Fall 2024

Identification of Inflammatory Factors Contributing to Foaming in Atherosclerosis

Kavya Podila

Identification of Inflammatory Factors Contributing to Foaming in Atherosclerosis

Kavya Podila[a], Alyssa Matz [b], Beiyan Zhou, PhD[b]

[a]*Holster Scholars program, University of Connecticut, Storrs, CT*
[b]*Zhou Lab, Department of Immunology, University of Connecticut School of Medicine, Farmington, CT*

Key Finding

AtheroSpectrum enabled identification of pathogenic foaming drivers that specifically enhance gene profiles associated with atherosclerotic cardiovascular disease.

Abstract

According to the National Institute of Health (NIH), atherosclerosis is attributed to the majority of cardiovascular disease (CVD) cases and is the leading cause of mortality in the US and worldwide. Despite significant medical advancements, the risk of cardiovascular events remains, warranting further mechanistic and pathogenic investigation. Advances in high-resolution omics profiling and the development of novel data-analytic bioinformatic tools previously enabled the novel discovery of two distinct programs in foaming, the central process in atherogenesis. The programs are pathogenic and homeostatic foaming, with the pathogenic foaming program being positively associated with disease. Using single-cell RNA sequencing (scRNA-seq) data from mouse aorta, transcription factors IRF7, ELL2, ARID3A, and MGA were identified as drivers of pathogenic foaming in atherosclerosis. Causal network analysis further revealed eight upstream regulators of these drivers. Expression analysis in symptomatic and asymptomatic sequencing data from independent human and mouse datasets confirmed the

significance of these 12 genes/complexes/channels/receptors. The results of this study provide specific causal regulators for the pathogenesis of atherosclerosis and potentially offer new targetable molecules and their downstream network for novel therapeutic designs that direct treatments toward targeting inflammation.

Introduction

50% of individuals ages 45 to 80 years old have some degree of atherosclerosis, which is the top contributor to cardiovascular disease (CVD)[1]. Atherosclerosis is characterized by the build-up of lipid material, or plaques, in arterial walls, leading to reduced blood flow. The central players in atherogenesis are macrophages, which are functionally heterogeneous immune cells that respond to environmental and cellular signals, especially in obesity-related conditions. One specialized form of macrophages is foam cells, which engulf fatty streaks in the arteries and are a major component of plaques[2]. Most therapeutic interventions for atherosclerosis today focus on lowering cholesterol levels. Despite significant medical advancements in the field, the risk of cardiovascular events remains. This suggests that there are some factors we are missing in controlling this disease.

Inflammation markers have been found to positively correlate with risk of recurrent cardiovascular events[3]. A trial study of Canakinumab, an anti-inflammatory therapeutic monoclonal antibody, was conducted and results revealed that the drug significantly lowered the risk of cardiovascular events[4]. Low doses of colchicine, an anti-inflammatory alkaloid, were also found to be an effective agent in managing atherosclerotic risk[5]. These findings that reducing

inflammation lowers CVD risk suggest inflammation is a missing piece in understanding the mechanisms of atherosclerosis.

The previous understanding in the field was that foam cell formation is linear, in that all foam cells are inflammatory. However, AtheroSpectrum, a novel macrophage annotation tool, recently provided more detailed insights into macrophage action in atherosclerosis. It characterizes cells based on the Macrophage-Derived Foam Index (MDFI) and the Macrophage Polarization Index (MPI). Using these measures of foaming and inflammation, respectively, AtheroSpectrum revealed 4 subpopulations of macrophages in atherosclerotic plaque: noninflammatory foam cell (A), inflammatory foam cell (B), inflammatory nonfoam cell (C), and noninflammatory nonfoam basal cell (D). Genetic information from these groups led to the discovery of 2 distinct foaming programs: homeostatic foaming, which is noninflammatory, and pathogenic foaming, which is inflammatory. Pathogenic foaming was confirmed to contribute to disease progression[6]. However, what separates the two foaming programs is unknown. This project therefore sought to identify the driver genes behind pathogenic foaming in atherosclerosis.

