THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Human Validation of Genes Associated With a Murine Atherosclerotic Phenotype

OPEN ACCESS

Human validation of genes associated with a murine atherosclerotic phenotype

Running title: Human validation of putative atherosclerosis genes

Authors: Gerard Pasterkamp[a,b,1,2], Sander W. van der Laan[a,1], Saskia Haitjema[a,1], Hassan Foroughi Asl[c], Marten A. Siemelink[a], Tim Bezemer[a], Jessica van Setten[a], Martin Dichgans[d,e], Rainer Malik[d], Bradford B. Worrall[f], Heribert Schunkert[g,h,], Nilesh J. Samani[i,j], Dominique P.V. de Kleijn[k,l,m], Hugh S. Markus[n], Imo E. Hoefer[a], Tom Michoel[o], Saskia C.A. de Jager[a], Johan L.M. Björkegren[c,p,q], Hester M. den Ruijter[a,3], Folkert W. Asselbergs[r,s,t,3]

Affiliations:
[a]Laboratory of Experimental Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands.
[b]Laboratory of Clinical Chemistry and Hematology, Division Laboratories and Pharmacy, University Medical Center Utrecht, Utrecht, the Netherlands.
[c]Vascular Biology Unit, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden.
[d]Institut für Schlaganfall- und Demenzforschung (ISD) Klinikum der Universität München, Munich, Germany.
[e]Munich Cluster for Systems Neurology (SyNergy), D-81377, Munich, Germany.
[f]Department of Neurology, University of Virginia, Charlottesville, United States of America.
[g]Klinik für Herz- und Kreislauferkrankungen, Deutsches Herzzentrum München, Technische Universität München, Munich, Germany.
[h]Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK) e.V., Partner Site Munich Heart Alliance, Munich, Germany.
[i]Department of Cardiovascular Sciences, University of Leicester, Leicester, United Kingdom
[j]Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Leicester, United Kingdom.
[k]Department of Surgery, National University of Singapore, National University Singapore, Singapore
[l]Cardiovascular Research Institute (CVRI), National University Health System, Singapore.
[m]Interuniversity Cardiology Institute of the Netherlands, Utrecht, the Netherlands.
[n]Department of Neurosciences, University of Cambridge, Cambridge, United Kingdom.
[o]The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom.
[p]Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, United States of America.
[q]Department of Physiology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia.
[r]Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, the Netherlands.

1

^sDurrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, the Netherlands.
^tInstitute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom.

[1] shared first author
[3] shared last author

[2] Corresponding author
Gerard Pasterkamp
Laboratory of Experimental Cardiology, Clinical Chemistry and Hematology
University Medical Center Utrecht
Room G02-523
Heidelberglaan 100
3584CX Utrecht
The Netherlands
E: g.pasterkamp@umcutrecht.nl
T: +31 88 755 86 54

Word count: 5978
Subject codes: [145]; [146]; [89]; [130]; [55]
Keywords: atherosclerosis, GWAS, mouse
Total number of figures and tables: 2
TOC category: Translational
TOC subcategory: Arteriosclerosis

2

Abstract

<u>Objective:</u> The genetically modified mouse is the most commonly used animal model for studying the pathogenesis of atherosclerotic disease. We aimed to assess if mice atherosclerosis related genes could be validated in human disease through examination of results from genome wide association studies.

<u>Approach and Results:</u> We performed a systematic review to identify atherosclerosis-causing genes in mice and carried out gene-based association tests of their human orthologues for an association with human coronary artery disease (CAD) and human large artery ischemic stroke (LAS). Moreover, we investigated the association of these genes with human atherosclerotic plaque characteristics. Additionally, we assessed the presence of tissue-specific *cis*-acting expression quantitative trait loci (eQTLs) for these genes in humans. Lastly, using pathway analyses we show that the putative atherosclerosis-causing genes revealed few associations with human CAD, LAS or atherosclerotic plaque characteristics, despite the fact that the majority of these genes have *cis*-acting eQTLs.

<u>Conclusions:</u> A role for genes that has been observed in mice for atherosclerotic lesion development could scarcely be confirmed by studying associations of disease development with common human genetic variants. The value of murine atherosclerotic models for selection of therapeutic targets in human disease remains unclear.

Introduction
Atherosclerosis is a multifactorial process that develops over decades, underlying the majority of cardiovascular diseases. Due to its slow progression, studying the natural history of atherosclerosis requires serial examinations, thus complicating the design of studies in humans. Most research on biological mechanisms of atherosclerosis has been performed in genetically modified mice, eliminating these challenges faced in human studies. Thus genetically modified mice elegantly allow the study of atherosclerosis in, arguably, the best-controlled model system possible. The most commonly used atherosclerotic murine models are Apolipoprotein E (*ApoE*) or Low-Density Lipoprotein Receptor (*LDLR*) gene knockouts. These mice clearly display an accelerated atherosclerotic phenotype with human-like vascular lesions[1,2]. Experimental modifications of these murine models may accelerate vascular plaque development resulting in advanced lesions within several weeks[3,4]. Consequently such models have been crucial in understanding the murine molecular and cellular basis of atherosclerosis. Yet, the relevance in human atherosclerotic disease remains elusive for a number of reasons. Primarily, the morphology of atherosclerotic plaque in mice differs from that of humans and acute events due to luminal thrombosis and evident plaque rupture are rarely observed. Secondly, clinical disease manifestations such as coronary artery disease (CAD) or ischemic stroke in mice are rare or lacking. Thirdly, it is arguable whether complete knockouts in mice correspond with expression-changing mutations in humans. Lastly, despite the knockout of individual genes, genetic redundancy on the pathway level further complicates the interpretation of results from animal models[5].

In recent years, millions of common single-nucleotide polymorphisms (SNPs) in the human genome were identified, and our understanding of these variants with respect to the genomic architecture has increased significantly[6]. This has opened up the possibility to agnostically assess the effects of genome-wide variation on human traits and disease[7]. Indeed, meta-analyses of human genome-wide association studies (GWAS) have identified many risk loci for CAD[8] and large artery ischemic stroke (LAS)[9]. These GWAS provide the unique opportunity to validate the putative disease-causing genes identified through murine models in humans.

