Systems Genetics Approaches to Compare Mechanisms of Smooth Muscle Cell Plasticity in Quiescent and Proliferative States in Coronary Artery Disease

A Technical Report submitted to the Department of Biomedical Engineering

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Diana Albarracin
Spring, 2022

Technical Project Team Members
Noah Perry, Ph.D. Candidate

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature  Date 05/06/2022
Diana Albarracin

Approved  Date 05/05/2022
Mete Civelek, Department of Biomedical Engineering

Systems Genetics Approaches to Compare Mechanisms of Smooth Muscle Cell Plasticity in Quiescent and Proliferative States in Coronary Artery Disease

Diana Albarracin, Noah Perry Ph.D. Candidate

Abstract

Vascular smooth muscle cells (VSMCs) are key players involved in atherosclerosis, the underlying cause of coronary artery disease (CAD). They can play either beneficial or detrimental roles in lesion pathogenesis, depending on the nature of their phenotypic changes. An in-depth characterization of their gene regulatory networks can help to better understand how their dysfunction may impact disease progression. A network preservation analysis of gene expression in aortic SMCs isolated from 151 multi-ethnic heart transplant donors cultured under quiescent or proliferative conditions was conducted. A total of 86 groups of co-expressed genes (modules) across the two conditions were identified and the focused were the 18 modules that are least preserved between the phenotypic conditions. The majority of the modules, however, were enriched for metabolic pathways including glycolysis related processes. We also created gene regulatory networks enriched for genes in glycolysis and predicted key regulatory genes driving glycolysis dysregulation. Our work provides evidence that dysregulation of VSMC metabolism participates in phenotypic transitioning which may contribute to disease progression and suggests that MPI may play an important role in regulating glycolysis related metabolism in SMCs.

Keywords: Coronary artery disease, atherosclerosis, network analysis, unpreserved pathways

Introduction

Cardiovascular diseases represent 31% of all deaths, making them the leading cause of death globally [1]. Coronary artery disease (CAD) makes up approximately half of these deaths. In addition, the prevalence of CAD is 6.7% in the US for adults who are 20 years of age or older [2]. It is projected that by 2030 CAD prevalence will increase by 16.6%. CAD is long recognized to be heritable, as 40-70% of the disease's risk can be attributed to genetics [3]. Gene sequencing efforts such as genome-wide association studies (GWAS) identified 175 loci associated with risk for CAD [4]. However, most of the loci associated with risk for CAD are associated with unknown mechanisms and are thought to function by regulation of gene expression in the vascular wall where smooth muscle cells (SMCs) play essential roles in the development of atherosclerosis, the underlying cause of CAD. Current therapies for CAD target its risk factors, such as blood lipid levels and blood pressure [5]. The goal of this project is to design a computational method that will aid to identify disease susceptibility mechanisms that function in the vessel wall where the disease develops. This is key to finding new therapeutic approaches. In addition to shining light on disease susceptibility mechanisms for CAD, the unpreserved modules analysis may be more comprehensive than other widely used methods to advance the genetic research in other heritable diseases.

Computational methods such as gene set enrichment analysis (GSEA) yield insights into different diseases by focusing on groups of genes that share commonalities such as chromosomal location, regulation or biological function [6]. Another approach for this type of analysis is differential expression analysis. This approach works by looking at the level at which specific genes are expressed between different conditions. This allows for understanding at which biological processes may be affected between different conditions [7]. Though these approaches have revealed important information about CAD and possible genes and mechanisms associated with it, there are still unknown mechanisms associated with the disease [4], [8].

To this end, this project aims to combine previously used methods to interrogate CAD and mechanisms associated with, specifically by investigating dysregulated gene co-expression networks between healthy and atherogenic SMCs. We will use weighted gene co-expression network analysis (WGCNA), which is a systems biology method for describing the correlation patterns among genes. WGCNA can be used for finding modules of highly correlated genes and relating said modules to each other or an external trait [9]. In addition, module preservation statistics have been shown to be useful for studying differences between the modular structure of networks [10]. To determine whether a pathway of genes is perturbed between the atherogenic and healthy conditions of CAD, modules whose connectivity patterns are not preserved between conditions can be studied. A way to do this is by studying module preservation by looking at the cross-tabulate module membership. However, this

approach often fails to consider preservation of connectivity patterns between nodes. Therefore, a better approach is to use module preservation statistics to determine preservation between modules.