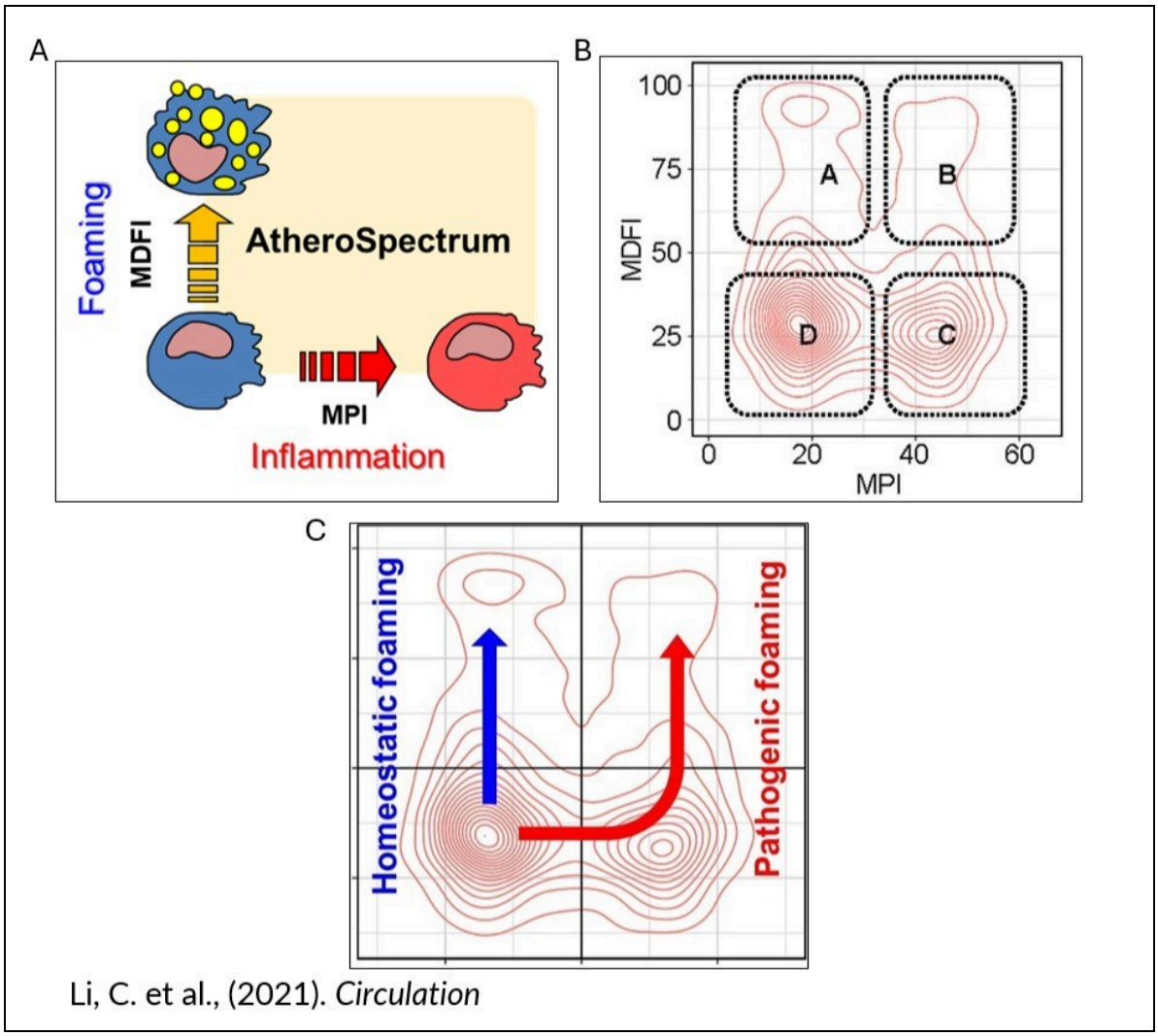


Figure 1 – AtheroSpectrum previously identified foaming programs

A) AtheroSpectrum provides macrophage annotations based on inflammation and foaming indices. B) 4 subpopulations of cells were novelly identified with distinct characteristics. C) Differentiation patterns of foam cells show separate pathways for homeostatic and pathogenic foaming

Methods

Single-cell RNA-sequencing (scRNA-seq) data was accessed from the Gene Expression Omnibus (GEO), an online database for gene expression profiling run by the NIH. The dataset used for discovery was GSE116240, which is from mouse CD45+ aortic cells. To process the data, R version 4.2.2 Patched was used in R Studio[7]. Packages for processing included data.table (1.14.10), dplyr (1.1.4), introdataviz (0.0.0.9003), patchwork (1.2.0), Seurat (5.0.1), stringi (1.8.3), tibble (3.2.1), and tidyr (1.3.0). For all plots, ggplot2 (3.4.4) was used.