We performed a systematic review to identify atherosclerosis-causing genes in mice and carried out gene-based association tests of their human orthologues for CAD and LAS. Moreover, we investigated the association of these genes with human atherosclerotic plaque characteristics. Furthermore, we assessed whether there are tissue-specific *cis*-acting expression quantitative trait loci (eQTLs) for the genes in humans. We report that putative atherosclerosis-causing genes reveal little association with human CAD, LAS or atherosclerotic plaque characteristics, despite the fact that the majority of these genes have *cis*-acting eQTLs.

Materials and Methods

Materials and Methods are available in the online-only Data Supplement.

Results

Of the 659 murine genes (table SI) a total of 486 genes (73.75%) were studied in knockout mice, 57 genes (8.65%) were studied in transgenic mice, and 116 genes (17.60%) were targeted by specific compounds (table SI). For 185 genes (28.07%) there are (pre-clinical) drugs available (table SI).

Validation within GWAS of CAD and LAS

We obtained summary statistics from GWAS on CAD[8] and large artery stroke[9] for SNPs ±50kb from the 5' and 3' gene borders of the 659 studied murine genes (Supplemental Material). We used these as input for a gene-based analysis using VEGAS which assigned an empirical p-value (after permutations) to each of the 659 genes based on the p-value of the SNPs in and ±50 kb around the genes of interest[10]. Thus, each gene was given a p-value of association to CAD or LAS based on the GWAS results while taking into account the correlation between SNPs that may exist. Out of the 659 studied genes that have been shown to affect atherosclerotic phenotypes in mice, 11 (1.7%) genes were associated with CAD after correction for multiple testing (p-value ≤ 0.05/659 ≤ 7.59×10^{-5}, table SIII). In contrast, none of the genes were associated with LAS after correction for multiple testing. The top 10 most significant genes for CAD and LAS are shown in table 1. A total of 84 (12.7%) and 41 (6.2%) of the genes were associated with CAD and LAS respectively, at a nominal p-value ≤ 0.05. The overlap of associated genes between LAS and CAD is limited; only the locus at 9p21 (containing *CDKN2A/B*), significantly associated with CAD, was also nominally associated with LAS (p-value < 0.0062). We did not observe any further overlap between CAD and LAS top-associated genes. When looking at model groups to which the genes were assigned (knockout, transgenic or compound), we did not observe significant differences between groups for the 11 significant genes (p-value = 0.513 using a Chi^2 test).

Validation using human atherosclerotic plaque

Subsequently, we conducted a similar gene-based analyses using VEGAS on seven plaque characteristics in the Athero-Express Biobank Study[11]. These human plaque characteristics have previously been associated with clinical presentation[12] and have been examined in many atherosclerotic murine models. Overall, out of the 657 genes a low number of genes nominally associated with a human plaque characteristic and followed the expectation under the null (range 4.1%-6.1%, table SV). Only two genes were significantly associated with human intraplaque macrophages, *F10* on chromosome 13 (p-value = 1.00×10^{-6}), and *TNFAIP8L2* on chromosome 1 (p-value = 5.80×10^{-5}) after correction of multiple testing. All gene-based association results for the 659 target genes with plaque characteristics are provided in table SVI.

Validation using pathway analyses

The genes that reached nominal significance in the gene-based analysis of CAD and LAS, were further analyzed using Ingenuity to identify canonical pathways associated with CAD and LAS (table SIV). Table 2 provides the 25 most significant canonical pathways based on the 659 murine target genes and the translation of

6

these genes to human CAD and LAS. The LXR/RXR activation pathway (involved in lipid metabolism) that has been extensively studied in atherosclerotic mice, was found to be significantly enriched for genes associated with CAD and LAS. In contrast, the NFκB inflammatory signaling pathway that has been extensively studied in murine models, revealed less genes that associated with CAD or LAS (table 2). Another example is the T-lymphocyte differentiation pathway that has been extensively studied in mice and associated with murine atherosclerosis, but for which we found little supportive evidence in our gene-based analysis associating with human CAD or LAS.

eQTL analysis of the murine genes
In murine models the 659 genes are clearly affected through knockout, transgenic techniques or targeted compound treatment. A close human analogue of such an effect would be *cis*-acting common genetic variants affecting gene expression through reducing or upregulating expression. Such variants are known as expression quantitative trait loci (eQTLs). We conducted *cis*-eQTL analyses using the STAGE study and queried three online public resources (table SVII) to identify common variants modulating the expression of the 659 genes in humans. Across the four datasets and 11 cell/tissue types we found eQTLs (p-value ≤ 7.31×10^{-7} after correction for 68,402 variants in and around these genes) for 411 out of the 659 genes (these genes are marked orange in table SVIII). For all genes we found SNPs in *cis* affecting expression (p-value < 0.05) in any of the four datasets queried (table SVIII). In a representative example of a canonical pathway (NFκB) we show that each target gene has a *cis*-eQTL, *i.e.* a common variant (SNP) in or around the gene that has a significant effect on tissue-specific gene expression in humans.

Not all genes will exert their effect on clinical outcome via gene expression, rather gene function. Thus, we compared the gene-based association results for CAD and LAS for the genes with a valid eQTL (p-value$_{eQTL}$ < 7.31×10^{-7}) and without (p-value$_{eQTL}$ ≥ 7.31×10^{-7}), but found no significant difference in enrichment for disease association in either CAD (p-value = 0.882) or LAS (p-value = 0.634). Similarly, we compared the gene-based association results between the model groups for the 84 and 42 genes that associated with CAD and LAS (knockout, transgenic, or compounds) and found a statistical difference between groups for CAD (p-value = 0.0013, Chi^2 = 13.291). A pairwise comparison of the three model groups revealed that knockout and compounds do not have statistically different gene-based results (p-value = 0.718), whereas the results between knockouts and transgenic mice (p-value = 0.00049, Chi^2 = 12.146) or transgenic mice and compounds do (p-value = 0.0029, Chi^2 = 8.852). We found no statistical difference for LAS (p-value = 0.637) when comparing the three model groups. We also stratified our pathway analyses based on these three model groups, but this revealed no additional significant pathways.

