

# Computational strategies in cardiometabolic diseases

Citation for published version (APA):

Lu, C. (2022). *Computational strategies in cardiometabolic diseases: a portal to deeper mechanistic understanding*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20221124cl>

## Document status and date:

Published: 01/01/2022

## DOI:

[10.26481/dis.20221124cl](https://doi.org/10.26481/dis.20221124cl)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

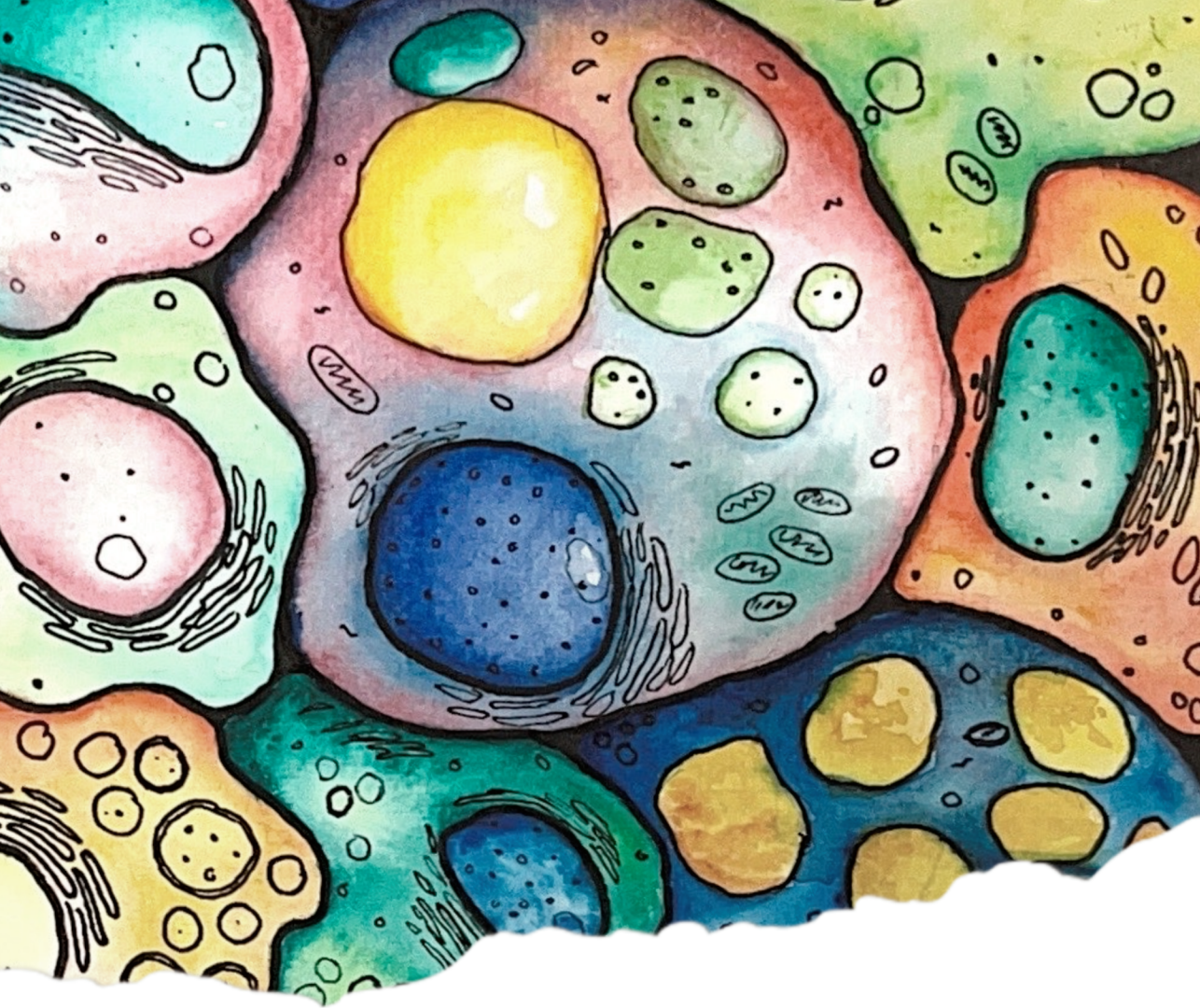
[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



# Computational strategies in cardiometabolic diseases

*a portal to deeper mechanistic  
understanding*

Chang Lu







**Computational strategies in  
cardiometabolic diseases:  
a portal to deeper mechanistic  
understanding**

**Chang Lu**

© Chang Lu, Maastricht 2022

All rights reserved. No part of this book may be reproduced, stored in a retrieval system of any nature, or transmitted in any form by any means (electronic, mechanical, photocopying, recording, or otherwise) without prior written permission of the author, or when appropriate, by the publishers of the publications.

ISBN	978-94-6469-127-6
Author	Chang Lu
Layout and figure design	Chang Lu
Cover design	Valeria Saar-Kovrov and Chang Lu
Printed by	DissertationMake   <a href="http://www.proefschriftmaken.nl">www.proefschriftmaken.nl</a>



The research presented in this thesis was conducted at the Department of Pathology, Maastricht University Medical Center+ (MUMC+). The author is financially supported by the China Scholarship Council (CSC, No. 201706990018).