The Seurat package was first used to filter the dataset for doublets and low-quality transcripts. Cells with a mitochondrial DNA percentage of less than 7% and a gene count ranging from 1000 to 5000 were retained using Seurat's "subset" function to remove low-quality and damaged cells. The data was then scaled, normalized, and clustered using Uniform Manifold Approximation and Projection (UMAP), an unsupervised non-linear dimensional reduction technique. The differentially expressed features from each cluster were identified using Seurat's "FindMarkers" function, which uses a non-parametric Wilcoxon rank sum test to run differential expression testing on each cluster compared to the rest. The types of cells present were determined by identifying the top markers for each cluster and inputting them into the Annotation of Cell Types web server[8]. Following the identification of each cell group, the clusters were subsetted to extract just the macrophage cluster. AtheroSpectrum was then run as a supervised dimensional reduction technique to identify cells in the homeostatic and pathogenic foaming programs.

To define the Differentially Expressed Genes (DEGs) between the foaming programs, various log2 fold changes and p-value cutoffs were tested. A log2 fold change not equal to 0 and

a p value less than or equal to 0.05 was decided on, based on the optimal number of cells to run pathway analysis using Quiagen's Ingenuity Pathway Analysis (IPA).

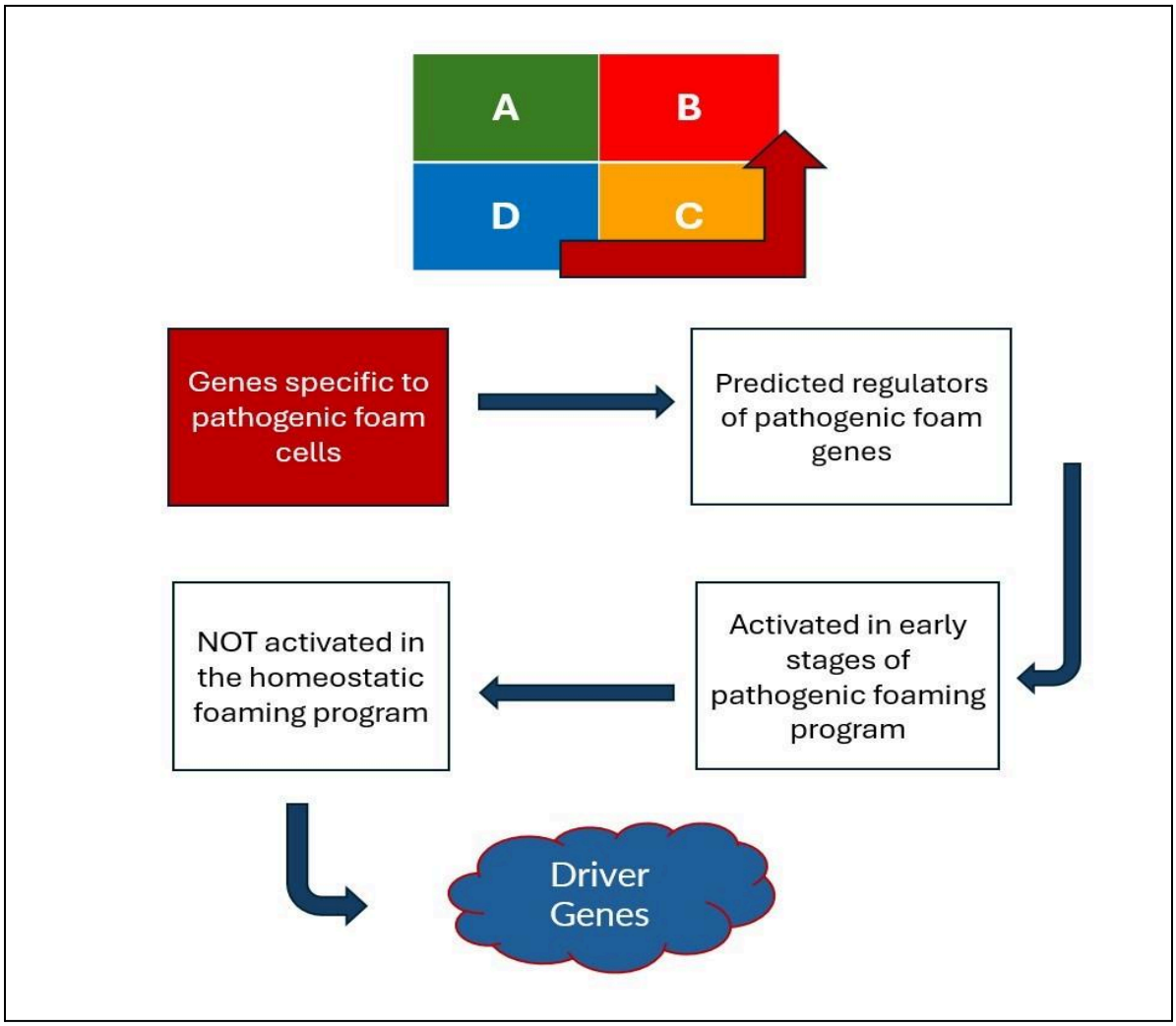


Figure 2 – Workflow for Identification of Driver Genes

Driver genes were defined as the upstream regulators of genes unique to the pathogenic foaming program that are present in early stages of pathogenic foaming (the transition from "D" to "C") and are not active in homeostatic foaming (the transition from "D" to "A").