Discussion

This study shows that putative atherosclerosis-causing genes identified in murine atherosclerosis models, reveal little association with human CAD, LAS or atherosclerotic plaque characteristics, despite the fact that the majority of these genes have *cis*-acting eQTLs.

Overall, the majority of genes associated with an atherosclerotic phenotype in mice do not carry variants that associate with human CAD, LAS or advanced plaque characteristics. In contrast, murine genes involved in lipid metabolism significantly associated with human CAD which is consistent with the known role of lipids in CVD risk. Indeed, lipid-lowering drugs, such as statins, act through *HMGCR* (chromosome 5q13.3) to lower circulating lipids[13]. A recent GWAS showing that variants in the *HMGCR* locus are associated with an increase of 2.84 mg/dL total cholesterol[14], effectively confirmed this drug action retrospectively. However, for most murine genes and pathways of innate and adaptive immunity, there was no association with human CAD, LAS or plaque characteristics.

ApoE[-/-] and LDLR[-/-] models are widely used to study the initiation and progression of atherosclerosis. To study the effect on plaque development[15] or therapeutic strategies[16], additional (double) knockout or transgenic atherosclerotic models have been developed. Yet, they lack plaque rupture thrombosis and subsequent cardiac or cerebral ischemia. Although therapeutic strategies have been developed based on animal models, failures in clinical utilization underscores the need for human verification and translation before initiating targeted drug development programs[17,18]. Indeed, a *post hoc* analysis by deCODE genetics and Amgen assessed the validity of results from human genetic studies as positive predictors of successful clinical trials[19]. Essentially all failed clinical trials targeting a gene (locus) lack any evidence from genetic association studies[19].

There are several potential explanations for the observed discrepancies of genes involved in atherosclerotic murine studies and human cardiovascular disease. First, they may be explained by differences in effect sizes. Common genetic variants associated with human disease often have a modest effect, in contrast to experimental gene manipulation (i.e. knockout) in animal models to study atherosclerotic disease. These genetic modifications in mice limit inferences regarding dose dependent effects which is relevant for predicting drug effects in human disease. Furthermore, the combined effect of human population history and selection may have yielded very little functional genetic variation and thus no association with CAD or LAS, even in a large sample. The low number of murine genes associated with plaque characteristics in the Athero-Express study may be explained by limited sample size and therefore should be interpreted with caution. However, for our gene-based analysis of CAD and LAS, we had access to GWAS results based on large mete-analyses, providing substantial statistical power to detect genes associated with disease. In addition, we studied the enrichment of murine derived atherosclerosis-related gene sets within canonical pathways. We then tested whether their human orthologues (that were nominally associated with disease based on a gene-based test) were also observed in these pathways. For the majority of pathways, weak or even absent evidence was found in humans. Of note is the LXR pathway which is proven to be linked with human cardiovascular disease,

and for which we found ample evidence using our gene-based and our pathway analysis. Furthermore, in atherosclerotic mouse models, plasma lipid levels are the main determinant of lesion development. The associations between mice and human GWAS studies may improve when human individuals are studied with (genetic) susceptibility to abnormal lipid metabolism. Such interactions have not been explored in the present study. However, such an association would imply that the atherosclerotic mouse models cannot represent disease development in the general population.

Second, one might argue that the SNP to gene mapping done by VEGAS is over-conservative and consequently excludes (regulatory) variants of greater effect that may lie as far as 1Mb. However, most variants significantly affecting gene expression are found within 50 kb of the gene body[20], and regulatory elements are usually found in intergenic regions. Nevertheless, it is unlikely that erroneous annotations alone can explain the strong discrepancy between murine genes and their human orthologues within pathways.

Third, undoubtedly the Ingenuity Knowledge Base is a comprehensive summary of the current knowledge from literature on gene networks and pathways. While it is constantly updated and manually curated, our Ingenuity based pathway analysis is biased. Indeed, genes (and thereby networks and pathways) that are not studied (in humans or any model system) would simply not exist in Ingenuity, thus partly explaining an apparent discrepancy between associated pathways in humans and mice. In addition, pathways are often cell and organ specific and a selective approach towards cell types that are considered to play a dominant role in atherogenesis may affect the readout of our pathway analyses.

Fourth, the annotation of candidate genes to GWAS loci is usually based on literature and proximity to the genome-wide significant SNP, but whether such a gene actually influences disease is still unknown. For our gene-based approach we mapped the mouse genes to their human orthologue and tested their association with CAD or LAS. Therefore, the genes that have been annotated to GWAS loci may not appear in our results even when the SNPs in their proximity meet the genome-wide significance threshold.

Lastly, a transgenic or compound-treated mouse may not be comparable to a knockout model, thus explaining the lack of association of these genes with human disease. Given the statistically different gene-based results, we stratified our pathway analyses based on these three different model types, and found no difference. Hence, it is unlikely that the type of murine intervention model explains the lack of association with human disease.

Previous studies have raised doubts upon the validity of translating murine models to human pathophysiology in studying immune responses[21-23] although conflicting results have been obtained[24]. Our study adds to the debate on the relevance of murine models for human atherosclerotic disease, and supports the view that the direction of the scientific process matters. The lack of association of murine atherosclerotic genes with human CAD, LAS or human plaque characteristics may be due to diverged gene expression patterns among mice and humans resulting in different phenotypic effects. Indeed, results from the recently published Mouse ENCODE project show that gene expression patterns diverge between mice and

9

humans[25,26]. Moreover, phenotypic effects of orthologous genes frequently differ between species. This suggests that a sensible approach would be to group genes based on mouse-human orthology in order to improve the translational power of putative murine genes. Genome-wide association studies agnostically provide human evidence for the involvement of genetic loci in the underlying mechanisms of atherosclerotic disease, including CAD and LAS. Therefore, from the outset, the rationale to initiate an examination of the role of genes in these loci in murine models of atherosclerosis could be supported by human data.

Our study may have several limitations. Individuals of non-European descent are underrepresented in the GWAS we examined. It remains to be investigated if results will differ for GWAS in non-European cohorts.