Materials and Methods

Smooth Muscle Cell Culture and Gene Expression

Smooth muscle cell (SMC) gene expression dataset has been described elsewhere [11]. We cultured the SMCs in complete media (containing 5% FBS) until 90% confluence. We then switched to either serum-free media for 24 hours to mimic the quiescent state of SMCs or continued to culture in complete media to mimic the proliferative state of SMCs [17]. Total RNA was extracted using the RNeasy Micro Kit (Qiagen) and the RNase-free DNase Set. RNA integrity scores for all samples, as measured by the Agilent TapeStation, were greater than 9, indicating high-quality RNA preparations. Sequencing libraries were prepared with the Illumina TruSeq Stranded mRNA Library Prep Kit and were sequenced to ∼100 million read depth with 150 bp paired-end reads at the Psomogen sequencing facility. We trimmed the reads with low average Phred scores (<20) using Trim Galore and mapped the reads to the hg38 version of the human reference genome using the STAR Aligner[11]. We quantified gene expression by calculating the transcripts per million (TPM) for each gene using RNA-SeQC[12] based on GENCODE v32 transcript annotations. In addition to protein-coding RNAs, we also measured the non-coding RNA since they have been shown to play significant roles in SMC biology [18]. We considered a gene as expressed if it had more than 6 read counts and 0.1 TPM in at least 20% of the samples. RNAseq data is available from GEO with the accession number GSE193817.

Gene set enrichment analysis

We performed Gene Set Enrichment analysis (GSEA) on all expressed genes shared between the quiescent and proliferative conditions using GSEA software version 4.1 and predefined gene sets from the Molecular Signatures Database version 7.5 [19], [20]. A gene set is a group of genes that shares pathways, functions, chromosomal location, or other features. For the present study, we used the Hallmark gene set, C2 curated gene sets including KEGG and Reactome, and C5 ontology gene sets including Gene Ontology (GO) Biological Process and Molecular Function sets. GSEA ranks all of the genes in the dataset based on mean value differences and calculates gene set significance using an enrichment score defined as the maximum distance from the middle of the ranked list. The enrichment score indicates whether the genes contained in a gene set are clustered towards the beginning or the end of the ranked list.

Differential gene expression and functional enrichment analysis

We included genes with > 6 reads in at least 80% of the samples for both of the conditions for differential expression analysis using DESeq2 [21]. Genes were considered to be differentially expressed between proliferative and quiescent conditions when Padj < 1x10^-3 and log2(fold-change)>0.5. To characterize the functional consequences of gene expression changes associated with proliferative and quiescent conditions, we performed Gene Ontology [22], [23], KEGG [24], [25], [26], Reactome [27], and Hallmark [28] gene set enrichment analysis on differentially expressed genes using the anRichment R package [29].

Weighted Gene Co-expression Network Analysis

A gene module is a cluster of densely interconnected genes in terms of co-expression. We used Iterative Weighted Gene Co-expression Network Analysis (iterativeWGCNA) [30], which uses hierarchical clustering and an adjacency matrix, to identify gene modules. The adjacency matrix is defined as the similarity between the i-th gene and j-th gene based on the absolute value of the Pearson correlation coefficient between the profiles of genes i and j. IterativeWGCNA follows the same principles as WGCNA [31] but re-runs WGCNA iteratively in order to prune poorly fitting genes resulting in more refined modules compared to WGCNA. Genes that are not assigned to any of the modules are designated to the grey module. Because these genes are not co-expressed, we did not consider them in our analyses.

Network preservation analysis

Preservation analysis was performed on modules constructed using iterativeWGCNA to study their changes across conditions. To determine whether a pathway of genes is perturbed between the proliferative and quiescent conditions, we studied modules whose connectivity patterns are not preserved between conditions as demonstrated by their module preservation statistics. For this analysis we used the summary statistic, medianRank, implemented in the WGCNA R package [31], [32] as a composite module preservation statistic. medianRank is a rank-based measure that relies on observed preservation statistics. medianRank is calculated as the mean of medianRank.density and medianRank.connectivity. Density is the mean adjacency (connection strength) across all nodes in the network. Connectivity is the sum of connection strengths with the other network nodes. In order to calculate

medianRank.density and medianRank.connectivity, for each statistic a in the reference network, we ranked modules in the test network based on the observed values obs_a^q. Thus, each module is assigned a rank $rank_a^q$ or each observed statistic. The median density and connectivity ranks are then calculated for each module, q, in the test network. The test and reference networks were then flipped in order to calculate preservation for each condition in the other. A module with a lower medianRank exhibits stronger observed preservation statistics than a module with a higher median rank. We identified the least preserved modules by defining the modules scoring in the bottom 20th percentile of preservation (modules with the highest medianRank score).