This filtering cutoff was applied to the differential expression values for genes in the homeostatic foaming program (A-D), genes in the later portion of the homeostatic foaming program (B-C), and the early portion of the pathogenic foaming program (C-D), resulting in 3 lists of DEGs. Then, using dplyr's "anti_join" function, DEGs that were in A to D and not B to C were identified. These were the genes unique to homeostatic foaming. The process was repeated to find the DEGs that were in B to C, but not A to D, to find genes unique to pathogenic foaming. This list of pathogenic foaming genes was then put into IPA. The analysis output was a list of upstream regulators (in human gene form), which are the genes that regulate the expression of other genes in the dataset that regulate the expression of input genes. The upstream regulators were then filtered by z score greater than or equal to |2|, a p-value less than or equal to 0.05, and only including transcription regulators. After filtering, the list was further filtered to only include genes that were also present in the list of DEGs from C to D because if they are drivers of pathogenic foaming, they must be present in the initial stages of the program. To finally get the drivers, this list was crossed with the upstream regulators of homeostatic foaming to ensure they are unique to pathogenic foaming. IPA was then revisited to identify the causal network master regulators of the drivers, which are not necessarily in the dataset, but are channels, complexes, and receptors projected to regulate these genes.

Results

4 drivers and 8 master regulators of the pathogenic foaming program in atherosclerosis were revealed. The 4 drivers are IRF7, ARID3A, ELL2, and MGA. Figure 3 shows the expression of each driver gene's mouse gene equivalent plotted in GSE116240. The visual difference in the gene expression of the pathogenic foaming program compared to the homeostatic foaming program confirms their significance. The 8 master regulators are: receptors

NR1H2, CHRNB2, and NR1H3, complexes P300/PCAF, LUBAC, IL1RL1/IL1RAP, and TP53 TETRAMER:ARID3A, and ion channel TRPC5.