The genetic association study on human plaque characteristics was executed in the Athero-Express study and included 1,443 patients with significant atherosclerosis. Although this study represents the largest collection of histologically investigated plaques, this study may suffer from limited power when examining the genetics of plaque characteristics.

We studied all published papers that applied the ApoE$^{-/-}$ and LDLR$^{-/-}$ as a model for atherosclerosis. However, alternative murine models have been studied in atherosclerotic disease that we did not include in our search and we cannot exclude that these reveal better associations with the human genetic outcome studies.

In conclusion, a role for genes that has been observed in mice for atherosclerotic lesion development could scarcely be confirmed studying associations of disease development with common human genetic variants indicating that knockout models of atherosclerosis are not a good reflection of the variation that underlies common forms of atherosclerosis. The value of murine atherosclerotic models for selection of therapeutic targets in human disease remains unclear.

10

Disclosures
Johan L.M. Björkegren is founder, main shareholder and chairman of the board for Clinical Gene Networks AB (CGN) and Tom Michoel is shareholder. CGN has an invested interest in microarray data generated from the STAGE cohort. Cavadis B.V. financed genotyping of AEGS1. Gerard Pasterkamp and Dominique P.V. de Kleijn are founders and stockholders of Cavadis B.V.

References

1. Plump AS, Smith JD, Hayek T, Aalto-Setälä K, Walsh A, Verstuyft JG, Rubin EM, Breslow JL. Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. *Cell.* 1992;71:343-353.

2. Ishibashi S, Goldstein JL, Brown MS, Herz J, Burns DK. Massive xanthomatosis and atherosclerosis in cholesterol-fed low density lipoprotein receptor-negative mice. *J Clin Invest.* 1994;93:1885-1893. doi:10.1172/JCI117179.

3. Thüsen von der JH, van Berkel TJ, Biessen EA. Induction of rapid atherogenesis by perivascular carotid collar placement in apolipoprotein E-deficient and low-density lipoprotein receptor-deficient mice. *Circulation.* 2001;103:1164-1170.

4. Cheng C, Tempel D, van Haperen R, van der Baan A, Grosveld F, Daemen MJAP, Krams R, de Crom R. Atherosclerotic lesion size and vulnerability are determined by patterns of fluid shear stress. *Circulation.* 2006;113:2744-2753. doi:10.1161/CIRCULATIONAHA.105.590018.

5. Kitami T, Nadeau JH. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nature Genetics.* 2002;32:191-194. doi:10.1038/ng945.

6. The International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005;437:1299-1320. doi:10.1038/nature04226.

7. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics.* 2005;6:95-108. doi:10.1038/nrg1521.

8. Schunkert H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics.* 2011;43:333-338. doi:10.1038/ng.784.

9. Traylor M, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. 2012;11:951-962. doi:10.1016/S1474-4422(12)70234-X.

10. Liu JZ, Mcrae AF, Medland SE, Wray NR, Brown KM, AMFS Investigators, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet.* 2010;87:139-145. doi:10.1016/j.ajhg.2010.06.009.

11. Verhoeven BAN, Velema E, Schoneveld AH, de Vries J-PPM, de Bruin P, Seldenrijk CA, de Kleijn DPV, Busser E, van der Graaf Y, Moll F, Pasterkamp G. Athero-express: differential atherosclerotic plaque expression of mRNA and protein in relation to cardiovascular events and patient characteristics.

12

Rationale and design. *Eur J Epidemiol*. 2003;19:1127-1133. doi:10.1007/s10564-004-2304-6.

12. Hellings WE, Peeters W, Moll FL, Piers SRD, van Setten J, van der Spek PJ, de Vries JPPM, Seldenrijk KA, De Bruin PC, Vink A, Velema E, de Kleijn DPV, Pasterkamp G. Composition of Carotid Atherosclerotic Plaque Is Associated With Cardiovascular Outcome: A Prognostic Study. *Circulation*. 2010;121:1941-1950. doi:10.1161/CIRCULATIONAHA.109.887497.

13. Stancu C, Sima A. Statins: mechanism of action and effects. *J Cell Mol Med*. 2001;5:378-387.

14. Teslovich TM, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-713. doi:10.1038/nature09270.

15. Plump AS, Masucci-Magoulas L, Bruce C, Bisgaier CL, Breslow JL, Tall AR. Increased atherosclerosis in ApoE and LDL receptor gene knock-out mice as a result of human cholesteryl ester transfer protein transgene expression. *ATVB*. 1999;19:1105-1110. doi:10.1161/01.ATV.19.4.1105.

16. Schwartz GG, et al. Effects of dalcetrapib in patients with a recent acute coronary syndrome. *N Engl J Med*. 2012;367:2089-2099. doi:10.1056/NEJMoa1206797.

17. Zacho J, Tybjaerg-Hansen A, Jensen JS, Grande P, Sillesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. *N Engl J Med*. 2008;359:1897-1908. doi:10.1056/NEJMoa0707402.

18. Voight BF, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012;380:572-580. doi:10.1016/S0140-6736(12)60312-2.

19. Kamb A, Harper S, Stefansson K. Human genetics as a foundation for innovative drug development. *Nature Genetics*. 2013;31:975-978. doi:10.1038/nbt.2732.

20. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavaré S, Deloukas P, Dermitzakis ET. Population genomics of human gene expression. *Nature Genetics*. 2007;39:1217-1224. doi:10.1038/ng2142.

21. Rice J. Animal models: Not close enough. *Nature*. 2012;484:S9.

22. de Souza N. Model organisms: Mouse models challenged. *Nature methods*. 2013;10:288-288. doi:10.1038/nmeth.2429.

23. Seok J, et al. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110:3507-3512.

doi:10.1073/pnas.1222878110.

24. Takao K, Miyakawa T. Genomic responses in mouse models greatly mimic human inflammatory diseases. *Proceedings of the National Academy of Sciences of the United States of America*. 2015;112:1167-1172. doi:10.1073/pnas.1401965111.