Coronary Artery Disease associated gene sets

We used two different curated gene sets based on the 175 genomic loci associated with coronary artery disease (CAD) risk in genome-wide association studies (GWAS) [33]. The CAD Candidate gene set includes 2051 genes representing all genes in the 175 CAD GWAS loci. The CAD Prioritized gene set contains 175 genes predicted to be causal at each genome-wide significant loci based on functional annotation, such as genomic location, biological pathway interpretation, literature reviews, and DEPICT gene prioritization [33], [34].

Pathway enrichment of co-expression modules

To interpret the biological significance of the least preserved co-expression modules, enrichment analysis was performed based on Gene Ontology, KEGG, Reactome, and Hallmark pathways using the anRichment R package. Biological Process and Molecular Function GO terms were focused on for downstream analysis.

Bayesian Network Construction and Key Driver Analysis

Gene expression from genes identified in co-expression modules were provided as input into the Reconstructing Integrative Molecular Bayesian Networks (RIMBANET) algorithm [35]. One thousand Bayesian networks (BNs) were reconstructed using different starting random seeds. Edges that appeared in greater than 30% of the networks were used to define a consensus network. Edges that were involved in loops were then removed from the consensus network. To identify key regulators for a given regulatory network, we performed key driver analysis (KDA) [36] which takes as input a set of genes (G) and a directed gene network (N). KDA first generates a sub network N_G, defined as the set of nodes in N that are no more than h-layers away from the nodes in G. We first computed the size of the h-layer neighborhood (HLN) for each node in the reconstructed BN. For the given network N, μ was defined as the average size of the HLN. A score was added for a specific node if the HLN was greater than $\mu + \sigma(\mu)$. Total key driver scores for each node were then defined as the summation of all scores at each h-layer scaled according to h.

Hypergraph Models

PANTHER version 14 [37] was used to annotate gene ontology terms (GO Term) in BNs. An FDR cutoff of 0.05 was used for enrichment. The HyperG R package was then used to create a hypergraph where each node represents a GO Term and each edge represents a gene or set of genes. The size of the node represents how many genes are in the pathway. Edges and gene labels are color coordinated [38].

Results

Gene expression modules in human smooth muscle cells

We constructed gene co-expression networks from RNA sequencing data of aortic smooth muscle cells (SMCs) isolated from 151 multi-ethnic heart transplant donors cultured in quiescent and proliferative conditions. The overall analysis workflow adopted in this work is summarized in Figure 1.

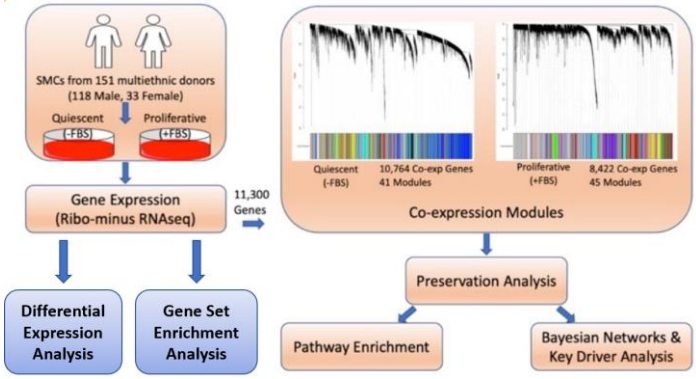

Figure 1. Schematic representation of the overall study design: Smooth muscle cells (SMCs) from the ascending aortas of 151 multiethnic donors were cultured with and without FBS to mimic the quiescent and proliferative phenotypes of SMCs. Cells were harvested and total RNA was extracted. There were 11,300 genes expressed in at least 80% of the samples across both culture conditions. Co-expression modules were created using the expression levels of the 11,300 genes. Network preservation analysis was performed to rank modules based on preservation. To interrogate the modules, pathway enrichment analysis was performed and Bayesian networks were created. Key driver analysis was performed to identify key regulating genes. Differential expression analysis and gene set enrichment analysis was also performed.

