Analysis of single-cell RNA-Seq reveals dynamic changes during macrophage state transition in atherosclerosis mouse model

Ahmed Safwat Abouhashem

Follow this and additional works at: https://fount.aucegypt.edu/etds

Recommended Citation

APA Citation

MLA Citation

This Master's Thesis is brought to you for free and open access by the Student Research at AUC Knowledge Fountain. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AUC Knowledge Fountain. For more information, please contact thesisadmin@aucegypt.edu.
Analysis of Single-Cell RNA-Seq Reveals Dynamic Changes During Macrophage State Transition in Atherosclerosis Mouse Model

A Thesis Submitted to the Biotechnology Program In Partial Fulfilment of the Requirements for The Degree of Masters of Science in Biotechnology

By: Ahmed Safwat Elsayed Abouhashem

Bachelor of Medicine, Bachelor of Surgery (MBBCh), Zagazig University

Under the Supervision of Dr. Hassan Azzazy

Professor and Chairman

Department of Chemistry

The American University in Cairo

December 2019
Dedication

I dedicate this research to my family for always providing me the time and facilities to focus on studying and research.
Acknowledgements

I would like to acknowledge the American University in Cairo, the Department of Biology for providing me the resources and materials to improve my knowledge. Also, I would like to acknowledge Indiana University, IUPUI, for providing a variety of wonderful teams in different disciplines. In particular, I would like to acknowledge Department of Surgery in Indiana University for their experience and guidance which were vital to the project. Also, I would like to acknowledge Dr Azzazy (AUC) for changing my mindset regarding how to identify problems and how to find existing solutions that fit and Dr Sen and Kanhaiya (Indiana University) for their continuous support from always appreciating my efforts to provide me with all the facilities to translate my ideas into experimental and clinical research through different lab members and multiple collaboration with other astonishing teams in Indiana University. I would like to acknowledge Dr Sarath Janga (Indiana University) for his guidance regarding interpreting the results. I would like to acknowledge Razan Masad.
Abstract

**Background:** Atherosclerosis is an arterial inflammation that causes ischemic heart disease, which is the first leading cause of death worldwide. Macrophages play major roles during disease development by having pro-inflammatory and anti-inflammatory functions. Lack of effective treatment is mainly due to incomplete understanding of the molecular mechanisms underlying disease progression and regression.

**Materials and methods:** The transcripts of the macrophages from two aortic samples from atherosclerotic region during disease progression and regression were analyzed using previously published dataset ([GEO Accession GSE123587](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123587)). Pre-processing, clustering of cells and identification of unique markers for each cluster were done using [Seurat](https://satiglab.org/seurat/) package implemented in R programming language. [Monocle](https://www.bioconductor.org/packages/vignettes/seurat/v1.0/brownseurat-vignette.pdf) package was used to order the cells in pseudotime and to detect the key molecules that changed dramatically during comparison between distinct macrophages states (pro-inflammatory and anti-inflammatory). [Ingenuity Pathway Analysis (IPA)](https://www.ingenuity.com/) software was used to analyze the pathways activity across macrophage states along the trajectory and to retrieve the transcriptional regulatory network between the genes determining the final states. Prediction of the miRNAs that might be involved in the disease progression was performed using [TargetScan](https://www.targetscan.org/) and [GSEA](https://www.gsea-msigdb.org/gsea/index.jsp) (Gene Set Enrichment Analysis). [Cytoscape](https://cytoscape.org/) application was used to visualize the regulatory network between the differentially regulated genes across macrophages states.

**Results:** Clustering analysis of macrophages revealed their presence in distinct 11 states. In addition, Two states were found to be dominant in the progression group macrophages, and one state was found to be dominant in the
regression group macrophages. Moreover, trajectory analysis showed a bifurcation point near the end of the trajectory, where macrophages fates were destined to be either pro-inflammatory or anti-inflammatory. Macrophages unique to the disease progression branch were found to activate STAT cascade, induce acute inflammatory response and upregulate inflammatory cytokines, denoting M1 polarization. In contrast, regression-branch specific macrophages were found to activate cholesterol efflux pathways and upregulate anti-inflammatory cytokines such as TSLP and CCL24. The transcription regulatory network between differentially regulated genes in both branches revealed changes in the transcriptional dynamics acquired during macrophage states transition. STAT1 (Signal transducer and activator of transcription 1) and IRF7 (Interferon Regulatory Factor 7) were found to be upregulated in the progression branch to maintain an inflammatory module resulting in production of distinct inflammatory cytokines. On the other hand, MAFB (MAF BZIP Transcription Factor B) and IGF1 (Insulin-like growth factor 1) were found to be upregulated in the regression branch to interrupt the inflammatory module at different levels. In addition, 10 miRNAs were predicted to be unregulated in progression-branch specific macrophages such as miR-344, miR-346 and miR-485.

**Conclusion:** Inflammatory sites in atherosclerosis lesions contain both pro-inflammatory and anti-inflammatory macrophages. Each subset of macrophage activates unique transcriptional program. Certain transcription factors and growth factors have potential to alter the whole transcriptional regulatory network, thereby shifting the macrophages from inflammatory to anti-inflammatory state. Understanding how macrophage state transition occurs from inflammatory to anti-inflammatory state will be a key step to better understanding and treating atherosclerosis.
# Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>VIII</td>
</tr>
<tr>
<td>List of Tables</td>
<td>IX</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>X</td>
</tr>
<tr>
<td><strong>CHAPTER 1: Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Atherosclerosis</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Epidemiology and Statistics</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Aetiology and Pathophysiology</td>
<td>1</td>
</tr>
<tr>
<td>1.1.3 Immune Cells in Atherosclerotic Plaques</td>
<td>3</td>
</tr>
<tr>
<td>1.1.4 Integrin Signalling in Atherosclerotic Plaque Development</td>
<td>6</td>
</tr>
<tr>
<td>1.1.5 Cytokines and Their Role in Atherosclerosis Development</td>
<td>6</td>
</tr>
<tr>
<td>1.1.6 Treatment and Management</td>
<td>7</td>
</tr>
<tr>
<td>1.2 Reverse Cholesterol Transport in Macrophages</td>
<td>8</td>
</tr>
<tr>
<td>1.3 Macrophages plasticity and polarization</td>
<td>9</td>
</tr>
<tr>
<td>1.4 miRNA in Atherosclerosis</td>
<td>10</td>
</tr>
<tr>
<td>1.5 Single Cell RNA Sequencing</td>
<td>11</td>
</tr>
<tr>
<td>1.6 Pseudotime</td>
<td>12</td>
</tr>
<tr>
<td><strong>CHAPTER 2: Hypothesis and Objectives</strong></td>
<td>13</td>
</tr>
<tr>
<td>2.1 Hypothesis</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Objectives</td>
<td>13</td>
</tr>
<tr>
<td><strong>CHAPTER 3: Materials and Methods</strong></td>
<td>14</td>
</tr>
<tr>
<td>3.1 Data Source</td>
<td>14</td>
</tr>
<tr>
<td>3.2 Single Cell RNA Sequencing Counts Matrix Pre-processing</td>
<td>15</td>
</tr>
<tr>
<td>3.2.1 Data Integration</td>
<td>15</td>
</tr>
<tr>
<td>3.2.2 Filtering Low Quality Cells</td>
<td>16</td>
</tr>
</tbody>
</table>
3.2.3 Dimensional Reduction

3.3 Clustering of Single Cells

3.4 Identification of Unique Markers for Each Cluster

3.5 Constructing Single Cell Trajectories

3.6 Differential Expression Analysis of the Terminal Branches

3.7 Pathway Activity Analysis

3.8 Transcription Regulatory Network Analysis

3.9 miRNAs Prediction

CHAPTER 4: Results

4.1 Pre-processing of Single Cell RNA Sequencing Counts Matrix

4.2 Clustering and Visualization

4.3 Identification of Clusters Markers

4.4 Constructing and Analysis of Single Cell Trajectories

4.5 Pathway Activity Analysis

4.6 Transcriptional Regulatory Network Analysis

4.7 miRNA Prediction

CHAPTER 5: Discussion

5.1 Distinct States of Macrophages During Disease Progression and Regression

5.2 Extracellular Matrix Remodelling

5.3 Dendritic Cell Maturation in the Progression Branch

5.4 Cytokines in Progression and Regression

5.5 Analysis of the Transcription Regulatory Network

5.5.1 Role of STAT1 and IRF7 in Maintaining The Inflammatory Module
List of Figures

Figure 1: Atherosclerosis resulting from chronic arterial wall injury 2
Figure 2: Distinct phenotypes of macrophages during atherosclerotic plaque development 4
Figure 3: Macrophages phenotypes M1 and M2 roles during inflammation and resolution 5
Figure 4: The analysis workflow 21
Figure 5: Quality check for single cells before and after filtering 23
Figure 6: Principal component analysis of single cells 24
Figure 7: Clustering of single cells and compositional analysis of the clusters 26
Figure 8: Heatmaps of the identified markers for clusters 4, 6 and 7 28
Figure 9: Single-cell trajectory analysis and dynamic changes in gene expression of branches-specific macrophages 30
Figure 10: Pathways and predicted miRNAs in macrophages across the pseudotime 33
Figure 11: Transcriptional regulatory network showing inflammatory module in progression-specific macrophages 36
Figure 12: Expression of genes responsible for the inflammatory and anti-inflammatory response 38
Figure 13: Expression level of downstream targets of STAT1 and IRF7 39
Figure 14: Altered downstream targets of MAFB gene 40
List of Tables