**Computational strategies in  
cardiometabolic diseases:  
a portal to deeper mechanistic  
understanding**

Dissertation

to obtain the degree of Doctor at Maastricht University,  
on the authority of the Rector Magnificus,  
Prof. dr. Pamela Habibović,

in accordance with the decision of the Board of Deans,  
to be defended in public on

Thursday, 24<sup>th</sup> of November 2022, at 16.00 hours

by

Chang Lu

**Supervisors**

Prof. dr. Erik A.L. Biessen

dr. Joël M.H. Karel

**Co-supervisor**

dr. Pieter Goossens

**Assessment committee**

Prof. dr. I.C.W Arts (Chair)

Prof. dr. J. Jankowski (Maastricht University, Netherlands; Rheinisch-Westfälische Technische Hochschule Aachen, Germany)

Prof. dr. T. Ulas (University of Bonn, Germany)

dr. C. Seiler

# TABLE OF CONTENTS

<b>Chapter 1</b>	General introduction	5
<b>Chapter 2</b>	Integrating multiplex immunofluorescent and mass spectrometry imaging to map tissue myeloid heterogeneity in its metabolic and cellular context	39
<b>Chapter 3</b>	Spatio-temporal metabolomics reveals metabolic changes in the early stage of non-alcoholic fatty liver disease in mice	89
<b>Chapter 4</b>	Cardiovascular disease but not health is hallmarked by sex-specific cytokine signaling pathways in circulating monocytes	119
<b>Chapter 5</b>	Identification of a gene network driving the hypertension associated dampened monocyte LPS response in coronary artery patients	149
<b>Chapter 6</b>	General discussion	179
<b>Chapter 7</b>	Summary   Samenvatting   总结	201
<b>Chapter 8</b>	Impact	211
<b>Appendices</b>	List of abbreviations Acknowledgements   致谢 Curriculum Vitae	217





# **Chapter 1**

General introduction





## Cardiometabolic diseases

Cardiometabolic diseases (CMDs), including cardiovascular disease (CVD), diabetes, heart attack, stroke, and non-alcoholic fatty liver disease (NAFLD), are the leading causes of death and disease burden in the world [1, 2]. Especially in the last decade, the incidence and prevalence of CMDs have increased dramatically, not only in developed countries, but also in developing countries with emerging economies [3–6]. According to new WHO predictions, more than 23 million people will die from CVDs by 2030 [7]. The prognosis for China may be at least equally alarming since by 2020, as there were already approximately 330 million people suffering from CVDs, including 13 million strokes and 8.9 million heart failures [8]. The demographics of NAFLD are even more concerning, as it is prevalent in a quarter of European adults [9]. These facts underscore that more effective and innovative diagnostic and therapeutic measures of CMDs are eagerly awaited.

The widespread prevalence of unhealthy lifestyles, such as high-fat diets, the lack of physical activity, smoking, and excessive alcohol consumption, are major risk factors for the dramatic rise in CMDs [10]. Other factors including age, gender, and genetics also influence their development and progression. For example, it has been well known that the elderly population is particularly susceptible to CVD [11, 12]. Although the lifetime risk of CVD is similar in men and women, the onset of CVDs in men approximately 5–10 years earlier than that in women [13]. Therefore, exploring the impact of risk factors on CMDs is pivotal for elucidating multifaceted mechanisms associated with many cardiometabolic complications, and opening up new therapeutic horizons.

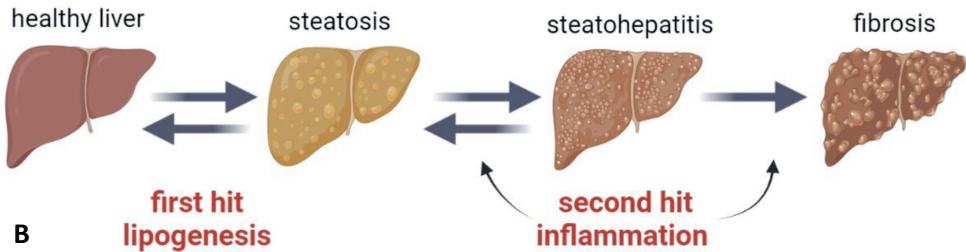
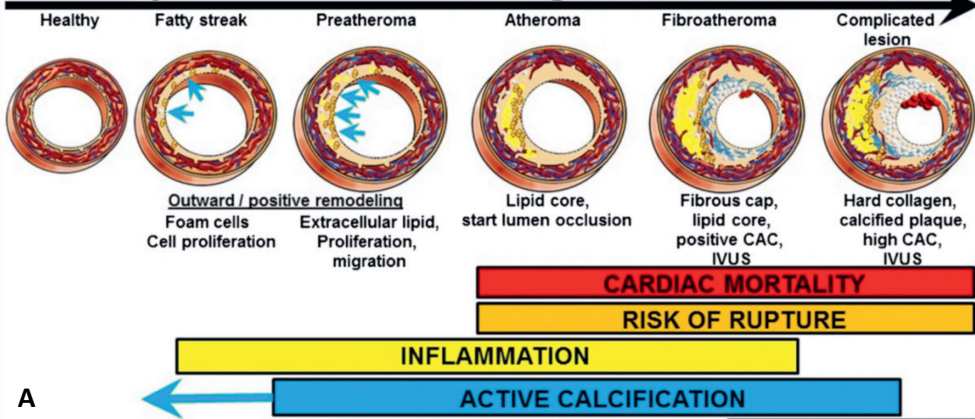
This thesis examines the disease mechanisms of two of the most common CMDs, i.e., CVD and NAFLD, by developing and employing multiple computational strategies. Therefore, I will first present the basic biomedical background involved in atherosclerosis (the major cause of CVD) and NAFLD, followed by the introduction of commonly used bioinformatics approaches.