After preprocessing and sample outlier detection, 151 samples with gene expression data for 11,330 genes were inputted into the weighted gene co-expression network

analysis (WGCNA) to create gene co-expression modules for both SMC phenotypic conditions. We performed module detection using iterativeWGCNA [39], [40]. To identify modules of co-expressed genes, we searched for genes with similar patterns of connection strengths to other genes or high topological overlap. A soft threshold power of three and six were used for quiescent and proliferative conditions, respectively, to ensure resulting co-expression networks are closer to a scale-free network frequently observed in large scale biological networks [41]. The quiescent condition resulted in 10,764 co-expressed genes segmented into 41 modules, and the proliferative condition resulted in 8,422 co-expressed genes segmented into 45 modules. The modules ranged in size from 34 to 2134 genes. Because co-expression modules are able to capture genes operating within similar biological pathways and functions, deciphering the modules that are context specific could lead to understanding genes and pathways operating in phenotype-specific context.

Preservation analysis of smooth muscle cell modules in distinct phenotypic states

We assessed whether the 41 modules identified in quiescent SMCs were preserved in the 45 modules identified in proliferative SMCs. We utilized statistics that do not depend on a gene's particular module assignment, but rather network properties such as density and connectivity which rely on connection strengths and topology among all genes [10]. Using a composite statistic of preservation from the WGCNA R package, medianRank, we ranked the preservation of each module across phenotypes. A low medianRank score represented a highly preserved module, whereas a high medianRank represented a less preserved module. To identify the genes and biological pathways most likely to be enriched for phenotype-specific functions, we identified the modules scoring in the bottom 20th percentile of preservation. This cutoff denoted the nine least preserved modules in both the quiescent condition and proliferative condition, represented by the modules beneath the red line (Figure 2A,B). Together these 18 modules contain 2,379 unique genes with topological connectivity representing phenotype-specific interactions. Of the 2,379 genes, less than 10% were shared across the conditions. Pathway analysis of the 18 modules revealed pathways representative of differential conditions that were not identified with differential gene expression or gene set enrichment analyses.

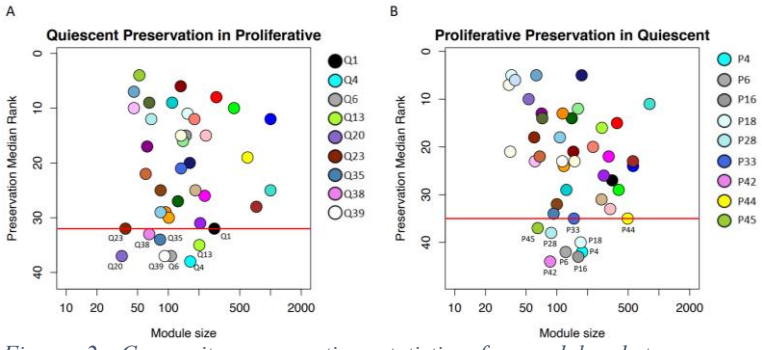

Figure 2. Composite preservation statistics for modules between quiescent and proliferative conditions: The composite statistic, medianRank (y-axis), as a function of the module size. Each point represents a module, labeled by color and a secondary numeric label. Low numbers on the y-axis indicate a high preservation. The red line denotes the bottom 20th percentile of preservation scores. Modules at or below the red line represent the least preserved modules. (A) medianRank scores for the preservation of quiescent modules identified in quiescent SMCs in proliferative modules identified in proliferative SMCs. (B) medianRank scores for the preservation of proliferative modules in quiescent modules. There are nine modules at or below the red line cutoff in each condition.

Metabolic pathway enrichment in unpreserved modules

A module was enriched for NADH regeneration and canonical glycolysis pathways (Q23) (Figure 3). Glycolysis plays an important role in the proliferation of VSMCs [42], [43]. Consequently, proliferative VSMCs demonstrate an increase in glycolytic flux so it is unsurprising that we identified a context specific function of glycolysis [44]. Exploring the glycolytic alterations of VSMCs, however, may provide new insights into the genes involved in the quiescent and proliferative functions of glycolysis. To address the potential rewiring of glycolysis, we compared network topology between the unpreserved Module Q23 in the quiescent condition, and a proliferative module, Module P17, that is also enriched for NADH regeneration and canonical glycolysis.