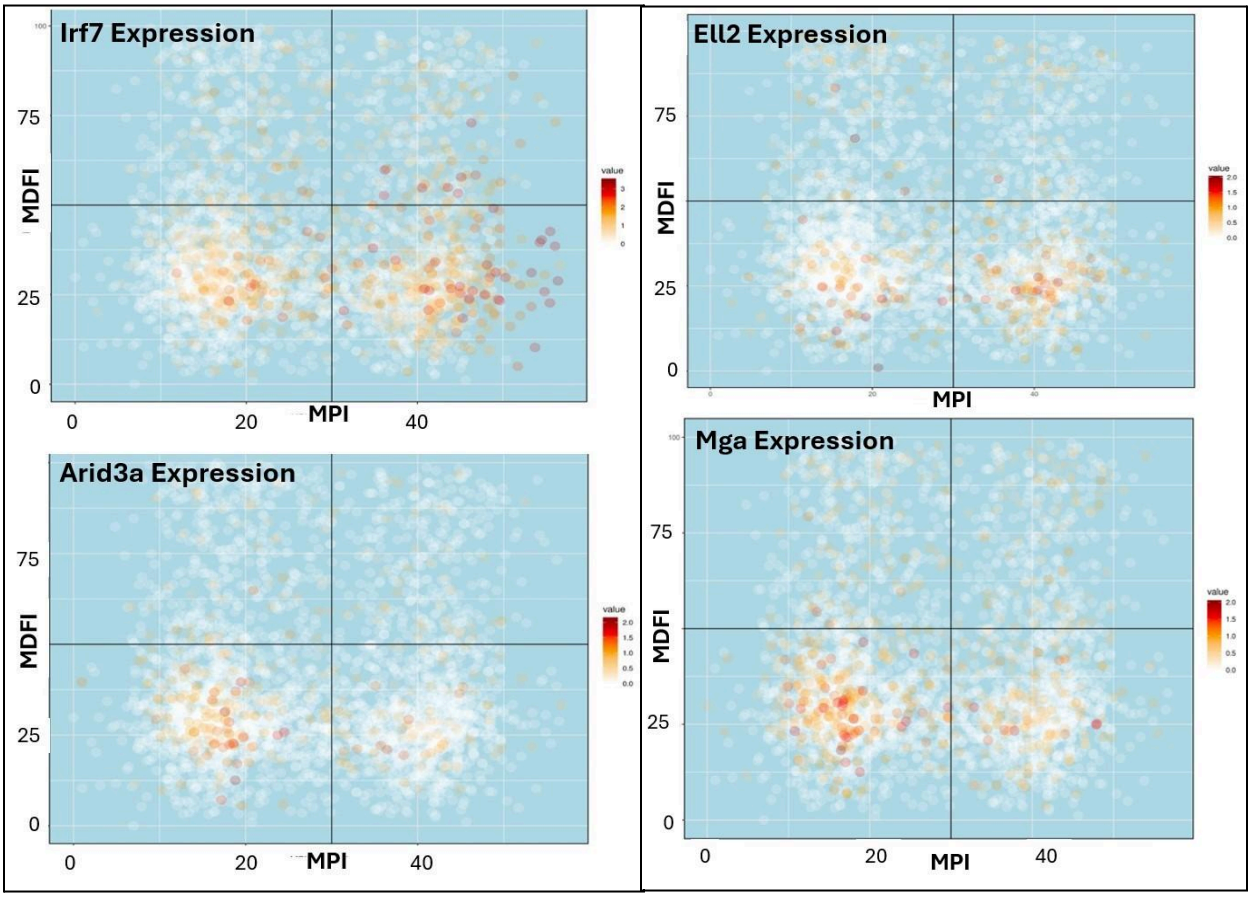


Figure 3 - Expression of identified driver genes in GSE116240

Plots show that drivers are significantly more expressed in quadrants associated with the pathogenic foaming program

To validate these 12 inflammatory factors, the expression of each one was compared across disease conditions in two additional macrophage/foam cell datasets via a Wilcoxon statistical test. First, human RNA-sequencing (RNA-seq) data of whole tissue from donors with and without plaque was accessed from the NIH's Genotype-Tissue Expression project (GTEx). As shown in Figure 4, a significant difference in expression was observed between conditions for

ARID3A ($p = 0.001$), IRF7 (3.89×10^{-12}), and MGA (1.907×10^{-9}). When testing the master regulators, the gene E2F1 was used to represent the TP53 complex because it is a direct part of the complex's pathway, and the SHARPIN gene was used to represent the LUBAC complex as a key component of it. As shown in Figure 4, a significant difference in expression was observed between the conditions for NR1H2 (8.47×10^{-12}), TRPC5 (0.037), NR1H3 (1.75×10^{-12}), EF21(3.98×10^{-5}), EP300 (3.17×10^{-14}), and IL1RL1 (7.39×10^{-9}).

Next, mouse scRNA-seq data from partial carotid artery ligation (PCAL) was accessed from the Zhou Lab (UConn Health, Department of Immunology). The two conditions are from the left and right aorta, with one side being normal blood flow, and the other being low flow from the PCAL, to simulate an atherosclerotic environment. As shown in Figure 5, a significant difference was observed between the conditions for Mga ($p = 0.025$). For the master regulators, a significant difference was observed for Nr1h2 ($p = 0.032$). There was a less significant difference in condition groups for the mouse data, despite visible differences in plots. This is likely a result of the dataset being relatively small and therefore experiencing high levels of dropout, which is when lowly expressed genes are not able to be captured in the sequencing process. Future steps should be to continue to validate results in other scRNA-seq and bulk RNA-seq data. Note that drivers/regulators not confirmed in validation sets are not being ruled out as insignificant now, given the high variability in RNA transcripts.