### The mechanism of atherosclerosis

CMDs are often shown as slowly progressing diseases, accelerated by chronic tissue inflammation [14]. For instance, atherosclerosis, as the main cause of CVD, develops gradually from early fatty streak to early fibroatheroma and thin-cap fibroatheroma with the risk of rupture. This process is accompanied by plaque thickening within the arterial wall, which has an inflammatory reaction and lipid build-up at its core. [15] (**Fig 1A**). Numerous studies have identified that vascular cells produce and respond to inflammatory cytokines [16–18], providing strong evidence that immune mechanisms are involved in atherosclerosis. Additional findings, such as the association of autoantibodies to oxidized low-density

lipoprotein (oxLDL) and heat shock protein (HSP) 60 with atherosclerotic CVD [19–22], also further supports this notion.

## Coronary Atherosclerosis Progression

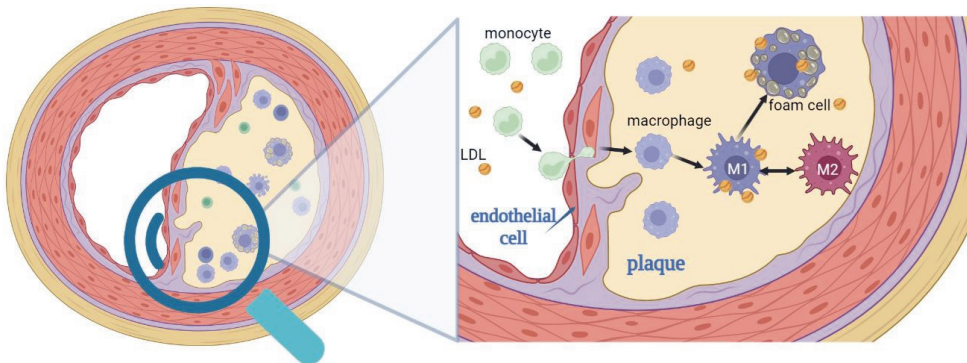


**Figure 1: Progression of atherosclerosis and NAFLD (A)** Progression from healthy arteries to complicated atherosclerotic lesions (copied from [23]) **(B)** Development and progression from healthy livers to steatosis (NAFLD) to non-alcoholic steatohepatitis (NASH) and advanced fibrosis. The figure was drawn using BioRender (<https://biorender.com/>)

Lipid accumulation within the plaque and immune cells, particularly macrophages and their circulating progenitors, *i.e.* monocytes, are the main driving factors of atherosclerosis [24, 25]. Based on the expression of the cell surface markers CD14 (Lipopolysaccharides (LPS) co-receptor) and CD16 (FcγIII receptor), human monocytes can be divided into three phenotypes: classical (CD14<sup>+</sup> CD16<sup>-</sup>), intermediate (CD14<sup>+</sup> CD16<sup>+</sup>), and non-classical monocytes (CD14<sup>-</sup> CD16<sup>+</sup>) [26, 27]. Their differential expression of adhesion molecules, chemokine receptors and functions leads to their differing roles in CVD [28–30]. Previous studies [30–32] have revealed that CD14<sup>+</sup> monocytes are more phagocytic, whereas CD16<sup>+</sup> monocytes appear to be specialized for anti-viral rather than anti-bacterial immunity. CD14<sup>+</sup> CD16<sup>-</sup> monocytes, the most abundant monocytes in humans, have been reported to be more susceptible to enter atherosclerotic lesions [29]. However, the precise roles of monocyte subpopulations in CVD, and in particular

how they are influenced by physiological characteristics and risk factors, have not been fully determined yet. Therefore, in the **chapter 4** and **5**, I analyse transcriptional profiles of monocytes from a CVD cohort collected by the Center for Translational Molecular Medicine (CTMM) [33], to investigate CVD-related sex differences in the monocytes (**chapter 4**) and the impact of CVD risk factors on the responsiveness of monocytes to inflammatory stimuli (**chapter 5**).

Activation of endothelial cells induces monocytes to roll on, adhere to and eventually migrate through the vessel wall into the subendothelial space, where they mature into macrophages under the influence of locally produced growth factors [34, 35] (**Fig 2**). Here, these macrophages take up and accumulate lipoproteins, resulting in their lipid-laden phenotype that is also known as foam cells [35] (**Fig 2**). Due to impaired mobility, foam cells gradually accumulate in the subendothelial space, contributing to plaque growth, and eventually die, resulting in a plaque-destabilizing necrotic core [36].