25. Koch L. Genomics: The best-laid plans of mice and men. *Nature Reviews Genetics*. 2014;16:1-1. doi:10.1038/nrg3873.

26. Yue F, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature*. 2014;515:355-364. doi:10.1038/nature13992.

Significance

While many putatively atherosclerosis-causing genes have been identified through the genetically modified mouse model, the translation to human atherosclerotic disease has been notoriously challenging. Here we systematically reviewed literature to identify putative atherosclerosis-causing genes in mice. We identified their human orthologs and performed gene-based association tests on human coronary artery disease, ischemic stroke and plaque characteristics. Of the 659 identified genes, 11 were associated with CAD, none with stroke, and 2 with intraplaque macrophages. Pathway analyses confirmed the limited association with human disease, despite the fact that many genes have a cis-eQTL. Our study underlines the need for human validation of murine atherosclerosis-causing genes.

Gene	Chr	start	end	CARDIoGRAM					METASTROKE				
				n SNPs	P	FDR	best SNP	P best SNP	n SNPs	P	FDR	best SNP	P best SNP
CDKN2A	9	21.957.750	21.984.490	100	**<2.00x10-6**	**<2.00x10-6**	rs10738604	2.27x10-16	120	*0,006*	0,380	rs634537	1.05x10-4
CDKN2B	9	21.992.901	21.999.312	92	**<2.00x10-6**	**<2.00x10-6**	rs10738604	2.27x10-16	107	*0,003*	0,349	rs634537	1.05x10-4
MRAS	3	139.549.314	139.607.067	81	**<2.00x10-6**	**<2.00x10-6**	rs2306374	3.34x10-8	93	0,637	0,911	rs10513053	0,026
APOA5	11	116.165.295	116.167.794	77	**2.00x10-6**	**3.29x10-4**	rs964184	8.02x10-10	114	0,533	0,891	rs5110	9.21x10-3
LPL	8	19.840.861	19.869.050	177	**5.00x10-6**	**6.57x10-4**	rs3779788	2.40x10-7	199	0,079	0,753	rs2410616	4.34x10-3
TGFB1	19	46.528.490	46.551.656	54	**7.00x10-6**	**7.67x10-4**	rs12327659	3.93x10-6	77	0,095	0,753	rs2217656	8.39x10-4
LDLR	19	11.061.056	11.105.505	101	**9.00x10-6**	**8.45x10-4**	rs1122608	9.73x10-10	119	0,044	0,749	rs1799898	3.14x10-3
LIPA	10	90.963.305	91.001.640	154	**1.30x10-5**	**1.07x10-3**	rs7922269	7.56x10-6	174	0,079	0,753	rs4933497	5.03x10-3
GIP	17	44.390.916	44.400.954	59	**1.50x10-5**	**1.10x10-3**	rs46522	3.57x10-6	69	0,121	0,794	rs4794015	0,020
APOA4	11	116.196.627	116.199.221	65	**1.70x10-5**	**1.12x10-3**	rs964184	8.02x10-10	101	0,353	0,891	rs5110	9.21x10-3
IL25	14	22.911.857	22.915.452	88	*0,046*	0,370	rs10143597	2.75x10-3	116	*0,001*	0,171	rs12894524	4.52x10-5
MMP12	11	102.238.673	102.250.922	114	0,055	0,408	rs7124926	4.25x10-3	132	*0,001*	0,171	rs660599	4.74x10-5
MMP3	11	102.211.737	102.219.552	135	0,079	0,454	rs2408489	5.72x10-3	158	*0,001*	0,171	rs475007	4.05x10-5
MAPK7	17	19.221.658	19.227.445	27	0,119	0,509	rs739830	1.53x10-3	46	*0,006*	0,380	rs1110467	1.27x10-4
MMP9	20	44.070.953	44.078.607	82	0,126	0,509	rs7270354	1.94x10-4	95	*0,005*	0,380	rs17448653	1.06x10-4
ALOX15B	17	7.883.082	7.893.176	90	0,202	0,635	rs4792214	6.92x10-3	111	*0,006*	0,380	rs4792203	9.36x10-5
CD40	20	44.180.312	44.191.791	123	0,253	0,689	rs1321003	8.17x10-3	142	*0,000*	0,091	rs4239702	1.41x10-5
TGFB3	14	75.494.194	75.517.845	67	0,902	0,965	rs8008060	0,090	82	*0,000*	0,120	rs3917187	4.16x10-5

Table 1: Top 10 gene-based association results of the 659 target genes for CAD and LAS. *CARDIoGRAM*: meta-analysis of GWAS of CAD. *METASTROKE*: meta-analysis of GWAS of LAS. Results are given per *gene* with its chromosomal (*chr*) *start* and *end* base pair

position. *n SNPs*: number of SNPs studied for that gene. Also given are the most significant (*best*) SNP for that gene and its p-value. P-values are in **bold** for genes with Bonferroni corrected p-value$_{gene}$ ≤ 7.59x10^{-5}, and *italic* for genes with nominal p-value < 0.05.