Table 1: Number of cells assigned to each cluster in both samples  27
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEAM</td>
<td>Branched Expression Analysis Modelling</td>
</tr>
<tr>
<td>CCA</td>
<td>Canonical Correlation Analysis</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular Matrix</td>
</tr>
<tr>
<td>DCs</td>
<td>Dendritic cells</td>
</tr>
<tr>
<td>FN</td>
<td>Fibronectin</td>
</tr>
<tr>
<td>GEO</td>
<td>Gene Expression Omnibus</td>
</tr>
<tr>
<td>GSEA</td>
<td>Gene Set Enrichment Analysis</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IPA</td>
<td>Ingenuity Pathway Analysis</td>
</tr>
<tr>
<td>LDL</td>
<td>Low density lipoprotein</td>
</tr>
<tr>
<td>LXR</td>
<td>Liver X receptors</td>
</tr>
<tr>
<td>KNN</td>
<td>K-nearest neighbour</td>
</tr>
<tr>
<td>miRNA</td>
<td>Micro-RNA</td>
</tr>
<tr>
<td>PC</td>
<td>principal component</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>SNN</td>
<td>Shared nearest neighbour</td>
</tr>
<tr>
<td>TSLP</td>
<td>Thymic stromal lymphopoietin</td>
</tr>
<tr>
<td>t-SNE</td>
<td>t-Distributed Stochastic Neighbour Embedding</td>
</tr>
<tr>
<td>UMAP</td>
<td>Uniform Manifold Approximation and Projection</td>
</tr>
</tbody>
</table>
CHAPTER 1: Introduction

1.1 Atherosclerosis

1.1.1 Epidemiology and Statistics

Atherosclerosis is a chronic inflammation of the arteries, which is initiated by low density lipoprotein (LDL) accumulation in defined regions in the arteries where the blood flow is disturbed and non-laminar. Atherosclerosis results in stroke, heart attacks and peripheral arterial disease (Dichgans M et al., 2019; Mohd Nor NS et al., 2019; Paul S et al., 2019). It is the major cause of cardiovascular diseases resulting in ischemic heart disease and ischemic stroke, which are the first and fifth causes of death worldwide; respectively (Ala-Korpela M et al., 2019; Watson M et al., 2019; Whelton SP et al., 2019; Carmona FD et al., 2019).

1.1.2 Aetiology and Pathophysiology

Multiple risk factors are involved in the development of atherosclerosis. The most common factors include elevated blood LDL (hypercholesterolaemia), diabetes mellitus, hypertension, smoking, sex as a male, old age (more than 45 years for males and more than 55 years for females) and family history of the disease (Reiss AB et al., 2019; Shafi S et al., 2019; Doodnauth SA et al., 2019). Atherosclerosis development started with endothelial cells dysfunction and infiltration of immune cells in the lesion micro-environment, resulting from arterial wall injury (figure 1).
Figure 1. Atherosclerosis resulting from chronic arterial wall injury. Chronic inflammation of endothelial cells resulting from atherosclerotic risk factors induces endothelial cells dysfunction and alters the arterial wall micro-environment. Macrophages recruitment to the lesion site results in foam cell formation after engulfing the deposited lipids in the arterial wall intima. Reproduced with permission (Rafieian-Kopaei et al., 2014). Image Source Reference; Copyright; see appendix.

Atherosclerosis is a continuous process of arterial wall lesions caused by lipid retention in the intimal part of the artery, which results in aggravation of the chronic inflammation. Following lipid retention, the intima develops fibrous plaques, resulting in disease complications such as stenosis or remote vessel occlusion due to rupture of the fibrotic plaque. Mechanisms of plaque rupture include shear stress injury (Gertz S et al., 1990), transient collapse of the stenotic lesion (Binns R et al., 1989) mechanical shear injury (Vito R et al., 1990) and vasa vasorum rupture (Barger A et al., 1991).
1.1.3 Immune Cells in Atherosclerotic Plaques

Atherosclerotic plaques contain several immune cells, including dendritic cells, macrophages and platelets (Libby et al., 2013). Dendritic cells have been found to increase dramatically as the disease progresses in the atherosclerotic plaques (Galkina E et al., 2006). They produce distinct chemokines in the lesion area resulting in recruiting lymphocytes and monocytes to the inflammation micro-environment (Huang DR, et al., 2001).

Macrophages in the plaque sites originate either from recruitment of blood circulating monocytes or from proliferation of tissue resident macrophages (Ginhoux and Jung, 2014). Macrophages have been found to play dual roles in the development of atherosclerosis as both of the phenotypes (M1 and M2) were found in atherosclerotic plaques (figure 2, 3) (Bouhlel et al., 2007, De Paoli et al., 2014).
Figure 2. Distinct phenotypes of macrophages during atherosclerotic plaque development. During early stages of the disease development, activation of endothelial cells occur. Following endothelial cells activation, monocytes are recruited into the lesion micro-environment and differentiated into macrophages. Distinct states of macrophages are found in the lesion site playing either inflammatory (M1 phenotype) or anti-inflammatory (M2 phenotype) roles. (Reproduced with permission (Sai Yang et al., 2019). Image Source Reference; Copyright; see appendix.)
Figure 3. Macrophages phenotypes M1 and M2 roles during inflammation and resolution. M1 phenotype is present during early stages of inflammation having roles in inducing inflammatory response. On the other hand, M2 phenotype is present during late stages of inflammation inducing resolution of the inflammatory response. Reproduced with permission (Amitava Das et al., 2015). Image Source Reference; Copyright; see appendix.

Pro-inflammatory (M1) macrophages have been found during plaque growth causing inflammation. They ingest lipid particles and form foam cells where their cytoplasm is filled with lipid droplets. Those foam cells secrete various signalling molecules and pro-inflammatory cytokines to further recruit monocytes, cause more lipid retention and result in extracellular matrix remodelling (Libby, 2002; Orekhov et al., 2014). On the other hand, anti-inflammatory macrophages (M2) have been found in the lesion area causing plaque regression and inflammation resolution (Nathan and Ding, 2010). Furthermore, M2 phenotype cells are more resistant to form foam cells and they contain fewer lipid droplets, indicating distinct metabolic activity between both phenotypes (Chinetti-Gbaguidi et al., 2011).
1.1.4 Integrin Signalling in Atherosclerotic Plaque Development

Distinct subsets of integrins have been found to be expressed by leukocytes and white blood cells to mediate their interaction with the endothelial surface markers (Hynes RO., 2002). These integrins have roles in the initial targeting leukocytes to the inflammation site, facilitating their migration along endothelial cells into the neointimal matrix (Ley K, et al., 2007) and regulation of their function upon transendothelial migration into the inflammation micro-environment (Libby P et al., 2015). Initial weak bonds are formed between integrins expressed by leukocytes and selectins present on the endothelial cell surface. Notably, Atherosclerotic plaque formation reduces dramatically following selectins deletion (Collins RG et al., 2000; Dong ZM et al., 2000). Following integrins activation, leukocytes make rolling movements resulting in firm cell adhesion and chemokines arrest on the endothelial surface (Shamri R et al., 2005). The resulting increase in chemokines induces inflammation and aid in further recruiting of more leukocytes (Boring L et al., 1998).

1.1.5 Cytokines and Their Role in Atherosclerosis Development

Cytokines are low-molecular weight proteins that are clustered into several groups as interleukins, tumour necrosis factors, interferons (IFN), transforming growth factors, chemokines and colony-stimulating factors. They are expressed in the atherosclerotic plaques by almost all the cells in the lesion micro-environment having either pro-inflammatory or anti-inflammatory role (Kleemann R et al., 2008; McLaren J.E et al., 2011; Ait-Oufella H et al., 2011). Cytokines and chemokines are potent signaling molecules mediating intercellular communication (Dinarello, 2007). During initial stages of the disease development, cytokines have a role in modulating endothelial cells permeability resulting in activating and opening up gaps between cells (Pober J.S. et al., 2007). The activated endothelial cells release chemokines
(chemoattractant cytokines) causing further recruitment of circulating leukocytes especially monocytes and T cells (McLaren J.E. et al., 2011). Chemoattractant chemokines modulate diverse intra-signalling pathways upon interaction with cell surface receptors involving activation of G proteins (Koenen R et al., 2010; Weber C et al., 2011).