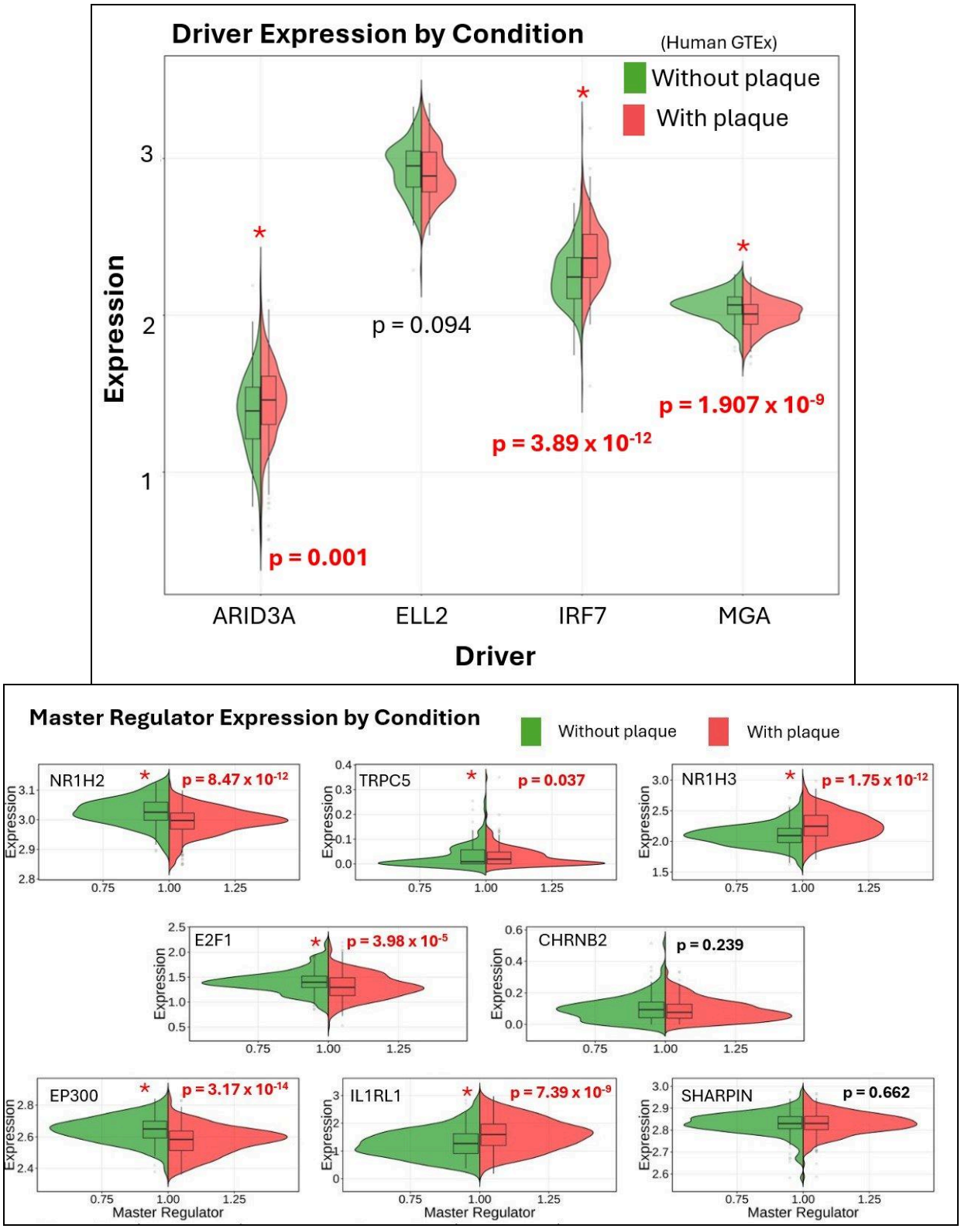
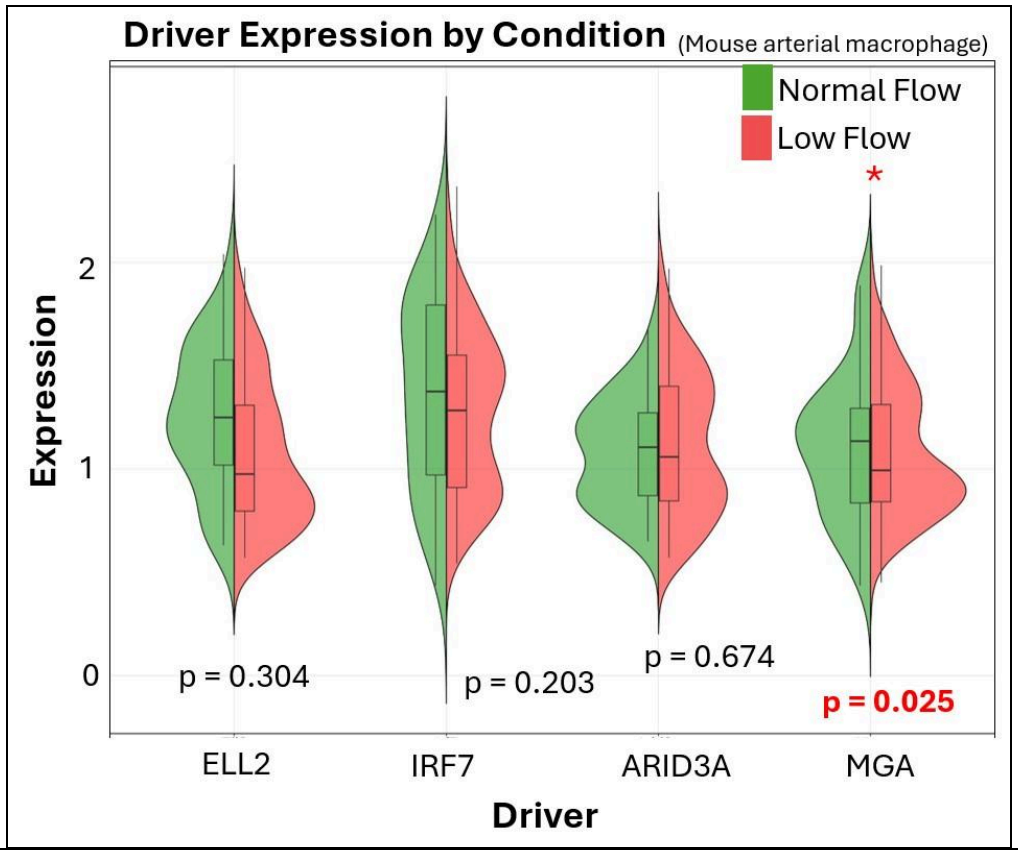