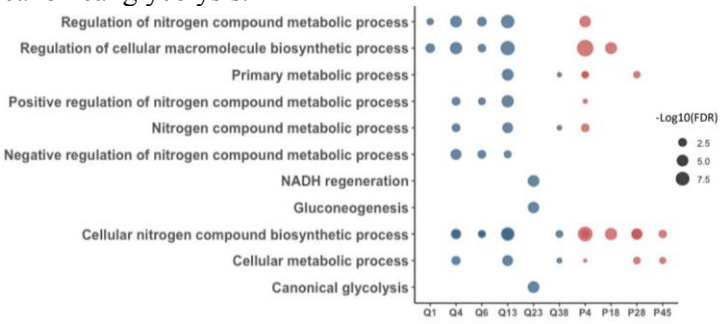

Figure 3. Presence of metabolic processes in the least preserved modules: Gene Ontology enrichment of metabolically related pathways in 10 of the least preserved modules. Each point is scaled according to -log10 (FDR) values. Blue points represent modules from the quiescent condition and red points represent modules from the proliferative condition.

We created BNs for each co-expression module to predict genetic regulatory function and identified KDs to isolate potential genes responsible for driving glycolytic rewiring (Figures 4A, 4B). There were 11 shared genes between the two BNs, 7 unique to the quiescent condition, and 11 unique to the proliferative condition. The BNs shared two CAD candidate genes, ENO2 and SPAG4, with the addition of RAB20 in the proliferative network. RAB20 expression was downregulated in proliferative SMCs (P-value <0.001). Downregulation of RAB20 has been shown to promote glycolysis and contribute to enhanced cell proliferation and motility [45]. KD analysis identified a novel KD gene in the proliferative BN, MPI. GO Term analysis of the genes present in each BN showed that all enriched pathways present in the quiescent condition were preserved in the proliferative condition. However, in the proliferative condition, there were more genes in each shared pathway and the addition of new pathways. Three of the new GO Term pathways in the proliferative BN were represented by the presence of the KD, MPI, suggesting that mannose metabolism could be driving glycolysis rewiring during SMC transition.

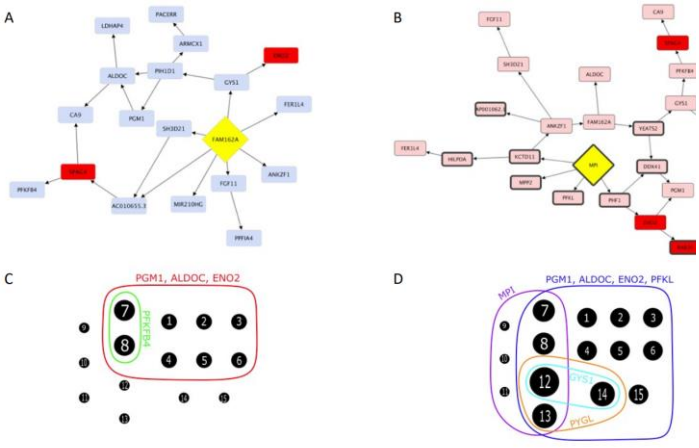

Figure 4. Rewiring of glycolysis metabolic pathway in SMC phenotypic transition:(A) Bayesian networks of genes in the (A) Q23 module and (B) P17 module. Red nodes represent CAD candidate genes and yellow diamonds represent the highest scoring key driver gene. Bold node outlines in (B) represent genes unique to the P17 BN. (C-D) Hypergraph representations of enriched GO Terms (FDR < 0.05) based on genes present in the (C) Q23 and (D) P17 BNs. Each node represents a GO Term. Nodes are scaled according to the number of genes functioning in the GO Term. Edges represent the gene or genes functioning in the surrounding nodes (GO Terms). GO Terms for each node: 1. Glycolytic process 2. ATP generation from ADP 3. ADP metabolic process 4. Purine ribonucleoside diphosphate metabolic process 5. Nucleoside diphosphate phosphorylation 6. Pyruvate metabolic process 7. Hexose metabolic process 8. Monosaccharide metabolic process 9. Mannose-6-phosphate isomerase activity 10. GDP-mannose biosynthetic process 11. GDP-mannose metabolic process 12. Nucleobase containing small molecule metabolic process 13. Carbohydrate metabolic process 14. Ribose phosphate metabolic process 15. Generation of precursor metabolites and energy.