1.1.5 Treatment and Management

Management of atherosclerosis is mainly by treating the risk factors as elevated blood LDL, diabetes and others. Patients are advised to eat healthy diet containing low saturated fat and less salts. In addition, they are encouraged to exercise 90 to 150 minutes per day. Furthermore, drugs such as statins are used to lower LDL cholesterol and reduce complications. Statins are a group of cholesterol-lowering compounds which act by inhibiting Hydroxymethylglutaryl Coenzyme A Reductase (HMG-CoA). This inhibition reduces the rate by which the enzyme is able to produce mevalonate, which is a major molecule in the cascade that produces cholesterol (Endo A, 1992). Although statins can lower blood cholesterol level, they have side effects such as muscle pain and abnormal blood levels of liver enzymes (Abd TT et al., 2011). In severe stenosis, surgical intervention is recommended such as coronary artery bypass grafting (Perez-Martinez P et al., 2019; Spannella F et al., 2019; Esper RJ et al., 2019; Arnett DK et al., 2019). The complicated roles of macrophages in atherosclerotic plaque micro-environment is poorly understood. Thus, better understanding of the molecular alterations caused by macrophages in the atherosclerotic plaque site is needed for improving the disease treatment.
1.2 Reverse Cholesterol Transport in Macrophages

Reverse cholesterol transport (RCT) was first described by Glomset (Glomset JA, 1968) as the process of returning the excess circulatory cholesterol to liver. RCT relationship with atherosclerosis development was first hypothesized as an imbalance which occurs between arterial cholesterol deposition and removal (Ross R & Glomset JA, 1973). In atherosclerotic plaques, macrophages are the primary cell type that are overloaded with cholesterol (Tall AR, 2001). Macrophages take up the dead cells that contain much cholesterol, lipoproteins and cellular debris (Rader DJ, 2005). The unesterified cholesterol has toxicity effect on macrophages leading to apoptosis (abas I, 2004). Thus, esterification of cholesterol to cholesteryl ester (CE) is the first line of defense against cholesterol toxicity in macrophages by the enzyme ACAT1 (acyl:coenzyme A cholesterol O-acyltransferase-1) (Chang TY, 2001). Cholesterol ester is stored in the cytoplasm and its accumulation results in foam cell formation.

The second line of defense against cholesterol toxicity is the cholesterol efflux. ABCA1 is an active cholesterol efflux pathway and its knock out in mice results in development of accelerated atherosclerosis (Aiello RJ, 2002). However, alternative pathways by which macrophages can be able to induce cholesterol efflux exist. ABCG1 was identified to promote cholesterol efflux in the form of HDL particles and its knock out results in intracellular lipid accumulation (Kennedy MA, 2005). Both ABCA1 and ABCG1 are regulated by the nuclear receptors liver X receptor LXR-α and LXR-β (Tontonoz P, 2003). Excess intracellular cholesterol induces formation of oxysterol which are the ligands for liver X receptors (Repa JJ, 2002). Thus, excess cholesterol upregulates major pathways for cholesterol efflux through ABCA1 and ABCG1 to protect the cells from the toxicity. Moreover, mice deficient in LXR-α/-β have been found to have lipid accumulation within
macrophages and tissues and have advanced atherosclerosis (Tangirala RK, 2002). Furthermore, peroxisome proliferator-activated receptor PPAR-α and PPAR-γ agonists were found to induce cholesterol efflux from macrophages through upregulation of LXRα (Chinetti G, 2001; Li AC, 2004; Chawla A, 2001). Following unesterified cholesterol efflux from macrophages, it associates with plasma HDL. Then, lecithin:cholesterol acyltransferase (LCAT) catalyzes its esterification to cholesteryl ester (CE) (Fielding CJ, 1995). As CE is hydrophobic, it moves to the lipoprotein particle core allowing mature HDL formation. LCAT-deficient mice were found to have very low level of HDL (Rader DJ, 1994) and its overexpression resulted in high HDL level and reduction in atherosclerosis (Hoeg JM, 1996).

### 1.3 Macrophages plasticity and polarization

Macrophages are plastic cells which can switch their state from one phenotype to another (Mantovani et al., 2004; Sica & Mantovani, 2012). Polarization of macrophages is a process in which macrophages change their transcriptional program in response to stimuli and signals from the surrounding micro-environment (Sica & Mantovani, 2012). Several classes of macrophages have been described based on their metabolic activity, surface markers expression and their released molecules. The two major sub-populations with opposing functions are M1 phenotype (inflammatory) and M2 phenotype (anti-inflammatory). The phenomenon of macrophages can change their state from M1 to M2 and the opposite is referred to term “macrophage polarization” (Cassetta, Cassol & Poli, 2011; Chittezhath et al., 2012).

M1 macrophages are induced by Th1 cytokines (IFN-γ and TNF-α) and they produce pro-inflammatory cytokines such as TNFα, IL1α, IL1β and IL6 (Biswas et al., 2012; Bashir et al., 2016). On the other hand, M2 macrophages are induced by
Th2 cytokines (Wang et al., 2014) and have anti-inflammatory roles. However, the M1/M2 phenotype does not reflect the distinct subsets of macrophages. They can be further subdivided into different sub-types (Chistiakov et al., 2015). Recent advances in single cell RNA-sequencing enables identification of phenotypes not previously identified (Hannah Van Hove et al., 2019; Jian-Da Lin et al., 2019).

1.4 miRNA in Atherosclerosis

miRNAs are small non-coding RNAs, ~22 nucleotides in length. Biogenesis of miRNAs starts with processing of RNA polymerase II/III transcripts mostly from introns and few from exons and intergenic regions (Ha M et al., 2014; de Rie D et al., 2017). miRNAs biogenesis is classified into canonical and non-canonical pathways. The canonical biogenesis pathway is the dominant pathway in which the pri-miRNAs are transcribed and processed into pre-miRNAs. Processing of pre-miRNAs are performed by the microprocessor complex (DGCR8, RNA binding protein DiGeorge Syndrome Critical Region 8 and Dorsha, a ribonuclease III enzyme) (Denli AM et al., 2004). DGCR8 recognizes motifs within the pre-miRNA (Alarcón CR et al. 2015) and Drosha cleaves the pri-miRNA resulting in formation of a 2 nt 3′ overhang on pre-miRNA (Han J et al., 2004). Then, the pre-miRNA is exported to the cytoplasm by exportin 5 (XPO5)/RanGTP complex and further processed by the RNase III endonuclease Dicer resulting in formation of the mature miRNA (Okada C et al., 2009). In addition, multiple non-canonical miRNA biogenesis pathways have been discovered. They can be classified into Drosha/DGCR8-independent and Dicer-independent pathways.

miRNAs recognize and bind to the 3′ UTR of the target mRNAs inducing translational repression and mRNA deadenylation and decapping (Huntzinger E et al., 2011). They have roles in both production and clearance of lipoproteins. miR-122 is
the first identified miRNA to have a role in lipoprotein metabolism (Esau C et al., 2006). Furthermore, overexpression of miR-30 in mice resulted in hyperlipidaemia and atherosclerosis development (Soh J et al., 2013). Other miRNAs have been identified to alter cholesterol efflux by targeting ABCA1 including miR-33 (Gerin I et al., 2010), miR-758 (Ramirez CM et al., 2011), miR-144 (De Aguiar Vallim TQ et al., 2013). Thus, miRNAs have potential roles in atherosclerosis progression and regression.

1.5 Single Cell RNA Sequencing

Cells express different transcriptional programs while being in different states even in similar cell types (Huang S et al., 2009; Li L et al., 2010; Shalek A K et al., 2014). The transient states in-between stable states are difficult to be detected using traditional experimental methods such as bulk RNA sequencing. Fortunately, single cell RNA Sequencing allows detection of cells in those transition states without the needs of their purification. It enables studying the heterogeneity among similar cellular populations (Montoro DT et al., 2018; Plasschaert LW et al., 2018).