Figure 4 - Validation of drivers and master regulators in human RNA-seq data (GTEx)



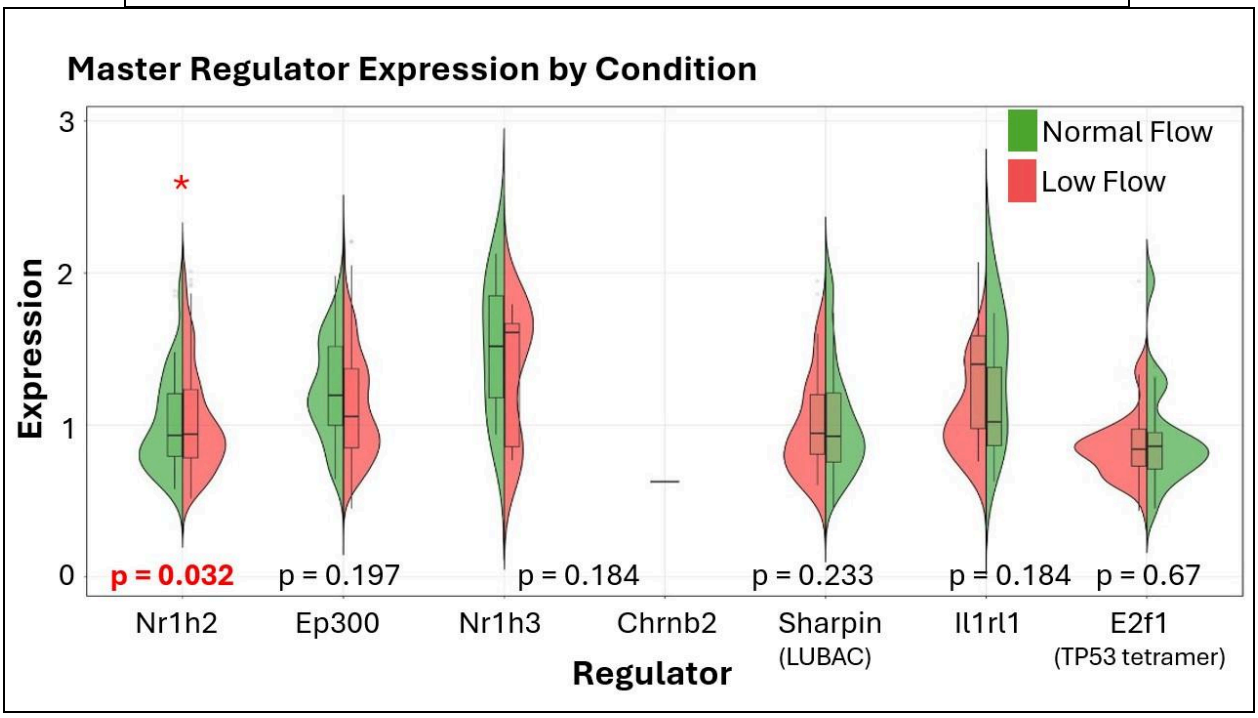


Figure 5 – Validation of drivers and master regulators in arterial mouse macrophage scRNA-seq data