Rank	Ingenuity canonical pathways	Overlap pathways	CARDIoGRAM		METASTROKE		Murine genes	
			N genes p ≤ 0.05	P	N genes p ≤ 0.05	P	N genes p ≤ 0.05	N genes p > 0.05
1	LXR/RXR Activation	59/121 (49%)	16	1,84E-08	5	0,098	19	40
2	FXR/RXR Activation	50/125 (40%)	12	4,22E-06	1	0,202	13	37
3	PPARα/RXRα Activation	40/165 (24%)	10	1,78E-05	2	0,278	12	28
4	Clathrin-mediated Endocytosis Signaling	27/184 (15%)	8	3,27E-05	4	0,034	10	17
5	Acute Phase Response Signaling	53/168 (32%)	11	4,32E-05	2	0,252	13	40
6	Glioma Invasiveness Signaling	11/57 (19%)	5	1,06E-04	1	0,329	6	5
7	Atherosclerosis Signaling	58/120 (48%)	11	1,17E-04	8	6,33E-03	17	42
8	Oncostatin M Signaling	13/34 (38%)	5	2,67E-04	3	0,021	7	6
9	IL-12 Signaling and Production in Macrophages	47/131 (36%)	9	3,79E-04	4	0,123	12	35
10	TR/RXR Activation	21/85 (25%)	6	3,93E-04	1	0,376	6	15
11	Hepatic Fibrosis / Hepatic Stellate Cell Activation	70/196 (36%)	11	5,12E-04	6	0,077	15	55
12	Adipogenesis pathway	22/124 (18%)	6	5,13E-04	2	0,207	7	15
13	Cellular Effects of Sildenafil (Viagra)	9/124 (7%)	4	6,09E-04	0	0,630	4	5
14	VEGF Signaling	23/89 (26%)	6	6,59E-04	0	0,307	6	17
15	Hepatic Cholestasis	51/158 (32%)	9	6,89E-04	3	0,222	11	40
16	Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	42/179 (23%)	8	8,06E-04	3	0,194	10	32
17	p70S6K Signaling	18/118 (15%)	5	1,37E-03	1	0,376	6	12
18	Relaxin Signaling	19/132 (14%)	5	1,77E-03	1	0,377	6	13
19	LPS/IL-1 Mediated Inhibition of RXR Function	38/208 (18%)	7	2,01E-03	2	0,277	9	29
20	Glioma Signaling	20/94 (21%)	5	2,24E-03	2	0,189	5	15
21	Nitric Oxide Signaling in the Cardiovascular System	20/95 (21%)	5	2,24E-03	1	0,377	6	14
22	Endothelin-1 Signaling	30/167 (18%)	6	2,71E-03	3	0,127	8	22
23	VEGF Family Ligand-Receptor Interactions	21/76 (28%)	5	2,80E-03	0	0,341	5	16
24	eNOS Signaling	21/135 (16%)	5	2,80E-03	1	0,376	6	15
25	Renal Cell Carcinoma Signaling	13/69 (19%)	4	2,82E-03	0	0,513	4	9

Table 2: Top-25 canonical pathways for murine target genes. For each *Ingenuity canonical pathway*, the number and percentage of murine genes overlapping with the total number of pathways was determined (*Overlap*). For coronary artery disease (*CARDIoGRAM*) and large artery stroke (*METASTROKE*) the number of murine genes reaching $p \leq 0.05$ (*N genes p ≤ 0.05*) for association with the respective disease is given. P indicates the binomial p-value and represents the proportion of murine genes associated with the disease compare to the total number of murine genes. For all murine genes the total *N genes p ≤ 0.05* and *N genes p > 0.05* for association with either disease is given.

Supplemental Material

Human validation of genes associated with a murine atherosclerotic phenotype

Gerard Pasterkamp, Sander W. van der Laan, Saskia Haitjema, Hassan Foroughi Asl, Marten A. Siemelink, Tim Bezemer, Jessica van Setten, Martin Dichgans, Rainer Malik, Bradford B. Worrall, Heribert Schunkert, Nilesh J. Samani, Dominique P.V. de Kleijn, Hugh S. Markus, Imo E. Hoefer, Tom Michoel, Saskia C.A. de Jager, Johan L.M. Björkegren, Hester M. den Ruijter, Folkert W. Asselbergs

Materials and methods

Literature search

A systematic search was performed at PubMed MEDLINE (http://www.ncbi.nlm.nih.gov) and Embase (http://www.embase.com) until July 1st 2014 using the key words (or synonyms and thereof): "ApoE$^{-/-}$", "LDLR$^{-/-}$", "atherosclerosis", "plaque", and "mice". After removal of duplicates we identified 11,219 publications (fig. SI).

Based on title and abstract, publications were manually selected that met the following criteria:

- Murine knockout model on an atherosclerotic background (either ApoE$^{-/-}$ or LDLR$^{-/-}$) that resulted in altered atherosclerotic plaque characteristics or altered plaque volume.
- Murine transgenic model on atherosclerotic background that resulted in altered atherosclerotic plaque characteristics or altered plaque volume
- Murine model on atherosclerotic background with a targeted intervention that resulted in altered plaque characteristics or plaque volume, either treated with a chemical compound or a biological entity (e.g. an antibody, hormone, siRNA or morpholino) that is recognized for protein or gene specificity.

Gene selection for human extrapolation

In total 2,076 papers met our predefined criteria (fig. SI) and we distilled 703 unique murine genes (table SI). Subsequently these genes were mapped to their human orthologues using an automatic script together with the search function of *GeneCards* (http://genecards.org). Results were manually checked for accuracy.

When a single murine gene mapped to multiple human orthologous genes, all human orthologues were included. We grouped each gene into one of three hierarchical categories (knockout, transgenic or compound) if any of the associated articles reported the gene in a knockout model, a transgenic model, or as a specific target of a chemical compound or biological entity (table SI). We excluded genes that could not be mapped by the gene-based association analysis software (n=21) or mapped to the X-chromosome genes (n=23). A total of 659 murine genes could be mapped to a human orthologue on genome build 36 and was thus available for downstream analyses (table SI).

Athero-Express Biobank Study: plaque collection and phenotyping

The details of the study-design and the plaque phenotyping have been described elsewhere[1]. In short, carotid plaque specimens were obtained from

carotid endarterectomy (CEA) patients during surgery. Plaques were immediately processed in the laboratory, where the culprit lesion with a length of 5 mm was fixed in 4% formaldehyde, subsequently followed by decalcification and embedding in paraffin. Cross-sections (5 µm) were sliced and routinely stained for different characteristics: atheroma size (based on interpretation of hematoxylin and eosin (HE), elastica von Gieson and collagen staining (picrosirius red)), macrophages (CD68), smooth muscle cells (α smooth muscle actin (SMA)), collagen, calcification (assessed using HE), intraplaque haemorrhage (HE and fibrin), and intraplaque vessel density (CD34)[2]. Collagen and calcified regions were semi-quantitatively scored as absent/minor vs. moderate/ heavy staining. Atheroma size was semi-quantitatively analyzed as <40% vs. >40% intraplaque fat content. CD68 and SMA were visualized with DAB (3,3'-diaminobenzidine), and were quantitatively analyzed using AnalySIS 3.2 software (Soft Imaging Systems GmbH, Münster, Germany) and expressed as % of plaque area. Likewise CD34 was visualized using DAB and the number of vessels per 3-4 hotspots per plaque was determined. Intraplaque haemorrhage was semi-quantitatively scored as no vs. yes.