Multiple steps are required to generate single-cell data from a biological sample and different protocols can be used. Generally, the steps incorporate dissociation of single cells, library constructions and sequencing (Vieth B et al., 2017; Macosko EZ et al., 2015; Rostom R et al., 2017). Droplet-based methods such as 10x Genomics (which was used in the original study) start with isolating each cell from the sample and capturing it in a microfluidic droplet. Although multiple cells or cellular debris can be captured within the same droplet, computational methods can be applied to overcome this issue by detecting and excluding those cells. Each droplet contains the necessary enzymes and chemicals to digest the cell membrane and perform library construction by reverse transcription of mRNAs generated from the
single cell. Additionally, each droplet contains a barcode of 16 nucleotide long which labels all the mRNAs from this single cell. After library construction, all the cDNAs from each single cell are labelled with a unique barcode and the libraries are pooled together for sequencing.

1.6 Pseudotime

Pseudotime is an abstract unit of measurement as it measures how much progress for each individual cell has been made through the process of state transition between cells. For the macrophages in atherosclerotic plaque, they were found to be in different states (Bouhlel et al., 2007; De Paoli et al., 2014). Monocle package in R uses reversed graph embedding algorithm to learn the sequence of change in gene expression each cell must go through across dynamic biological processes in single cells experiments (Qi Mao et al., 2016; Xiaojie Qiu et al., 2017). Thus, it enables constructing a pseudo-temporal path to order the cells based on the gradual continuous change of their transcriptomes (Trapnell C et al., 2014; Haghverdi L et al., 2016).
CHAPTER 2: Hypothesis and Objectives

2.1 Hypothesis

Macrophages play a vital role in atherosclerosis during progression and regression phases of the disease. They are able to dramatically change their states between pro-inflammatory (M1) and anti-inflammatory (M2) phenotypes. This results in secretion of distinct set of cytokines and activation of different metabolic programs, which affect their ability to digest the lipid particles in the inflammatory micro-environment.

Therefore, it is hypothesized that using single cell RNA sequencing help in identifying the molecular signature of the macrophages during disease progression and regression, the change in pathways activity during macrophages transition across different states and the key molecules which drive the transition decision for the macrophages to be either in an inflammatory (M1) or an anti-inflammatory (M2) state.

2.2 Objectives

1. To cluster macrophages from each sample into distinct states (11 states).

2. To detect the macrophages states unique to disease progression (2 states) or regression (1 state).

3. To order the cells in pseudotime to detect the bifurcation points where macrophages go through different fates.

4. To calculate the pathways activity in different fates across the pseudotime.

5. To identify the key genes that drive the macrophages to either state.

6. To predict the miRNAs that may be involved in states transition.
CHAPTER 3: Materials and Methods

3.1 Data source

The data was obtained from the GEO (http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE123587. The data was produced by Jian-Da Lin et al, a research group in USA in 2019 (Jian-Da Lin et al., 2019). Their paper is entitled “Single-cell analysis of fatted-mapped macrophages reveals heterogeneity, including stem-like properties, during atherosclerosis progression and regression”. The paper was published in JCI Insight Journal, Volume 8, Issue 8, 2019. The original study aimed to address the differentiation of CX3CR1+ CCR2+ cells into macrophages in atherosclerotic plaques during progression and regression of the disease.

The authors generated a chimeras of LDLr-/- (LDL receptor) mice and placed all the eight weeks old mice onto a Western diet (contains high fat content) for 18 weeks. Then, the mice were divided into two groups: a progression group which continued on Western diet and a regression group which was switched to a chow diet (less fat) and injected with apolipoprotein B (ApoB) anti-sense oligonucleotide intraperitoneal (50mg/kg) twice per week to lower the atherogenic lipoproteins. After 2 weeks, they sorted the macrophages from the aortic arches of progression and regression groups. This enabled excluding cells from lymphoid lineage such as T cells, B cells, eosinophils, neutrophils and natural killer cells.

Then, the sorted cells were processed to single-cell RNA sequencing. Finally, by combining four samples from each group with total number of 10,000 to 12,000 cells for each group, cells were loaded on 10x Genomics instrument and libraries were prepared as described (GX et al., 2017). Then, the authors used CellRanger Single Cell Software Suite (version 1.3) for processing the raw reads and generating
3.2 Single Cell RNA-Seq Counts Matrix Pre-processing

The generated count matrix from the two samples was used to perform quality check and further downstream analysis. Seurat package in R (v3.1.0) was used to process the counts matrix, barcodes and genes for each sample (Stuart T et al., 2018; Butler et al., 2018).

3.2.1 Data Integration

The data from the two samples was integrated to enable the identification of common cellular states and to allow comparative analysis between them as described in Butler et al., 2018. Briefly, integration of the two samples required multiple steps starting from data normalization to identification and integration of anchors between cells from the two samples. For normalization, log 10-transformation of expression values for each gene was performed, using 10,000 molecules for each cell as a size factor. Then, standardization of expression value of each gene across all the cells was performed (linear transformation) to shift the expression of each gene, so that the mean is 0 and the variance is 1 across all the cells. This gives equal weight, so that no domination for the highly expressed genes in the downstream analysis.

The top 2000 highly variable genes in both samples were chosen to perform canonical correlational analysis (CCA) and dimensionality reduction to detect cells from the 2 samples with common biological states. K-nearest neighbours (KNNs) were identified and each pair of cells from the two samples represent an anchor (mutual nearest neighbours). Further filtering of anchors was performed by giving a score to each anchor using shared nearest neighbour (SNNs) graphs between the two
samples. Finally, small number of anchors surrounding each cell were considered for identifying the integration anchors. The identified anchors and the weights matrix was used to integrate the two samples as described in Haghverdi L et al., 2018.

3.2.2 Filtering low quality cells

As suggested before, there is a positive correlation between expression level of mitochondrial encoded genes and cell death (Detmer SA et al., 2007; Galluzzi L et al., 2012, Silvia Márquez-Jurado et al., 2018). Thus, cells that exhibit mitochondrial contamination (more than 5% of the reads are mitochondrial encoded genes) were filtered. Also, cells with very high or very low total number of gene counts (more than 7000 or less than 2500) were filtered as high counts may represent cell doublets or multiplets and low counts may represent empty droplets or represent premature cell rupture leading to mRNA leakage.

Moreover, cells identified as non-macrophages were excluded from the analysis. Identification of each cell type was performed using SingleR package implemented in R (Aran D et al., 2019). The algorithm computes the similarity between the dataset and a reference dataset for unbiased recognition of cell types. The reference dataset chosen was ImmGen (Immunological Genome Project), which includes 830 microarrays classified into 20 main cell types and 253 subtypes. Briefly, Spearman coefficient was calculated for each cell expression with each of the reference datasets samples. Then, correlation analysis was performed followed by aggregation of multiple correlation coefficients per cell types using 80% percentile of the correlation values. Next, the correlation analysis was rerun using only the top cell lines resulted from the previous step and the variable genes between those cell types, while removing the lowest cell type. Finally, the previous step was repeated until two cell types remained and the top one was assigned to the cell.
3.2.3 Dimensionality Reduction

Principal component analysis (PCA) was performed using the previously identified 2000 highly variable genes in both samples. Each principal component represents information about a correlated gene set. To further determine how many principal components to be used for dimensionality reduction, Jackstraw procedure was used (Chung NC et al., 2015). Jackstraw procedure is a resampling test, in which a subset of the data is randomly permuted and PCA is rerun, allowing for construction of a “null distribution” of genes scores. This method allows identification of the significant PCs that have low p-values. In addition, a heuristic method based on the percentage of variance represented by each gene was used to generate elbow plot. Consequently, the first 24 principal components were chosen to reduce the data dimensions for downstream analysis to enable using the highly variable genes instead of using all the genes as features.

3.3 Clustering of Single Cells

A graph-based clustering approach was used for clustering the cells. Cells were embedded in a graph structure (KNN graph) based on Euclidean distance in PCA space. Then, edge weights were refined between any pair of cells based on Jaccard similarity (the shared overlap between the cells local neighbourhoods). Edge refinement was performed using the first 24 PCs as previously determined to be the dimensionality of the data. After refinement edge weights, edges were drawn between cells having similar gene expression patterns. Louvain algorithm was applied as a modularity optimization technique to group cells together. Finally, Uniform Manifold Approximation and Projection (UMAP) was used to visualize the cells using the reduced dimensions (Becht et al., 2018; McInnes et al., 2018).
3.4 Identification of Unique Markers for Each Cluster

To identify unique markers for each cluster, Seurat was used to perform differential expression analysis by comparing each cluster with the rest of other clusters using Wilcoxon Rank-Sum Test. Log2 fold change cut off value of ± 0.5 and an adjusted p value less than 0.05 were used to detect the differentially expressed genes within each cluster.