Discussion

There is an urgent need to address the persisting risk of cardiovascular events. The results of this project provide specific targets and a new approach to atherosclerosis treatments. Targeting inflammation is a difficult task; inflammation is a natural and important process, and its exact pathways in atherogenesis are not yet understood. Further, as demonstrated by the existence of a homeostatic foaming program, foaming is also an important process, and targeting it does not necessarily reduce atherosclerosis. The identification of IRF7, ARID3A, ELL2, and MGA as driver genes provides insights into the regulatory pathways involved in inflammatory foam cell progression. Still, these 4 drivers are in the nucleus and hold many responsibilities, making it difficult to target or stop their actions specifically. The identification of NR1H2, CHRNB2, NR1H3, P300/PCAF, LUBAC, IL1RL1/IL1RAP, TP53 TETRAMER:ARID3A, and TRPC5 as regulators of the drivers provide even more insights into the mechanisms of this disease. These 8 master regulators are a wider, more accessible list of potential therapeutic targets. Future studies can test the effects of drugs known to regulate the master regulators of the drivers on plaque formation in atherosclerosis. Such work will move treatments away from only focusing on lipid content, and rather, turn the field toward a more holistic approach to treatment that considers the immunological systems at play.

Acknowledgements

I sincerely thank Mr. and Mrs. Holster for their generous funding of this project through the Holster Scholars program. I also want to express my gratitude to Dr. Beiyan Zhou and Alyssa Matz for their expert mentorship throughout this project, and to the Zhou Lab as a whole for

welcoming me into the lab. Finally, I am grateful to Dr. Vincent Moscardelli and the Honors program at the University of Connecticut for giving me this opportunity to enrich my undergraduate experience.

References

1. *Atherosclerosis.* 2024, https://my.clevelandclinic.org/health/diseases/16753-atherosclerosis-arterial-disease
2. Karlinsey, K., Qu, L., Matz, A. J., & Zhou, B. (2022). A novel strategy to dissect multifaceted macrophage function in human diseases. *Journal of Leukocyte Biology, 112*(6), 1535–1542. 10.1002/jlb.6mr0522-685r
3. Ridker, P. M., Rifai, N., Pfeffer, M. A., Sacks, F. M., Moye, L. A., Goldman, S., Flaker, G. C., Braunwald, E., & Gotto, A. M. (1998). Inflammation, Pravastatin, and the Risk of Coronary Events After Myocardial Infarction in Patients With Average Cholesterol Levels. *Circulation,* https://www.ahajournals.org/doi/10.1161/01.CIR.98.9.839
4. Ridker, P. M., Everett, B. M., Thuren, T., Macfadyen, J. G., Chang, W. H., Ballantyne, C., Fonseca, F., Nicolau, J., Koenig, W., Anker, S. D., Kastelein, J. J. P., Cornel, J. H., Pais, P., Pella, D., Genest, J., Cifkova, R., Lorenzatti, A., Forster, T., Kobalava, Z., . . . Glynn, R. J. (2017). Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *New England Journal of Medicine, 377*(12), 1119. 10.1056/nejmoa1707914
5. Buckley, L. F., & Libby, P. (2024). Colchicine's Role in Cardiovascular Disease Management. *Arteriosclerosis, Thrombosis, and Vascular Biology, 44*(5), 1031. 10.1161/atvbaha.124.319851
6. Li, C., Qu, L., Matz, A. J., Murphy, P. A., Liu, Y., Manichaikul, A. W., Aguiar, D., Rich, S. S., Herrington, D. M., Vu, D., Johnson, W. C., Rotter, J. I., Post, W. S., Vella, A. T., Rodriguez-Oquendo, A., & Zhou, B. (2021). AtheroSpectrum Reveals Novel Macrophage Foam Cell Gene Signatures Associated With Atherosclerotic Cardiovascular Disease Risk. *Circulation, 145*(3), 206–218. 10.1161/circulationaha.121.054285
7. RStudio Team. (2020). RStudio, PBC. Boston, MA:http://www.rstudio.com/
8. Quan, F., Liang, X., Cheng, M., Yang, H., Liu, K., He, S., Sun, S., Deng, M., He, Y., Liu, W., Wang, S., Zhao, S., Deng, L., Hou, X., Zhang, X., & Xiao, Y. (2023). Annotation of cell types (ACT): a convenient web server for cell type annotation. *Genome Medicine, 15*(1)10.1186/s13073-023-01249-5