (v11.2) [14] was used (Text S1).

Supporting Information

**Figure S1** Principles of the cost function in the SPC algorithm. The superparamagnetic clustering (SPC) algorithm uses a cost function with a temperature parameter (T) to assign genes into different clusters. Genes could belong to many clusters (right) or to no cluster at all (left). At a certain temperature the clusters are different. Genes could belong to many clusters (right) or to no cluster at all (left). At a certain temperature the clusters are robust and stable against noise (middle). Found at: doi:10.1371/journal.pgen.1000754.s001 (1.21 MB EPS)

**Figure S2** LDB2 proteins and CD68 staining in serial sections of human carotid plaques. Consecutive human carotid plaque sections were incubated with goat anti-LDB2 antibody and rat monoclonal anti-mouse CD68 at 4 °C overnight. LDB2 is co-localized with CD68. Found at: doi:10.1371/journal.pgen.1000754.s002 (3.17 MB EPS)

**Table S1** Gene expression cluster relation to surrogate measurements of atherosclerosis (QCA and IMT). Found at: doi:10.1371/journal.pgen.1000754.s003 (0.04 MB XLS)

**Table S2** 49 RefSeqs corresponding to 48 genes of the atherosclerotic arterial wall/IMA cluster in Figure 2. Found at: doi:10.1371/journal.pgen.1000754.s004 (0.02 MB XLS)

**Table S3** 59 RefSeqs genes of the visceral fat cluster in Figure 3. Found at: doi:10.1371/journal.pgen.1000754.s005 (0.03 MB XLS)

**Table S4** 55 RefSeqs corresponding to 54 genes of the carotid lesion cluster in Figure 4. Found at: doi:10.1371/journal.pgen.1000754.s006 (0.02 MB XLS)

**Table S5** 129 RefSeqs corresponding to 128 genes in the A-module.
References