3.5 Constructing Single Cell Trajectories

Monocle package in R (v2.10.1) uses reversed graph embedding algorithms to order the cells in pseudotime and construct the trajectory (Trapnell C et al., 2014). The significant markers from all the clusters with log2 fold change ± 0.5 were used to order the cells (437 genes). Briefly, the ordering algorithm reduces the data dimensionality, learns the smooth manifold that generates the data and finally places every cell at the proper position in the trajectory through unsupervised method. In addition, if there are multiple outcomes, it reconstructs a branched tree representing different cellular fates (Qi Mao et al., 2015).

3.6 Differential Expression Analysis of the Terminal Branches

Furthermore, Monocle enables detection of branch points along the trajectory. The significant branch-dependent genes are the genes that explicit different expression dynamics along each branch. Branched expression analysis modelling (BEAM) algorithm implemented in Monocle, was used to find all the genes that differ between the two terminal branches after the bifurcation point. Then, hierarchical clustering of the resulting top 457 genes with (q value less than 1e10-4) was performed for further downstream analysis. Finally, EnrichR
(https://amp.pharm.mssm.edu/Enrichr/) was used to enrich the identified clusters with the significant gene ontology biological processes (Maxim V et al., 2016).

3.7 Pathway Activity Analysis

Pathway activity analysis was performed using Ingenuity Pathway Analysis (IPA) software (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). The algorithm considers the role of each gene within the pathway, so that each pathway has a pattern to be considered fully activated. If the gene is consistent with its pattern specific to pathway \( i \), score of 1 is assigned to the gene in pathway \( i \), and if it is inconsistent, a score of -1 is assigned for this gene in pathway \( i \). Finally, for determining the activity of pathway \( i \), Z score was calculated for each pathway in each pseudotime point using the following equation, where \( z \) is the activation score:

\[
Z = \frac{X}{\sigma_x} = \frac{\sum_i x_i}{\sqrt{N}} = \frac{N_+ - N_-}{\sqrt{N}}
\]

\( X_i \) is the score of a gene in the pathway. \( N_+ \) and \( N_- \) are the number of the consistent and inconsistent genes with the pathway pattern; respectively. \( N \) is the total number of genes in pathway \( i \). Then, the 457 genes differentially expressed between the 2 branches were used with 20 pseudotime points representing the whole pseudotime. The gene was considered upregulated or downregulated using cut off ± 0.5.
3.8 Transcription Regulatory Network Analysis

The genes that were identified to be differentially expressed between the two different fates of macrophages using BEAM algorithm (457 genes) were used to construct the transcriptional regulatory network. First, IPA was used to retrieve the experimentally validated relationships between the genes. Then, causal analysis was performed using IPA software to detect the subnetworks consistent with the results as previously described (Andreas et al., 2014). Finally, Cytoscape application was applied to visualize the resulting network using the values of the last pseudotime points from both branches (Shannon P et al., 2003).

3.9 miRNAs Prediction

TargetScan (http://www.targetscan.org/) was used to predict the miRNAs that may play a role in regulating the expression level of the differentially regulated genes in progression versus regression branch. The Predicted targets of all known mouse miRNAs were downloaded based on their alignment with the 3’ untranslated regions of each transcript (Rosenbloom KR et al., 2013; Agarwal V et al., 2015). Then gene set enrichment analysis (GSEA) was performed using the gene expression data at the end of the progression branch (http://software.broadinstitute.org/gsea/index.jsp) (Subramanian et al., 2005).
CHAPTER 4: RESULTS

The authors in the original study performed scRNA-seq using 10x Genomics technology and performed the alignment using CellRanger software (Jian-Da Lin et al., 2019). The goal of the current study is to start from performing the quality control and clustering of the cells to trajectory inference, pathway activity analysis and miRNA prediction (Figure 4).

Figure 4. The analysis workflow. Sequencing of single cells were performed using 10x Genomics technology. CellRanger was used for mapping of reads to the mouse genome. Seurat package in R was used for pre-processing of the resulting count matrix, clustering and identification of unique markers for each cluster. Then, Monocle in R was used for trajectory inference and analysis of the bifurcation point. The resulting differentially expressed genes were used for different downstream analysis applications. EnrichR was used for GO biological processed enrichment of the resulting clusters. Ingenuity Pathway Analysis (IPA) software was used to analyse pathway activity over pseudotime and for retrieving the transcriptional regulatory network. TargetScan was used along with Gene Set Enrichment Analysis (GSEA) for miRNA prediction. Cytoscape was used for network visualization.
4.1 Pre-processing of Single Cell RNA Sequencing Counts Matrix

Data from a total number of 3157 cells for progression group and 2198 cells for regression group were analysed. Using SingleR package, most of the cells were identified as macrophages, while few number of cells were assigned to dendritic cells, B cells, monocytes, stem cells and stromal cells. 1565 cells during atherosclerosis progression and 1421 cells during atherosclerosis regression remained after excluding cells with mitochondrial contamination, very high or low total counts and non-macrophages cells (figure 5; figure 7-A & B; section 3.2.2 in methods).
Figure 5. Quality check for single cells before and after filtering. (A) Violin plots showing abundance of cells in each sample regarding the number of detected genes (features), total number of gene counts per cell and the percentage of mitochondrial genes expression (left: before filtering and right: after filtering). Each dot represents one cell. (B & C) Scatter plots before and after filtering; respectively showing number of gene counts against percentage of mitochondrial genes (left) and against total number of detected genes per cell (right).
Following log10-normalization and scaling of the gene counts, PCA was performed using the top 2000 variable genes from both samples. Jackstraw procedure was utilized to assign p value for each principal component as shown in figure 6-A. Moreover, elbow plot was generated using a heuristic method based on the percentage of variance represented by each gene as in figure 6-B. Analysis of the principal components suggest that the majority of the signal was captured in PCs from 1 to 24.

![Figure 6](image_url)

**Figure 6. Principal component analysis of single cells.** (A) Jackstraw plot showing the first 50 principal components from the PCA analysis with their corresponding p-values. (B) Elbow plot showing the standard deviation of the first 50 PCs.
4.2 Clustering and Visualization

A graph-based clustering approach was performed to cluster the cells using the first 24 principal components. Then, UMAP algorithm was used as a non-linear dimensionality reduction technique to visualize the data as shown in figure 7-C. Figure 7-D represents the percentage of the cells assigned to each cluster in respect to each condition (progression and regression groups). Clusters with more than 100 cells and represent more than 70% of their cells from either group were considered to be condition-specific clusters. Cluster four consists of 71% of its cells from the regression group, while 82% and 83% of clusters six and seven; respectively arose from the progression group (Figure 7-D, table 1).
Figure 7. Clustering of single cells and compositional analysis of the clusters. (A) t-distributed stochastic neighbor embedding (t-SNE) projection of the raw data before filtering (5355 cells) color-coded using a scale of gray to violet to plot the percentage of mitochondrial genes (right), total number of gene counts per cell (left) and number of unique genes per cell (middle). Each cell is represented as a dot. (B) t-SNE plot of all cells showing the cells which passed the quality control (green) and the excluded cells (black) from the downstream analysis. (C) Uniform manifold approximation and projection (UMAP) projection of the filtered cells showing 11 clusters in each sample. (D) Compositional analysis of each cluster showing the percentage of the cells from each group per cluster.
Table 1. Number of cells assigned to each cluster in both samples

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Number of cells</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Progression</td>
<td>Regression</td>
</tr>
<tr>
<td>Cluster 0</td>
<td>264</td>
<td>417</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>321</td>
<td>278</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>271</td>
<td>185</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>191</td>
<td>133</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>70</td>
<td>177</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>72</td>
<td>120</td>
</tr>
<tr>
<td>Cluster 6</td>
<td>157</td>
<td>34</td>
</tr>
<tr>
<td>Cluster 7</td>
<td>122</td>
<td>25</td>
</tr>
<tr>
<td>Cluster 8</td>
<td>56</td>
<td>38</td>
</tr>
<tr>
<td>Cluster 9</td>
<td>32</td>
<td>6</td>
</tr>
<tr>
<td>Cluster 10</td>
<td>3</td>
<td>13</td>
</tr>
</tbody>
</table>

4.3 Identification of Clusters Markers

Seurat was used to compare each cluster with the rest of other clusters using Wilcoxon Rank-Sum Test. Log2 fold change cut off value of ± 0.5 and adjusted p value less than 0.05 were chosen to detect unique markers for each cluster, resulting in a total number 437 genes for all the 11 clusters. Figure 8 represents the identified markers from clusters four, six and seven.
Figure 8. Heatmaps of the identified markers for clusters four, six and seven. (A) Comparison of average expression of selected markers between cluster four (regression-specific) and clusters six and seven (progression-specific). Markers are classified into cytokines, transcription factors, enzymes, peptidases, transmembrane proteins, transporters, phosphatases and kinases. High expression is represented by red color, while low expression is represented by green color.
4.4 Constructing and Analysis of Single Cell Trajectories