Gene set enrichment analysis (GSEA) and differential expression analysis was performed on this data. The top enriched GO Terms for both methods mainly captured DNA replication processes. These standard methods were unable to detect metabolic dysregulation like the preservation analysis method as seen on Figure 5.

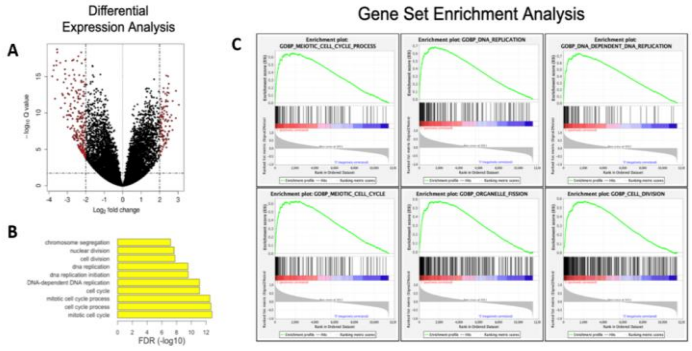

Figure 5. Differential Expression and Gene Set Enrichment Analysis: (A) Volcano plot showing differentially expressed genes between quiescent and proliferative conditions. (B) Top 10 enriched gene ontology (GO) Terms from differentially expressed genes. (C) Top 6 enriched GO Terms from GSEA between quiescent and proliferative conditions.

Discussion

Smooth muscle cells are a major cell type present at all stages of an atherosclerotic plaque. Previously thought to only play a protective role in stabilizing the fibrous cap, lineage-tracing studies have highlighted that SMCs can adopt alternative phenotypes that contribute both positively and negatively to disease progression [46], [47], [48]. Investigating the genetic architecture of quiescent SMCs compared to proliferative SMCs could lead to identifying biological pathways being rewired during phenotypic transformation, and therefore, lead to mechanistic predictions driving SMC plasticity in atherosclerosis, the underlying cause of CAD.

Since it is not possible to study these processes in detail in the arteries of living humans, we cultured VSMCs isolated from an ethnically-diverse population of 151 heart transplant donors in two conditions that are believed to represent the quiescent and proliferative state of the cells. Differential gene expression analysis confirmed the gene expression profiles that are consistent with the phenotypic state of the cells [11]. Because co-expression networks are representative of functionally related genes, a network preservation approach was able to capture dysregulated

pathways whose gene-gene interactions were rewired as a result of phenotypic transition of SMCs

Preservation analysis identified the least preserved gene co-expression modules between quiescent and proliferative SMCs. Unpreserved modules enriched for metabolic functions were present in both our quiescent and proliferative conditions representing genetic rewiring of metabolic pathways contributing to a phenotype specific role of metabolism. Thus, our results support the claims that metabolism of VSMCs are correlated with phenotype switching.

Although glycolytic metabolism in VSMCs is not fully understood, previous reports have identified their potential role in VSMCs and atherosclerosis [12]. We are the first to hypothesize mannose metabolism as a potential mechanism contributing to proliferative VSMCs. Mannose is not a significant energy source in humans but it is required for protein glycosylation [13]. Mannose treatment was shown to attenuate weight gain, improve glucose and lipid homeostasis, and reduce gene expression of inflammatory markers in adipocytes of high fat diet mice [14]. Mannose supplementation in normal mice, however, did not have as pronounced of an effect on weight gain attenuation suggesting that mannose has a greater effect in pathological-like conditions. In addition, plasma levels of mannose have recently been shown to be a biomarker of CAD and a more vulnerable plaque phenotype [15]. It is not clear whether mannose is related to CAD because it marks insulin resistance or because of an intrinsic biological property [16]. As VSMCs respond to vascular injury and transition to a more proliferative, disease-like phenotype, mannose metabolism may be responding to a pathological-like phenomenon and mediating changes in metabolism due to imbalances in energy uptake, thus contributing to disease development through a discrete biological mechanism.