Genome-wide association study summary statistics of CAD and LAS
To test the association of genes with atherosclerotic disease, we obtained summary statistics from GWAS on the traits of interest as follows. Data for CAD were downloaded from the CARDIoGRAMplusC4D website (http://www.cardiogramplusc4d.org). These data are the results from CARDIoGRAM, a meta-analysis of 14 GWAS on CAD comprising of 22,233 cases and 64,762 controls of European descent[3]. Data for LAS were obtained from METASTROKE[4] (http://www.strokegenetics.com/members-area/meta-stroke), a meta-analysis of data from 15 ischemic stroke cohorts with a total of 12,389 cases and 62,004 controls, all of European ancestry. In METASTROKE 2,167 cases were determined to be of the "large artery stroke" subtype according to the TOAST classification system[5], with matching 49,159 controls. More details on genotyping, imputation, and study inclusion of the CARDIoGRAM and METASTROKE meta-analyses of GWAS can be found in the respective publications[3,4].

Athero-Express Biobank Study: genotyping
The targeted SNP based analyses *in-silico* analyses were performed using data from two imputed genome-wide genotyping experiments carried out in 1,858 consecutive patients from the AE. For these experiments DNA was extracted from blood or plaque samples (when no blood was available) following standardized in-house validated protocols. The first dataset (Athero-Express Genomics Study 1, AEGS1) was genotyped using Affymetrix Genome-Wide Human SNP Array 5.0, the second dataset (Athero-Express Genomics Study 2, AEGS2) was genotyped using the Affymetrix Axiom® GW CEU 1 Array. We adhered to community standard quality control and assurance (QCA) procedures to clean the whole-genome data obtained in AEGS1 and AEGS2[6]. We used the HapMap 2 CEU release encompassing over 2.5 million SNPs as the reference panel for imputation for autosomal missing genotypes in the 1,443 individuals that passed the QC. Depending on the phenotype we applied linear or logistic regression models adjusting for

age, sex, year of surgery, chip-type, and 10 principal components. We assumed an additive genetic model.

We obtained summary level data from targeted SNP-based analyses (focused on SNPs in and around [±50kb] the 659 genes) on 7 plaque characteristics in the Athero-Express Biobank Study (AE). The local ethical committee approved the study and all patients gave written informed consent after the nature and possible consequences of the study were explained.

eQTL of human orthologues of selected murine genes
There is no human analogue for murine genetically modified models, other than naturally occurring genetic variation that affects gene expression (known as expression quantitative trait locus, eQTL). Therefore, we identified eQTLs through two approaches. First, we queried three online resources (table SVI) for eQTLs in lymphoblastoid cells[7,8], monocytes[9], subcutaneous adipose tissue[7], and skin tissue[7]. Secondly, we performed eQTL analyses in 7 tissue types that could be relevant for disease development in the STAGE study[10].

The Stockholm Atherosclerosis Gene Expression (STAGE) Study
In the STAGE study, seven vascular and metabolic tissues of well-characterized coronary artery disease (CAD) patients were sampled during coronary artery bypass grafting (CABG) [10]. The samples from atherosclerotic arterial wall (AAW), internal mammary artery (IMA), liver, skeletal muscle (SM), subcutaneous fat (SF), visceral fat (VF), and fasting whole blood (WB), were obtained during CABG for DNA and RNA isolation. Patients were included if they were eligible for CABG and had no other severe systemic diseases (e.g. widespread cancer or active systemic inflammatory disease).

In order to prepare inferred genotypes for STAGE for genotype imputation, SNPs were quality controlled for minor allele frequency MAF < 5%, Hardy-Weinberg equilibrium (HWE) p-value < 1×10^{-6}, and call rate of 100%. Thereafter, genotypes for the STAGE study were imputed using IMPUTE2 using 1000 Genomes EUR as the reference[11]. Quality control measures for imputed genotypes used an additional filter of IMPUTE2 INFO score of < 0.3. This yielded a total of 5,473,585 SNPs. The Ethical committee of the Karolinska Hospital approved the study, and all patients gave written informed consent after the nature and possible consequences of the study were explained.

As previously described(9), an expression trait was tested for association with each genotyped and imputed SNP using Kruskal-Wallis test and false discovery rate to correct for multiple testing. First, all *cis*-pairs of SNPs within 50kb of the transcription start or end site for each gene were identified. Next, *cis* SNP-gene pairs were tested for association in all seven STAGE tissues using kruX[12]. The p-value for eQTL inclusion in kruX was set at 0.05. Finally, an empirical FDR estimate for each eQTL-gene pair was calculated using ten permutations by shuffling patient IDs on genotype data. As a result, the most significant eQTL-gene association in each tissue was reported.

Gene based association study of target genes
We used "a versatile gene-based association study" (VEGAS)[13] to calculate gene-based association statistics from the summary statistics of each target

gene for each trait. The details of the methods applied by VEGAS have been described elsewhere[13]. In short, SNPs are mapped to the gene (in and around ±50kb from 5' and 3' gene borders), and using the GWAS p-value a gene-based test statistic is calculated corrected for the underlying population linkage disequilibrium structure (based on HapMap 2 CEU). Finally using simulations an empirical gene-based *p*-value of association with the phenotype is calculated per gene[13].

Pathway analysis
The 659 genes were analyzed through the use of QIAGEN's Ingenuity Pathway Analysis (IPA, 2014 winter version, QIAGEN Redwood City, www.qiagen.com/ingenuity)[14]. We used IPA to identify canonical biological pathways within the Ingenuity Knowledge Base to which the murine gene targets were mapped. Only direct relationships that were experimentally observed in humans were included in the analysis.