Monocle was used to order the cells in pseudotime and to construct the trajectory using the previously identified markers (figure 9-A). Branched expression analysis modelling (BEAM) algorithm was performed to find all the genes that differ between the two branches after the bifurcation point. Then, hierarchical clustering of the top differentially expressed genes resulted in nine clusters with distinct patterns across the pseudotime (Figure 9-B). Figure 9-C represents the gene ontology biological processes enrichment for the nine identified groups.
Figure 9. Single-cell trajectory analysis and dynamic changes in gene expression of branches-specific macrophages. (A) Cells along the trajectory are divided into distinct stable states and ordered according to the pseudotime development along the trajectory. At the branching point (highlighted with a black dot), cells divide to go through two fates. Cells are colored based on their position across the pseudotime (above) or based on the sample (below). (B) Heatmap of the top 457 differentially expressed genes between the two branches represents the transcriptional dynamics across the two branches. Genes (rows) are clustered and cells (columns) are ordered according to the pseudotime development. The middle part of the heatmap shows the pre-branch points and the outermost left and right parts show the regression and the progression groups; respectively. (C) The patterns of each cluster from the heatmap along with enrichment of the GO biological processes. All the enrichment terms have p-value < 0.05 and the star denotes corrected p-value < 0.25.
4.5 Pathway Activity Analysis

IPA software was used to compute pathway activity analysis for both branches. Figure 10-A represents a heatmap of the z scores for each pathway across the pseudotime in both branches.
Figure 10. Pathways and predicted miRNAs in macrophages across the pseudotime (A) Heatmap showing each pathway activity across the pseudotime for both the progression and regression branch-specific macrophages. The middle points of the heatmap represent the pre-branching pseudotime points, while the left and right outermost points represent the progression and regression-specific macrophages; respectively. (B) Gene set enrichment analysis plots for the top significant enriched miRNAs in the progression group. (C) The predicted upregulated miRNAs and their target genes. Colors represent the upregulated (red) and downregulated (green) genes along the trajectory in the progression branch-specific macrophages.
4.6 Transcriptional Regulatory Network Analysis

The transcriptional regulatory network was retrieved from the IPA database considering only the experimentally approved relationships between the differentially expressed genes in both branches. Causal analysis were performed using the software to detect the consistent subnetworks. Figure 11 B and C represent the identified network with gene values from the last pseudotime point from the progression and regression branches; respectively.

4.7 miRNA Prediction

**Gene set enrichment analysis (GSEA)** was performed for differentially altered genes in both branches. The gene expression values from the last pseudotime point in the progression branch was used to perform the enrichment. Ten miRNAs were enriched with p value less than 0.05 and bonferroni corrected p value less than 0.1 as shown in figure 10-B.
Figure 11. Transcriptional regulatory network showing inflammatory module in progression-specific macrophages (A) UMAP plotting color-coded using a scale of green to red to black to represent the gene expression level within each cluster (left). The relationship between some genes of interest (right). The transcription regulatory network in the last pseudotime point of progression group (B) and regression group (C). High expression is represented by larger node size and red color, while lower expression is represented by smaller size and blue color.
Figure 12. Expression of genes responsible for the inflammatory and anti-inflammatory response (A) Trajectory analysis of cells within clusters four, six and seven. (B) Violin plots showing expression level of genes involved in maintaining the inflammatory module in the main network. (C) Violin plots showing expression level of genes involved in interrupting the inflammatory module.
Figure 13. Expression level of downstream targets of STAT1 and IRF7 (A) Violin plots showing expression level of downstream targets of IRF7 and STAT1 in progression-specific clusters six and seven and in regression-specific cluster four. (B) Subnetwork from the main transcriptional regulatory network in progression-branch specific macrophages (left) and in regression-specific branch macrophages (right). High expression is represented by larger node size and red color, while lower expression is represented by smaller size and blue color.
Figure 14. Altered downstream targets of MAFB gene MAFB and its downstream targets in regression-specific macrophages (A) and progression-specific macrophages (B). High expression is represented by larger node size and red color, while lower expression is represented by smaller size and blue color.
CHAPTER 5: DISCUSSION

5.1 Distinct States of Macrophages During Disease Progression and Regression

Previous studies to evaluate the macrophage population in atherosclerosis demonstrated that symptomatic plaques contain macrophages of M1 inflammatory phenotype (Cho KY et al., 2013). In contrast, M2 anti-inflammatory phenotype was found to be enriched in asymptomatic lesions indicating the opposing role of macrophages in disease progression and resolution (Monica de Gaetano et al., 2016). Each phenotype stimulates distinct transcriptional program and differs regarding its metabolic activities. For instance, inflammatory phenotype expresses surface markers and secretes specific pro-inflammatory cytokines transmitting the inflammatory signals to neighbour cells in the plaque micro-environment. Also, their transcriptional state results in inefficient lipid metabolism enabling aggravating the inflammation and foam cell formation, while the M2 phenotype expresses certain genes enabling dampening the inflammation and prevention of intracellular lipid retention.

Clustering of macrophages from both samples resulted in identification of 11 clusters. Although, macrophages during disease progression and regression shared common states, condition-specific clusters were identified. Cluster four was found to be more abundant in the regression group (71%), while clusters six and seven were abundant in the progression group, 82% and 83%; respectively, (figure 7-C, table 1). The two clusters (six and seven) arose mainly during atherosclerosis progression indicating their role in the inflammatory process. Moreover, trajectory inference revealed two distinct fates of macrophages, where cluster four acquired a path differs from that of clusters six and seven (figure 12-A). Furthermore, analysis of the differentially regulated genes and pathways from those unique clusters revealed new insights into the regulatory mechanisms underlying macrophages states during
disease progression and regression. For instance, the upregulated genes along the trajectory towards the progression branch-specific macrophages were enriched for biological processes that enhance the disease progress (figure 9-C). They are enriched for cholesterol storage, foam cell differentiation, STAT cascade, INF signalling and positive regulation of acute inflammatory response. On the contrary, regression branch-specific upregulated genes were enriched for bleb assembly and cholesterol efflux indicating the existence of distinct transcriptional program and metabolic activity for each group of macrophages.

5.2 Extracellular Matrix Remodelling

Pathways activity analysis revealed increase in integrin and actin cytoskeletal signalling activity in the progression branch-specific macrophages while decrease in their activity in the regression branch-specific macrophages as shown in figure 10-A. Integrins are involved in the disease initiation and progression by mediating the interaction with endothelial cell surface receptors (Hynes RO, 2002). Although, the plaque extracellular matrix contains lipids and fibronectins (FN), the source of the fibronectins which are one of the earliest proteins deposited in the plaque site is not clear. Notably, deletion of FN alleles from mice significantly reduced the plaque size and prevent formation of the fibrous cap which results in the disease complications. (Ina Rohwedder et al., 2012).

Interestingly, there is inter-cluster communication as shown in the two unique clusters for the progression group (cluster six and seven). progression-specific clusters produce distinct kinds of inflammatory cytokines (figure 8) indicating their cooperation to boost the inflammatory process. For instance, cluster six produces the cytokine CCL2 and cluster seven expresses the G-protein coupled receptor CCR2, the binding partner for CCL2 (figure 11-A). CCL2-CCR2 interaction stimulates signal
transduction through CCR2 receptor resulting in overexpression of FN1 in cluster seven and fibronectin deposition, which is one of the earliest signs during atherosclerosis development. This finding indicates that a small subset of macrophages is a source of the deposited fibronectin in the plaque microenvironment. On the other hand, the regression-specific macrophages (cluster four) downregulates CCR2, so that no signal transduction can be stimulated even in the presence of CCL2 in the lesion site. Consequently, identification of the possible mechanisms of modulating ECM components by macrophages could help in better understanding the disease progression.

5.3 Dendritic Cell Maturation in the Progression Branch

Dendritic cells (DCs) during initial stages of atherosclerosis have been found to be accumulated dramatically in the plaques (Galkina E et al., 2006). Pathways activity analysis reveals that dendritic cell maturation signaling is upregulated in the progression branch and not in the regression branch (figure 10-A). In fact, DCs have been found to be able to accumulate lipids resulting in disease progression (Paulson KE et al., 2010) and produce distinct pro-inflammatory cytokines such as CCL2 and CCL4 to further recruit monocytes in the lesion area. (Huang DR et al., 2001; Braunersreuther V et al., 2007). Furthermore, CCR2, was found to be positively correlated with DCs level in atherosclerotic plaque (Combadiere C et al., 2003; Combadiere C et al., 2008; Liu P, 2008) which is higher in progression branch as previously mentioned. However, the regression branch has low expression of the transmembrane receptor CD74, resulting in inhibiting DCs motility in mice (Faure-Andre G et al., 2008). Besides, mice lacking CD74 have marked reduction of the atherosclerotic plaque development (Sun J et al., 2010). This investigation offers
insight into a potential role of macrophages on activating and suppressing of dendritic cells during atherosclerosis development.