This study demonstrates the power of using network preservation statistics in identifying differences between two biological states. We provide new evidence supporting the role of metabolism as a potential regulator of VSMC plasticity. Further studies need to be conducted to discern whether dysregulated metabolism in VSMCs is a byproduct or a driving mechanism of phenotypic plasticity. Specifically, considering MPI in regulating mannose and glycolysis metabolism in VSMCs.

Limitations

Limitations of this project include lack of a representative dataset of the population in terms of ancestry and sex. In addition, since the project is purely computational the pathways identified as being pathways of interest will have to be validated through further experiments.

References

[1] "Low bone mineral density and coronary artery disease: A systematic review and meta-analysis - ScienceDirect." https://www.sciencedirect.com/science/article/pii/S2352906721001792 (accessed Nov. 09, 2021).
[2] M. Vecsey-Nagy *et al.*, "Association between affective temperaments and severe coronary artery disease," *J. Affect. Disord.*, vol. 295, pp. 914–919, Dec. 2021, doi: 10.1016/j.jad.2021.08.063.
[3] A. V. Khera and S. Kathiresan, "Genetics of coronary artery disease: discovery, biology and clinical translation," *Nat. Rev. Genet.*, vol. 18, no. 6, Art. no. 6, Jun. 2017, doi: 10.1038/nrg.2016.160.
[4] van der Harst Pim and Verweij Niek, "Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease," *Circ. Res.*, vol. 122, no. 3, pp. 433–443, Feb. 2018, doi: 10.1161/CIRCRESAHA.117.312086.
[5] "Medications for Coronary Artery Disease (CAD) | Winchester Hospital." https://www.winchesterhospital.org/health-library/article?id=19085 (accessed Nov. 09, 2021).
[6] A. Subramanian *et al.*, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," *Proc. Natl. Acad. Sci.*, vol. 102, no. 43, pp. 15545–15550, Oct. 2005, doi: 10.1073/pnas.0506580102.
[7] "Differential gene expression (DGE) analysis," *Training-modules*. https://hbctraining.github.io/Training-modules/planning_successful_rnaseq/lessons/sample_level_QC.html (accessed Dec. 16, 2020).
[8] J. Guo *et al.*, "Association of expression of ZNF606 gene from monocytes with the risk of coronary artery disease," *Clin. Biochem.*, vol. 60, pp. 44–51, Sep. 2018, doi: 10.1016/j.clinbiochem.2018.08.005.
[9] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 559, Dec. 2008, doi: 10.1186/1471-2105-9-559.
[10] P. Langfelder, R. Luo, M. C. Oldham, and S. Horvath, "Is my network module preserved and reproducible?," *PLoS Comput. Biol.*, vol. 7, no. 1, p. e1001057, Jan. 2011, doi: 10.1371/journal.pcbi.1001057.

[11] Aherrahrou, R. *et al.* Genetic regulation of human aortic smooth muscle cell gene expression and splicing predict causal coronary artery disease genes. 2022.01.24.477536 (2022) doi:10.1101/2022.01.24.477536.

[12] Shi, J., Yang, Y., Cheng, A., Xu, G. & He, F. Metabolism of vascular smooth muscle cells in vascular diseases. *Am. J. Physiol.-Heart Circ. Physiol.* 319, H613–H631 (2020).

[13] Stanley, P., Taniguchi, N. & Aebi, M. N-Glycans. in *Essentials of Glycobiology* (eds. Varki, A. et al.) (Cold Spring Harbor Laboratory Press, 2015).

. [14] Sharma, V. *et al.* Mannose Alters Gut Microbiome, Prevents Diet-Induced Obesity, and Improves Host Metabolism. *Cell Rep.* 24, 3087–3098 (2018).

[15] Ferrannini, E. *et al.* Mannose as a biomarker of coronary artery disease: Angiographic evidence and clinical significance. *Int. J. Cardiol.* 346, 86–92 (2022).

[16] Mardinoglu, A. *et al.* Plasma Mannose Levels Are Associated with Incident Type 2 Diabetes and Cardiovascular Disease. *Cell Metab.* 26, 281–283 (2017).

[17] Zimmermann, O. *et al.* Serum starvation and growth factor receptor expression in vascular smooth muscle cells. *J. Vasc. Res.* 43, 157–165 (2006).