**Figure 2: Potential contributing macrophage sources in atherosclerotic plaques in metabolic disease.** The figure was drawn using BioRender (<https://biorender.com/>)

The ability of macrophages to develop focused responses to a wide variety of stimuli in their local microenvironment is known as plasticity [37]. A traditional dichotomous theory divided macrophages into M1 (classically activated) and M2 (alternatively activated) phenotypes [38, 39]. The former phenotype is induced by interferon- $\gamma$  (IFN- $\gamma$ ) and displays pro-inflammatory and antigen-presenting functions, while the latter is stimulated by interleukin 4 (IL-4) or interleukin 13 (IL-13), playing an anti-inflammatory and wound healing role [40–43]. Recent insights, however, show that M1 and M2 are rather two extremes of macrophage polarization, and a spectrum of functional intermediate transition states exists between them in different tissues and conditions [44–46]. For instance, a macrophage phenotype induced by oxidized phospholipids (OxPLs), called Mox, [47] features a metabolism that differs from that of M1 and M2 [48]. In parallel, plenty of other macrophage phenotypes have been described in atherosclerotic plaques [49]. Therefore, there is an urgent need to develop effective methods for



the systematic understanding of macrophage heterogeneity and plasticity, with the potential of future targeted interventions [50]. This is exactly what will be studied in **chapter 2**.

### The mechanism of NAFLD

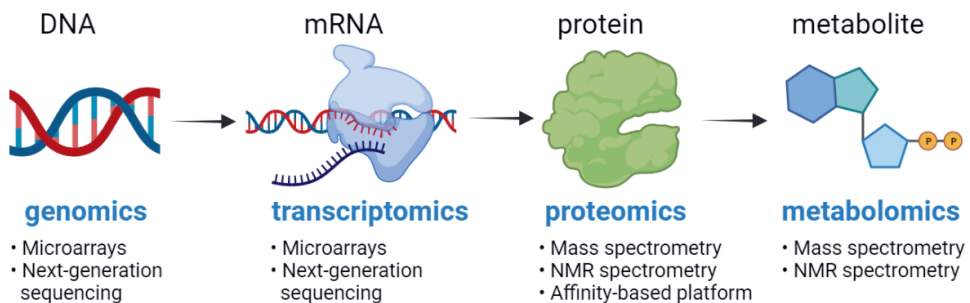
As one of the most common chronic liver diseases, NAFLD encompasses a range of progressive liver lesions, including simple steatosis, steatohepatitis (NASH), and fibrosis [51] (**Fig 1B**). The current consensus about the mechanism of NAFLD is a ‘two-hit hypothesis’: 1) accumulation of lipids in the hepatocytes, resulting in an increased sensitivity of immune cells; 2) lobular inflammation that characterizes NASH aggravating disease progression [52–54]. As lipid accumulation and inflammation progress, the disease progress and histological changes become increasingly complex and may vary considerably between individuals and within any given individual. Early lesions may regress spontaneously, but intermediate and late stage lesions appear to be progressing [55] (**Fig 1**). Therefore, studying the gradual changes in diseased tissues helps to improve our understanding of NAFLD.

Similar to atherosclerosis [25, 56], NAFLD is also a complex disease with pathogenesis involving dietary, genetic, environmental, and metabolic factors that progress through the disease stages [57, 58]. Unhealthy dietary habits, especially excessive sugar and fat intake, accelerate the prevalence of NAFLD. Fructose enters the liver and is metabolized to triglycerides (TG) via the de novo lipogenesis [57, 59]. NAFLD arises when fatty acid (FA) and TG uptake from circulation and de novo synthesis of lipids (lipogenesis) exceeds the metabolic capacity of the liver leading to lipid accumulation [60, 61]. Therefore, studying changes in metabolism and their role in the pathogenesis of NAFLD may provide new clues for interventions in NAFLD.

Additionally, the analysis of metabolite trajectories should not ignore the heterogeneity of liver tissue since the liver includes 2 main types of cells: parenchymal and nonparenchymal cells. Of these, parenchymal cells are the main component of the liver, accounting for 90% of the total cell mass [62, 63]. Nonparenchymal cells, including Kupffer cells, hepatic sinusoidal endothelial cells, and stellate cells, not only perform different functions but also exhibit a distinctly different spatial distribution from parenchymal cells [64]. Studying this spatial heterogeneity enable precise dissection of the specific functions of different liver compartments. Therefore, **chapter 3** analyses the metabolic trajectories of mice in the early stages of NAFLD in the whole liver tissue and in individual liver compartments.

## Bioinformatics in Cardiometabolic diseases

Rapid advances in high-throughput technologies nowadays allow us to measure molecular profiles within tissues or cells at an unprecedented resolution. For example, next-generation sequencing (NGS) technologies have made it possible to rapidly assess the expression of tens of thousands of genes [65], while steadily progressing technologies in mass spectrometry (MS) enable us to not only measure but also to localize peptides, metabolites and lipids [66, 67]. The collection of techniques for unbiasedly detecting expressions or abundance of genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in organisms is called ‘omics’[68] (**Fig 3**). Each omics dataset provides a unique view of a biological system, and they have demonstrated their powerful ability at uncovering CMD-causing characteristics. In the remainder of this chapter, I introduce the development of omics technologies, the most common corresponding computational methods, and their application in CMD research and to the studies in my thesis.