5.4 Cytokines in Progression and Regression

Cytokines in atherosclerotic plaques have both pro-inflammatory and anti-inflammatory roles (McLaren J E, 2011; Ait-Oufella H, 2011). The results revealed upregulation of distinct set of cytokines in each branch. The progression branch shows upregulation of CXCL1, CXCL2, CXCL10, CCL3, CCL4, CCL5, CCL7, CCL9, CCL12, IL1B, IL6, IL1RN and IFNB1, while the regression branch shows upregulation of TSLP, PF4 and CCL24 and downregulation of the progression-specific cytokines (figure 8).

During initial stage of the disease where capturing and rolling of monocytes occurs, CCL5 and CXCL1 interact with proteoglycans and P-selectins on endothelial cells causing immobilization (Ley K et al., 2007; Moore K J et al., 2013; Soehnlein O et al., 2013). Both genes (CCL5 and CXCL1) have higher RNA expression level in the progression branch. Additionally, CCL5 deposition in the plaques promotes monocyte arrest (Von Hundelshausen P et al., 2001) and knockout of Y-box binding-protein 1 which regulates CCL5 expression level results in reduction of neointimal formation in ApoE −/− mice models (Krohn R et al., 2007). CXCL10, CCL3, CCL7, IFNB1 and IL1-Beta, which are higher in the progression branch, have been found to be involved in atherosclerosis progression and their downregulation reduces the pro-inflammatory genes and atherosclerosis plaque burden (Kirii H et al., 2003; Heller E A, 2006; Kennedy A et al., 2012; de Jager S C et al., 2013; Alexander M R et al., 2012, Denes A et al., 2012; Ting-Ting Chang et al., 2016). Moreover, IL-6 injection in ApoE −/− mice increases plaque size and pro-inflammatory cytokines, while its inhibition enhances atherosclerosis via reducing plaque size, decreasing activation of
endothelial cells and reducing monocytes recruitment (Huber S A et al., 1999, Madan M et al., 2008).

Furthermore, CCL2 was found to be higher in the progression branch. In addition to its role in recruiting leukocytes to the inflammation site, it represses the expression of the molecules responsible for cholesterol efflux from the cells resulting in impairment of reverse cholesterol transport mechanism (Weber C et al., 2011; Huang C X et al., 2013). Additionally, Pro-inflammatory cytokines cause inhibition of LXR (Liver X receptors) signaling disrupting cholesterol efflux and resulting in foam cell formation. (Pascual-García M et al., 2013). LXR signalling was found to have less activity in progression branch and more activity in regression branch from the pathway activity analysis (figure 10-A).

On the other hand, among the three cytokines found to be upregulated in the regression branch, thymic stromal lymphopoietin (TSLP) was found to attenuate atherosclerosis development (Kunwu Tu et al., 2013) and to promote macrophages polarization towards M2 phenotype (Debin et al., 2019). Additionally, it further upregulates CD63 tetraspanin (Wang B et al., 2013) which was also found to be secreted in M2 macrophages exosomes (Peiming Zheneg et al., 2018). The second cytokine, CCL24, also has been shown to be expressed in M2 phenotypes (Martinez FO et al., 2006). The functionality of the third cytokine (PF4) found in the regression group awaits further investigation. Taken together, the results indicate the production of anti-inflammatory cytokines from the macrophages in the regression branch driving the process of atherosclerosis resolution.
5.5 Analysis of the Transcription Regulatory Network

5.5.1 Role of STAT1 and IRF7 in Maintaining the Inflammatory Module

The progression branch-specific macrophages activate a transcriptional program resulted in producing distinct kinds of cytokines and other molecules aggravating the inflammatory response in the atherosclerotic plaque lesion site as shown in figure 11-B. Together, STAT1 and IRF7 transcription factors upregulate multiple pro-inflammatory cytokines such as CXCL10 and CCL12 (Vogel SN et al., 2003; Miyagawa F et al., 2016) and downregulate the cytokine TSLP which is upregulated in the regression-specific macrophages. Moreover, they upregulate the expression of three peptidases (PSMB8, PSMB9 and USP18), which play a part in the immunoproteasome machinery system in immune cells (figure 13-A, B) (Barton LF et al., 2002; Goubau D et al., 2009). Notably, immunoproteasome is required for protein processing to present antigenic peptides on MHC class I molecules (Kloetzel PM, 2004) and it was found to be involved in autoimmune diseases pathogenesis (Gutcher I et al., 2007). In fact, peptidases upregulation in the progression branch-specific macrophages sheds light towards their involvement in the continuous inflammatory process during atherosclerosis development.

Interestingly, STAT1 and IRF7 also upregulates interferon induced proteins and IFIT family proteins (Interferon Induced Protein with Tetratricopeptide Repeats) as shown in figure 11 and 13-A. IFIT family members were found to be expressed during viral infection and following IFN treatment (Diamond et al., 2012). They have unique structure as they lack any enzymatic domains or activity. Rather, they have multiple copies of TPR motif (tetratricopeptide), whose presence in proteins indicates their role in cell cycle regulation, protein folding and protein transport. (D’Andrea LD et al., 2003). IFIT proteins role in inflammation has contradicting results. Some
studies reported their role in reducing the inflammation (Berchtold S et al., 2008; Li Y et al., 2009), while others reported their role in augmenting the inflammatory response. (McDermott JE et al., 2012). Recently, IFIT1, IFIT2 and IFIT3 were found to be severely upregulated in M1 phenotype macrophages (Huang C et al., 2018), indicating their role in inflammatory response. Accordingly, this investigation suggest that IFIT family proteins might have a role in atherosclerosis development. In addition, IFI35 has been previously identified to have a role in neointimal formation and to serve in endothelial cell migration and proliferation (Jian D et al., 2018). Moreover, interferon-stimulated exonucleases (ISG15 and ISG20) and IFI44 upregulation by IRF7 (Lazear HM et al., 2013) in atherosclerosis needs further investigation as they show high expression level in progression-specific macrophages (figure 13-A).

Surprisingly, once the inflammatory module is initiated, distinct feed-forward loops and trans-activation patterns between STAT1 and IRF7 have been developed to maintain the inflammatory process (Figure 11-B), while inhibiting the expression of other cytokines such as TSLP that was found to be upregulated in the regression branch-specific macrophages.

5.5.2 Interrupting the Inflammatory Module

On the contrary, regression branch-specific macrophages interrupt the inflammatory network at multiple levels as seen in figure 11-C. Macrophages in the regression branch show early downregulation of the transcription factor ID2 (Inhibitor of DNA binding 2), which has been reported to upregulate IRF5 (Li J et al., 2017), preventing its downstream upregulation of IRF7, IFNB1 and STAT1 (Lazear HM et al., 2013; Steinhagen F et al., 2015). Additionally, upon upregulation of IGF1 by S1PR1 stimulation (Straub AC et al., 2009), signal transduction through IGF1
interrupts the inflammatory network in multiple ways. For instance, IGF1
downregulates a number of pro-inflammatory cytokines CCL3, IL6 and IL1B as
shown on figure 11-C (Santini MP et al., 2007; Chand HS et al., 2012). Furthermore,
IGF1 can downregulate STAT1 downstream targets such as PSMB8 and PSMB9 (the
immunoproteasome subunits) (Pan Y et al., 2013) indicating their involvement in the
disease progression (figure 13-B).

Likewise, upregulation of IGF1 further results in breaking the inflammatory
module by decreasing activation of STAT1 and by inhibiting expression of IL6
(Bernabei P et al., 2003). IL6 inhibition resulted in upregulation of the anti-
inflammatory cytokine TSLP through downregulation of its inhibitor CCR2 (Bromley
SK et al., 2013). Consequently, TSLP upregulates the expression of the M2
macrophage surface marker CD63 denoting polarization towards M2 anti-
inflammatory phenotype (Wang B et al., 2013).

The analysis of the transcriptional regulatory network offers new insights into
potential roles of different transcription factors, growth factors and signal
transduction through G-protein coupled receptors in controlling expression of a wide
range of genes differentially expressed in macrophages in both branches.