[18] Leung, A., Stapleton, K. & Natarajan, R. Functional Long Non-coding RNAs in Vascular Smooth Muscle Cells. *Curr. Top. Microbiol. Immunol.* 394, 127–141 (2016).

[19] Subramanian, A. *et al.* Gene set enrichment analysis: 2 A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550 (2005).

[20] Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* 34, 267–273 (2003).

[21] Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 | Genome Biology | Full Text. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8.

[22] Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29 (2000).

[23] The Gene Ontology Consortium *et al.* The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 49, D325–D334 (2021).

[24] Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30 (2000).

[25] Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551 (2021).

[26] Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci. Publ. Protein Soc.* 28, 1947–1951 (2019).

[27] Fabregat, A. *et al.* Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics* 18, 142 (2017).

[28] Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* 1, 417–425 (2015).

[29] Miller, J. A. *et al.* Strategies for aggregating gene expression data: The collapseRows R function. *BMC Bioinformatics* 12, 322 (2011).

[30] Greenfest-Allen, E., Cartailler, J.-P., Magnuson, M. A. & Stoeckert, C. J. iterativeWGCNA: iterative refinement to improve module detection from WGCNA co-expression networks. 234062 (2017) doi:10.1101/234062.

[31] WGCNA: an R package for weighted correlation network analysis | BMC Bioinformatics | Full Text. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559.

[32] Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is My Network Module Preserved and Reproducible? *PLOS Comput. Biol.* 7, e1001057 (2011).

2 [33] Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–1177 (2020).

2 [34] Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).

[35] Zhu, J. *et al.* Integrating Large-Scale Functional Genomic Data to Dissect the Complexity of Yeast Regulatory Networks. *Nat. Genet.* 40, 854–861 (2008).

[36] Zhang, B. & Zhu, J. Identification of Key Causal Regulators in Gene Networks. 4 (2013).

[37] Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47, D419–D426 (2019).

[38] Feng, S. *et al.* Hypergraph models of biological networks to identify genes critical to pathogenic viral response. *BMC Bioinformatics* 22, 287 (2021).

[39] Greenfest-Allen, E., Cartailler, J.-P., Magnuson, M. A. & Stoeckert, C. J. iterativeWGCNA: iterative refinement to improve module detection from WGCNA co-expression networks. 234062 (2017) doi:10.1101/234062.

[40] WGCNA: an R package for weighted correlation network analysis | BMC Bioinformatics | Full Text. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559.

[41] Barabási, A.-L. Scale-Free Networks: A Decade and Beyond. *Science* 325, 412–413 (2009).
[42] Hall, J. L., Chatham, J. C., Eldar-Finkelman, H. & Gibbons, G. H. Upregulation of glucose metabolism during intimal lesion formation is coupled to the inhibition of vascular smooth muscle cell apoptosis. Role of GSK3beta. *Diabetes* 50, 1171–1179 (2001).
[43] GLUT1-induced cFLIP expression promotes proliferation and prevents apoptosis in vascular smooth muscle cells | American Journal of Physiology-Cell Physiology. https://journals.physiology.org/doi/full/10.1152/ajpcell.00213.2009.
[44] Salabei, J. K. & Hill, B. G. Mitochondrial fission induced by platelet-derived growth factor regulates vascular smooth muscle cell bioenergetics and cell proliferation. *Redox Biol.* **1**, 542–551 (2013).
[45] Liu, B. H. M. *et al.* TPI1-reduced extracellular vesicles mediated by Rab20 downregulation promotes aerobic glycolysis to drive hepatocarcinogenesis. *J. Extracell. Vesicles* 10, e12135 (2021).
[46] Feil, S. *et al.* Transdifferentiation of vascular smooth muscle cells to macrophage-like cells during atherogenesis. *Circ. Res.* 115, 662–667 (2014).
[47] Hoofnagle, M. H., Thomas, J. A., Wamhoff, B. R. & Owens, G. K. Origin of neointimal smooth muscle: we've come full circle. *Arterioscler. Thromb. Vasc. Biol.* 26, 2579–2581 (2006).
[48] Bentzon, J. F. & Majesky, M. W. Lineage tracking of origin and fate of smooth muscle cells in atherosclerosis. *Cardiovasc. Res.* 114, 492–500 (2018).