**Figure 3: Main omics procedures and analytical platforms used currently in medicine and in CMD research.** The figure was drawn using BioRender (<https://biorender.com/>)

### Bulk omics techniques and analysis approaches

Bulk omics refers to the characterization of genes, proteins and metabolites extracted from homogenized tissues and cell lysates, which has provided substantial breakthroughs in our understanding of biological processes [66, 69–72]. Microarrays and NGS (such as RNA sequencing (RNA-Seq)) continue to be the major techniques for producing bulk transcriptomics data, while MS is a commonly used tool for proteomics and metabolomics (**Fig 3**)[72]. Among these, transcriptomics has obtained a central role in the study of CMDs due to the ease of use and cost advantages of data acquisition [64,65,67].

A DNA microarray (or gene chip) is a small slide with short DNA probes on it [73]. Gene expression will be detected upon RNA hybridization to these probes [74–76]. RNA-Seq on the other hand is a high-throughput technique based on simple counting of reads that can be reliably aligned to a reference sequence [77, 78].

Although microarray technology is still used to study CVD [79], with the development of NGS, RNA-Seq is gradually eroding its dominance in recent years as it addresses some of the limitations of microarrays [70, 75, 80]. The most obvious advantage of RNA-Seq is that instead of prior probe selection, RNA-Seq allows to sequence the whole transcriptome, implying its advantage in the discovery of novel transcripts, allele-specific expression and splice variants. Additionally, compared to a standard whole-genome microarray, RNA-Seq allows for higher resolution and much lower detection limits for differentially expressed genes [75, 81]. These powerful benefits have resulted in the development of a great number of bioinformatics tools for RNA-Seq analysis [82]. However, microarrays often remain the first choice when studying larger cohorts, because of their lower assay cost as well as more limited data storage and easier analysis [75, 83]. Several studies have also already shown the high correlation of differential expression profiling between RNA-Seq and microarrays, indicating their similar analysis results based on known genome sequences [77, 81].

The pre-processed bulk microarray or RNA-Seq data can be represented as a gene expression matrix, where each row represents a gene and each column represents a sample. Each entry in the matrix represents the expression level of a specific gene in a given sample (**Fig 4**). A traditional analysis pipeline for the gene expression matrix can be summarised as two steps: (1) Identification of characteristic genes (or gene modules) that relate to experimental, clinical or treatment conditions, followed by (2) downstream analysis for biological interpretation.

### Differential gene expression analysis

The purpose of differential gene expression analysis is to calculate changes in gene expression under specific conditions. A simple and traditional approach is to calculate the (log 2) fold change for each gene in the two sets of samples. However, this method does not take into account the variance of the expressions [84]. Therefore, the selection of differential expressed genes (DEGs) always needs to be combined with hypothesis testing [85]. A widely used method is linear-model-based limma [86], which tests for differential gene expression by moderated t-statistics. In contrast to t-statistics calculated independently for each gene, moderated t-statistic takes into account both the gene under test and genes with similar magnitude of variation to shrink the standard errors towards a global value. Each gene's log 2-fold change, t-statistics, and corrected p-value for multiple comparisons can be used for generating gene expression signatures (GES) for subsequent downstream analysis.

### Gene co-expression analysis

Gene co-expression analysis is a favoured strategy to identify relevant gene sets [87]. The workflow includes the following steps [87]: A correlation network is first

constructed based on pairwise correlations between genes [88–90], where a node presents a gene, and an edge indicates the correlation of expression between genes. Next, gene modules (co-expressed gene clusters) are identified using clustering methods. Finally, the modules that correlate to one or several experimental or clinical conditions are extracted for further downstream analysis. A well-established co-expression analysis method is Weighted Gene Correlation Network Analysis (WGCNA) [89], which defines a scale-free network by non-linearly enhancing the gap between the edges with high and low correlation as follows:

$$w_{ij} = |\text{corr}(X_i, X_j)|^\beta \quad (1)$$

Where  $w_{ij}$  is the weight between gene  $i$  and  $j$ ,  $\text{corr}(X_i, X_j)$  the Pearson correlation between  $i$  and  $j$ , and  $\beta$  ( $>1$ ) is a soft power of nonlinear enhancement. Gene modules are clustered using hierarchical clustering with the topological overlap measure (TOM) as proximity measure. The main idea behind TOM is that the more similar two genes are and the more neighbours they share, the more likely they are to belong to the same module. In brief, compared to differential expression analysis, co-expression analysis can infer the association between a gene module and multiple factors (such as disease, treatments), as well as the interactions among genes, which allows to dissect a gene's functional status identified from a systematic perspective.

### Gene regulatory network (GRN) reconstruction

Like co-expression network analysis, GRN reconstruction is also a network-based tool, but rather aimed at understanding how transcription factors (TFs) drive changes in gene expressions [91]. The purpose of GRN reconstruction is to infer the target genes of TFs, based on gene expressions of a set of samples [91]. In the past decades, a large amount of machine learning-based methods was proposed for GRN reconstruction. Some unsupervised methods such as MRNET [92] and ARACNE [93] are well-received in the biology community since they capture non-linear dependencies between genes using Mutual Information (MI) and computational cost is low [94]. To be specific, ARACNE first generated a co-expression network using MI and discarded indirect interactions using Data Processing Inequality (DPI). DPI states that if genes  $i$ ,  $j$  and  $k$  are linked as a triplet, then the interaction between  $i$  and  $j$  will be eliminated if [93]:

$$I(X_i, X_j) \leq \min(I(X_i, X_k), I(X_j, X_k)) \quad (2)$$

where  $I(X_i, X_j)$  is the MI between gene  $i$  and  $j$

### Enrichment analysis

Gene expression signatures or gene modules of interest that were obtained from differential expression analysis or co-expression analysis, can be further used for

assessing their role in the cell's molecular mechanisms, such as pathway activity, cytokine production, transcription factor activation and even for drug repurposing [95–98]. Some approaches to these analyses are completed by mapping a reference gene list from an *a priori* knowledge database to the GES or gene modules of interest and calculating their similarity or overlap, to identify terms which are statistically over- or under-represented within the list of interest. In this section, I briefly introduce the methods used in this thesis, including overrepresentation analysis (ORA) [99], gene set enrichment analysis (GSEA) [96] for pathway analysis, DoRotheEA [98] and VIPER [100] for TF activity inference, cytosig [101] for inference of responses to cytokines and growth factors, and Connectivity Map for drug repurposing.

ORA and GSEA, the most often used statistics-based methods for pathway enrichment analysis, are commonly used in combination with Gene Ontology (GO) [102] or pathway databases (e.g. KEGG [103], Wikipathways [104]). ORA computes if genes from a GO term or a pathway are overrepresented in the gene sets of interest (such as DEGs) using a one-sided version of Fisher's exact test as follows [105]:

$$p = 1 - \sum_{t=0}^{q-1} \frac{\binom{M}{t} \binom{N-M}{n-t}}{\binom{N}{n}} \quad (3)$$

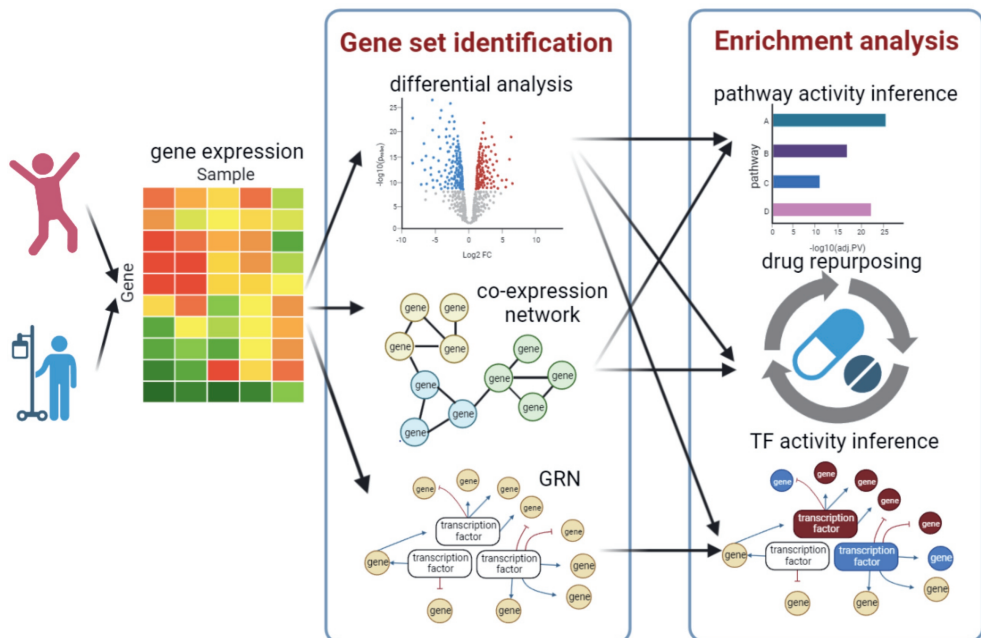
where  $N$  is the count of background genes (which are usually the size of GES);  $M$  is the count of overlap between background genes and genes in a pathway;  $n$  is the number of genes of interest and  $q$  is the size of overlap between genes of interest and the genes in this pathway. However, ORA regards all differentially DEGs evenly, which may lose the information from the genes with weak differential expressions. GSEA calculates the extent to which genes in predefined GO terms or pathways as a whole are up- or down-regulated in the GES list, contributing to detecting small but consistent changes in a set of genes.

TF activity inference algorithms aim to find those transcription factors that strongly regulate their target genes in response to specific stimuli (e.g. diseases, drugs) [106, 107]. Since assessing TF activation directly (e.g. Western blot or immunofluorescence) can be challenging, most computational approaches predict TF activities based on differential expression data and a regulatory network [107]. This regulatory network can be constructed using models of TF binding sites (TFBSs) [108] or reverse engineering based on gene expression profiles [93]. The activities of TFs can be then inferred using a statistics-based enrichment analysis. For example, VIPER [100] first generated a regulatory network using ARACNE and calculated an enrichment score for each TF on the GES using an analytic Rank-based Enrichment Analysis algorithm (aREA), by rewarding the TFs whose targets have the same directions of regulations as GES and penalizing the TFs that have opposite

regulations. However, the reliability of TF-target networks inferred by a single computational method remains to be verified. To address this issue, Garcia-Alonso *et al.* recently integrated TF-target interactions inferred from literature curated resources, ChIP-seq peaks, TF binding site motifs, and reverse engineering as a resource called DoRotheEA [98], which illustrates the validity of combining literature-curated resources with computational prediction results.