5.5.3 Role of MAFB in Atherosclerosis

MAFB is a basic leucine zipper transcription factor involved in development
and lineage-specific hematopoiesis. (Kelly et al., 2000). It has been reported that
MAFB drives macrophages to M2 anti-inflammatory phenotype by altering various
number of inflammatory related genes (Víctor D et al., 2017) as shown in figure 14-
A. MAFB upregulates the solute carrier SLC40a1 (Ferroportin-1) which has a role in
iron efflux from the cells. Indeed, deposition of iron in atherosclerotic lesion occurs
following intra-plaque haemorrhage (Kolodgie, F D et al., 2003). SLC40a1 was found to be downregulated in the macrophages from the progression branch indicating loss of their ability to export iron outside the cell. Consequently, overloading with iron catalyses reactive oxygen species generation and lipid oxidation in the foam cells, which in turn aids in further progression of the disease (Ramakrishna G et al., 2003). Additionally, macrophages enriched with CD163 in atherosclerotic plaques were found to have less lipid contents and less oxidative damage (Finn et al., 2012).

5.6 Role of the predicted miRNAs

Ten miRNAs were predicted to be upregulated to cause such alteration in the gene expression in the progression branch-specific macrophages as shown in figure 10-B. Some of them were previously identified to play a role in atherosclerosis progression. The role of miRNA-155 in atherosclerosis remains controversial as some studies reported that its overexpression in ApoE\(^{-}\) mice reduces the atherosclerosis (Zhu et al., 2012) and others reported that its deficiency reduces the lesion size in ApoE\(^{-}\) mice (Nazari-Jahantigh et al., 2012). Recently, it has been reported that mice with ApoE\(^{-}\) / miR-155\(^{-}\) double knockout have less atherosclerotic lesion (Anthony Virtue et al., 2017). The exact role of miR-155 in atherosclerosis is not clear. The results indicate that miR-155 can bind to the 3’ untranslated region of multiple mRNAs suppressing the expression of a variety of genes such as LRP1, TGFBR2, NRP1 and MAFB (figure 10-C). Mutation of LRP1 (Low-Density Lipoprotein Receptor-Related Protein 6) was found to cause a number of metabolic diseases such as hyperlipedemia, type 2 diabetes (Mani A. et al 2007). Particularly, LRP6 mutation is associated with increase plasma LDL level (Tomaszewski M et al., 2009) which is one of the major risk factor of atherosclerosis development. NRP1 (Neuropilin 1) deficiency shifts the energy metabolism of macrophages towards glycolysis (Ariel
Molly Wilson et al., 2018) as in the M1 phenotype resulting in more lactate and reactive-oxygen species production and consequently more inflammation and inefficient fatty acids burning (West AP et al., 2011).

ABCA1 (ATP Binding Cassette Subfamily A. Member 1) was predicted to be targeted with miR-148, miR-140 and miR-758. ABCA1 was previously identified to stimulate cholesterol efflux (Wang N et al., 2000, Oram JF et al., 2000). Mutations of ABCA1 resulted in increasing cholesterol deposition in cells and tissues, especially in macrophages (Assmann et al., 1995) and a significant reduction in plasma HDL level (Marcil M et al., 1999). Anti-miR-148a shows reduction in the atherosclerotic plaque formation (Rotllan N et al., 2018). MiR-410 inhibition was found to rescue endothelial cells apoptosis by regulating oxidized-LDL levels (Ming-Yan Hu et al., 2018). MiR-758 was identified before to target ABCA1 resulting in suppressing cholesterol efflux from the cell (Ramirez CM et al., 2011)

IGF1, MAF and MAFB have been identified to be targeted with the predicted miRNAs, resulting in suppressing their role in resolving the inflammation in the progression-specific macrophages.

MiR-129-5p upregulation was suggested to have a protective role in atherosclerosis through suppression of endothelial cell autophagy (Zhaohua Geng et al., 2016), which was predicted in the results to be upregulated in the progression-specific macrophages. This contradicting finding needs further validation.

The miRNAs prediction results confirmed the previously identified findings from the literatures and revealed additional roles of those miRNAs. In addition, the results offer new insights on other miRNAs involvement in the disease progression such as miR-344-3p, miR-346 and miR-485-5p.
CHAPTER 6: CONCLUSION AND FUTURE PERSPECTIVES

6.1 Conclusion

Atherosclerosis inflammatory sites contain both pro-inflammatory and anti-inflammatory macrophages. Each subset of macrophage activates unique transcriptional program with distinct pathways and metabolic activities. Activating integrin signalling and maturation of dendritic cells by the inflammatory macrophages aggravates the inflammation by ECM remodelling. Inflammatory macrophages express distinct set of genes to activate an inflammatory module aggravating the atherosclerosis inflammation and disrupting cholesterol metabolism resulting in intracellular cholesterol retention and foam cell differentiation. By upregulating certain genes such as STAT1 and IRF7, macrophages can activate various pathways to produce inflammatory cytokines which have positive regulatory effects on those genes to maintain the inflammatory process. In contrary, upregulation of genes responsible for cholesterol efflux and prevention of foam cell formation were found in the anti-inflammatory macrophages. In addition, interrupting the inflammatory module in M2 macrophages can be achieved by altering expression level of certain transcription factors and key genes that act systematically to dampen the inflammatory network by interrupting it at various levels. Anti-inflammatory macrophages upregulates IGF1 and MAFB and downregulate ID2 transcription factor, which in turn interrupt the inflammatory network through downregulating STAT1, IRF7, IL6, IL1B and their targets as well. Understanding how anti-inflammatory macrophages suppress inflammation is a key step to better understand atherosclerosis resolution and offers new insights on how to force the transition of macrophages from inflammatory phenotype to anti-inflammatory phenotype.
6.2 Future Perspectives

This study has some limitations regarding the validation of the detected candidates and the techniques used. It is limited to the computational analysis of distinct macrophage states during atherosclerosis progression and regression. Future work to experimentally validate the candidates is required. Additionally, although single cell RNA sequencing allows detection of the heterogeneity among the total population, the technique has shallow coverage and not like deep bulk RNA sequencing. For instance, the average number of genes detected per cell is ranging from 1000 to 2000 genes per cell in this dataset. Also, alternative splicing events are difficult to be detected among the clusters because of the low number of reads per gene in each cell. It is recommended for future studies to take into consideration performing bulk RNA sequencing on cluster-basis. For instance, a combination of unique surface markers can be identified for progression-specific clusters (cluster six and seven) and the regression-specific one (cluster four) for sorting them and performing deep RNA sequencing. This will extend the network analysis to include more candidates and pathways that may be identified as therapeutic targets. Furthermore, understanding how macrophages gradually change their states from inflammatory to anti-inflammatory phenotypes can be used to force them to enter anti-inflammatory state. Future research and experimental validations are required using atherosclerotic mice models to confirm the findings.
References


Agarwal V, Bell GW, Nam J, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015;e05005.


Chung NC, Storey JD. Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics. 2015;31:545-54.


Kloetzel PM. Generation of major histocompatibility complex class I antigens: functional interplay between proteasomes and TPPII,” Nature Immunology. negative cross-talk between liver X receptors (LXRs) and STAT1: effects on IFN-γ-induced inflammatory responses and LXR-dependent gene expression. 2005;5:661–669.


Ramakrishna G, Rooke TW, Cooper LT. Iron and peripheral arterial disease: revisiting the iron hypothesis in a different light. Vasc Med. 2003; 8:203-10.


APPENDIX

Figure 1 copyright agreement

Copyright: © International Journal of Preventive Medicine

This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Dec 01, 2019

This Agreement between Dr. Ahmed Abouhashem ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number 4718701417644

License date Nov 30, 2019

Licensed Content Publisher Elsevier

Licensed Content Publication Clinica Chimica Acta

Licensed Content Title Macrophage polarization in atherosclerosis

Licensed Content Author Sai Yang, Hou-Qin Yuan, Ya-Meng Hao, Zhong Ren, Shun-Lin Qu, Lu-Shan Liu, Dang-Heng Wei, Zhi-Han Tang, Ji-Feng Zhang, Zhi-Sheng Jiang

Figure 2 copyright agreement
Figure 3 copyright agreement

Modified image

ELSEVIER LICENSE
TERMS AND CONDITIONS

Dec 01, 2019

This Agreement between Dr. Ahmed Abouhasem ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number 4720160992252
License date Dec 01, 2019
Licensed Content Publisher Elsevier
Licensed Content Publication The American Journal of Pathology
Licensed Content Title Monocyte and Macrophage Plasticity in Tissue Repair and Regeneration
Licensed Content Author Amitava Das, Mithun Sinha, Soma Datta, Motaz Abas, Scott Chaffee, Chandan K. Sen, Sashwati Roy