Statistical analyses
This study focused on 659 human orthologous genes for the atherosclerotic murine genes and thus, in our gene-based analyses of the atherosclerotic disease traits (CAD, LAS, and the 7 plaque characteristics), we corrected conservatively for multiple testing (p-value$_{gene}$ = 0.05/659 = 7.59x10^{-5}). However, to allow optimal description of relevant biological pathways (in IPA) with the trait we also report data using a nominal p-value < 0.05. In the public eQTL sources there is no information on the number of SNPs mapped in and around ±50kb of the genes, we only were able to obtain the total number of eQTLs. STAGE uses 1000G imputed data and thus has denser SNP coverage per gene. In STAGE we mapped a total of 68,402 independent common SNPs in and around ±50kb of the genes for the *cis*-eQTL analysis. Therefore based on these numbers, we conservatively set the threshold at p-value$_{eQTL}$ = 0.05/68,402 = 7.31x10^{-7}. We queried the 4 eQTL studies and noted the total number of nominally associated (p<0.05) eQTLs (within the ±50kb range) with the target gene expression. We also report the most significantly associated eQTL for each gene. The three mouse model groups (knock-out, transgenic and compound) were compared using chi-square tests.

List of Figures and Tables

Fig. SI. Flow-chart showing the filtering and selection of relevant papers and the identification of murine atherosclerotic genes mapped to human orthologues.

Table SI. All 703 human orthologous genes identified in murine models.

Table SII: All gene-based association results of the 659 target genes for CAD and LAS.

Table SIII: Canonical pathways to which any of the 659 genes were mapped.

Table SIV: Number of genes with $p < 0.05$ in the gene-based association study of plaque phenotypes.

Table SV: Gene-based association results for the atherosclerotic plaque characteristics in the Athero-Express Biobank Study.

Table SVI: Background information of the publically, online available eQTL datasets queried.

Table SVII: *Cis*-eQTL results from the analysis in STAGE and the online query of various datasets.

Figure SI. Flow-chart showing the filtering and selection of relevant papers and the identification of murine atherosclerotic genes mapped to human orthologues. A total of 11,219 papers were screened, and 2,076 met our predefined criteria. From those 2,076 papers we distilled 703 unique genes or compounds targeting genes on which we applied filters as described in the methods. Finally we mapped 659 murine genes to their human orthologue. *VEGAS*: a versatile gene-based association study software. *GWAS*: genome-wide association study.
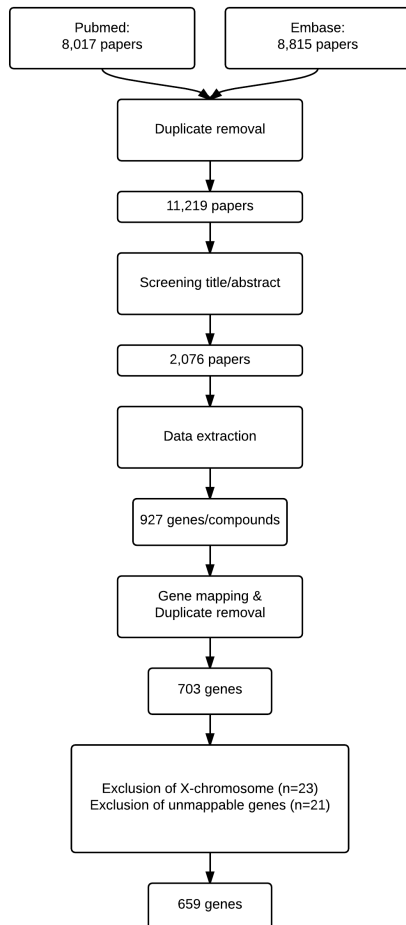
Supplemental Tables

Supplemental tables SI through SVII are available in a separate Excel-file.

Supplemental References

1. Verhoeven BAN, Velema E, Schoneveld AH, de Vries J-PPM, de Bruin P, Seldenrijk CA, de Kleijn DPV, Busser E, van der Graaf Y, Moll F, Pasterkamp G. Athero-express: differential atherosclerotic plaque expression of mRNA and protein in relation to cardiovascular events and patient characteristics. Rationale and design. *Eur J Epidemiol*. 2003;19(12):1127-1133. doi:10.1007/s10564-004-2304-6.

2. van den Borne P, van der Laan SW, Bovens SM, Koole D, Kowala MC, Michael LF, Schoneveld AH, van de Weg SM, Velema E, de Vries J-P, de Borst GJ, Moll FL, de Kleijn DPV, Quax PHA, Pasterkamp G. Leukotriene b4 levels in human atherosclerotic plaques and abdominal aortic aneurysms. *PLoS ONE*. 2014;9(1):e86522. doi:10.1371/journal.pone.0086522.

3. Schunkert H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics*. 2011;43(4):333-338. doi:10.1038/ng.784.

4. Traylor M, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. 2012;11(11):951-962. doi:10.1016/S1474-4422(12)70234-X.

5. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24(1):35-41. doi:10.1161/01.STR.24.1.35.

6. Laurie CC, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol*. 2010;34(6):591-602. doi:10.1002/gepi.20516.

7. Nica AC, et al. The Architecture of Gene Regulatory Variation across Multiple Human Tissues: The MuTHER Study. Barsh G, ed. *PLoS Genet*. 2011;7(2):e1002003. doi:10.1371/journal.pgen.1002003.

8. Liang L, Morar N, Dixon AL, Lathrop GM, Abecasis GR, Moffatt MF, Cookson WOC. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Research*. 2013;23(4):716-726. doi:10.1101/gr.142521.112.

9. Zeller T, Wild PS, Schillert A, Maouche S, Schnabel RB, Ziegler A, Blankenberg S, Cambien F. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS ONE*. 2009;5(5):e10693-e10693. doi:10.1371/journal.pone.0010693.

10. Hägg S, et al. *PLoS Genet*. 2009;5(12):e1000754. doi:10.1371/journal.pgen.1000754.s013.

11. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet*. 2009;5(6):e1000529. doi:10.1371/journal.pgen.1000529.

12. Qi J, Asl HF, Björkegren J, Michoel T. kruX: matrix-based non-parametric eQTL discovery. *BMC Bioinformatics*. 2014;15(1):11. doi:10.1186/1471-2105-15-11.

13. Liu JZ, Mcrae AF, Medland SE, Wray NR, Brown KM, AMFS Investigators, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*. 2010;87(1):139-145. doi:10.1016/j.ajhg.2010.06.009.

14. Calvano SE, et al. A network-based analysis of systemic inflammation in humans. *Nature*. 2005;437(7061):1032-1037. doi:10.1038/nature03985.