The advantages of data integration have also been applied on cytokine activity analysis. With the prime example of Cytosig [101], a tool including an integrated database of target genes regulated by cytokines and a ridge-regression-based model for cytokine signalling activity prediction.

Drug repurposing is an effective strategy to discover novel indications for approved or experimental drugs by in silico or by experimental screening pipelines. An extensively used database for this analysis is Connectivity Map (Cmap) and its successor LINCS [97], both of which include a reference collection of gene-expression signatures that connect genes, drugs and disease states. The candidate drugs were selected using weighted connectivity score (WTCS), a bi-directional version of GSEA[96]. Drugs with negative score indicate that their GESs are opposite to those from the gene set of interest, implying the potential to treat this disease.



**Figure 4: approaches for bulk transcriptomics analysis.** The figure was drawn using BioRender (<https://biorender.com/>)

## Uncovering mechanisms of CVD using bulk analysis methods

1

The approaches above have been used for the discovery of CVD-causing genes, proteins, or metabolites, and for exploring their association with CVD-related genetic and environmental factors, to uncover the disease mechanisms and develop novel therapeutic strategies. For example, Jin et al. [109] identified sex-specific plaque-phenotype-related genes that were implicated in fibrosis and inflammation, by comparing unstable plaque segments to stable segments from an all-male MaasHPS (Maastricht Human Plaque Study) transcriptomics cohort of carotid endarterectomy plaques, and then validated on STARNET aortic plaque cohort. A recent study based on arrays of plasma metabolites constructed sex-specific metabolite–metabolite association networks using ARACNE, observing sex-related variability in branched-chain amino acids, ketone bodies, and propanoate metabolism [110]. Furthermore, Nath et al. identified 5 molecular pathways related to cardiovascular mortality in heart failure based on differential expression genes from 944 patient’s whole blood transcriptomic profiling and then inferred candidate drug targets using mortality-related transcript patterns [111]. These studies demonstrated the strength of bulk omics analysis in revealing molecular mechanisms underlying CVD. In the **chapter 4** and **5**, I apply multiple approaches above to analyse the microarray gene expression profile from monocytes of CTMM cohort, including 450 CVD patients to identify genes (modules) associated with gender (**chapter 4**) or CVD risk factors (**chapter 5**), and to infer the TFs, pathways and cytokines that related to these genes.

### From bulk to single cell: towards to cell heterogeneity

Bulk analysis provides an averaged profile of a mixed pool of a huge number of cells in tissues, but is not able to precisely capture the expression profiles of the cellular constituents of the tissue, although they can be inferred to some degree by deconvolution[72, 112–114]. In recent years, however, technological advances have made it possible to explore omics information at a single-cell resolution, resulting in a better understanding of the function of each individual cell [114, 115]. One of the most well-known transcriptomic techniques is single-cell RNA sequencing (scRNA-Seq), which can detect RNA molecules in individual cells with high resolution and on a genomic scale [116]. In the field of proteomics, flow cytometry [117] and CyTOF [118] are the most universal single-cell techniques, currently allowing to detect up to 40 biomarkers simultaneously. From the perspective of data structure, similar to the bulk technology, the single-cell data can also be processed as an expression matrix, only here the rows of the matrix change from a sample to a cell.

In addition to the analyses mentioned above for bulk transcriptomics, cell clustering and visualization are crucial steps for distinguishing individual cell phenotypes within a sample [119]. Some classical machine-learning-based clustering methods

(such as k-means and hierarchical clustering) have been widely applied to group cells into several phenotypes based on the similarity of their expression [120–122]. To allow interpretation of the often huge-sized single cell expression data, graph-based clustering algorithms (e.g. Louvain [123, 124] and Shared Nearest neighbour(SNN) [125, 126]) have become popular, due to their scalability and efficiency.

Cell clusters are usually visualized by projecting high-dimensional omics data into a two- or three-dimensional space, using dimensionality reduction[127]. For example, Principal component analysis (PCA) [128], one of the most long-established linear dimensionality reduction methods, finds a set of mutually orthogonal vectors (i.e., principal components) by a linear transformation, and then ranks the principal components in decreasing order according to variances they can explain. By projecting the data onto the first few principal components of the row, visualization can be achieved while retaining as much global information as possible. In case of a complex nonlinear structure, however, PCA is not the optimal method for visualization because the local structure (clusters) of data may be lost. Therefore, nonlinear dimensionality reduction techniques, such as t-distributed Stochastic Neighbour Embedding (tSNE) [129] and more recent alternatives like Uniform Manifold Approximation and Projection (Umap) [130] have become the most common approaches to visualize cells discriminated into distinct phenotypes.

With the booming variety of single-cell computational tools, in the context of atherosclerosis, the quest for single-cell approaches has also made great progress. Zerneck *et al.* identified multiple cell subpopulations in healthy and atherosclerotic aortas, by integrating nine scRNA-Seq datasets [131]. Cochain *et al.* focused on macrophages from mice aorta, finding IL-1 $\beta$ -rich inflammatory macrophages and TREM2-high macrophages are enriched in atherosclerotic aortas and exhibit functions related to lipid metabolism [132]. A study based on the expressions of scRNA-Seq and CyTOF data from leukocytes in the aorta of atherosclerotic mice detected and validated function-specific phenotypes of leukocytes [133]. These studies demonstrate the effectiveness of single-cell transcriptomics and proteomics in the dissection of cellular heterogeneity in an atherosclerotic environment. Therefore, in **chapter 2** I utilize multiple clustering and visualization approaches described above, to identify macrophage heterogeneity.

## From single image to image cube: exploring the cellular microenvironment

Although single-cell techniques have demonstrated success in profiling cell heterogeneity, these approaches require tissue dissociation, resulting in all spatial contexts of cells within the tissue being lost. This information however is critical for understanding the complex heterogeneity of cells within one tissue, such as an



atherosclerotic plaque [67, 114, 134]. On the other hand, some traditional tissue staining techniques, such as haematoxylin and eosin (H&E) stain, and immunohistochemistry (IHC) or immunofluorescence (IF) stain [135] can help biologists and pathologists observe the cells' tissue localization and identify differences in cell morphology and intensities of a specific protein marker, but the limited number of features that can simultaneously be assessed constrains the applicability in the study of cell heterogeneity. To address this problem, several spatially resolved omics techniques have been proposed [136–138]. Multispectral or hyperspectral imaging (or “lambda scanning”) of a tissue that has undergone multiplex IHC or IF staining microscopically reveals the presence and abundance of multiple biomarkers in parallel [136]. Imaging Mass Cytometry (IMC) is able to visualize up to 40 markers simultaneously by labelling antibodies with tags that can be detected by mass spectrometry imaging [137]. Furthermore, the development of spatial transcriptomics [139] techniques has gradually reached a sub-single-cell resolution in the last two years. For example, high-definition spatial transcriptomics (HDST) enables the measurements of gene profile at a resolution of 2  $\mu\text{m}$  [140], and Seq-Scope can even reach nanometer resolution (0.5-1  $\mu\text{m}$ ) [141]. These novel techniques will greatly advance the *in-situ* characterization of different cell types and subsets within their tissue context. All these techniques eventually generate a 3D image cube, stacking multiple layers of 2D images (the spatial x and y coordinates) with the third dimension representing individual protein markers, mass to charge (m/z) ratios or gene expression levels.

For those spatial omics techniques that have not yet (completely) reached a single-cell resolution, some approaches cluster an individual pixel (or spot) to analyse the spatial heterogeneity [142–145]. For the spatial omics data at sub-single-cell resolution (e.g., multiplexed imaging), prior cell segmentation is required to quantify the signals over the individual cells. Next, based on the physical locations of cells, cell-cell interactions can be studied [146, 147] by identifying the interactions in local cellular niches (the first-order neighbours) or cells within a fixed distance from each other [146, 148].

The application of spatial omics in CMDs is still in its early stages, despite the fact that spatial information is a key factor in understanding cell heterogeneity [149, 150]. Wu *et al.* analysed cells and their neighbourhoods from pancreas in type 2 diabetic patients by quantifying 34 markers of pancreatic exocrine, islet, and immune cells and stromal components using IMC [151]. Saldarriaga *et al.* combined 5 antibodies targeting macrophages into a multiplex protocol and observed the number of macrophages significantly increased in the HCV-positive patients with advanced fibrosis [152]. It is conceivable that the tremendous advantages offered by spatial omics will break new ground in the exploration and treatment of CMDs. In **chapter 2**, we aim to link 3 types of imaging data (cube) from the same mouse, including Matrix-assisted laser desorption ionization-Mass spectral imaging (MALDI-

MSI) data (containing the signal intensities (aka. a mass spectrum, >20,000 dimensions) of every pixel on a mouse section), sets to multispectral images (similar to MALDI-MSI but with higher resolution and lower dimensionality (15 features)), and histology (H&E-stained images; 1 image per sample), ultimately aiming to identify molecular patterns associated with the presence of deleterious cell types or activation states.

## From single omics to multi-omics

Although omics studies have identified genes, proteins and metabolites associated with some diseases or phenotypes, single omics remain limited to a one-sided view of the full picture [153, 154]. For example, protein levels are a sum of the upstream transcriptional activity and subsequent post-translational modifications, which are difficult to observe from transcriptomic profiles. Metabolites are the end products of complete biological processes with several of the intermediate steps potentially responsive to genetic or environmental changes [155, 156]. Therefore, an integrative analysis of data from multiple omics can help researchers to fully understand molecular interactions, bridging the gap between various omics data [153, 157].

For the studies of CMDs, integrative analysis across multi-omics layers has been underway for more than a decade [158]. However, most of these studies just utilize the multi-omics data for cross-platform mutual validation on a specific gene set, and do not take full advantage of the information from multiple layers. Therefore, Jin et. al integrated bulk transcriptomic, proteomic, and peptidomics data derived from plaques and compared the difference between low-versus high-risk atherosclerotic lesion segments, revealing suppression of a regulatory network driven by serum response factor controls lesion stability in vulnerable plaque [159] using Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) [160].

Due to the success of spatial transcriptomics technologies, multi-omics data analysis in single-cell spatial resolution will be the next wave to explore spatial heterogeneity of tissues and to understand the underlying pathogenesis of many complex diseases such as CMDs [161, 162]. The disparities in spatial resolutions across various omics measurement platforms, in addition to the existing issues in traditional multi-omics analysis, such as heterogeneous data modalities and the curse of dimensionality, pose a more significant challenge to spatial integration. For instance, the spatial resolution of the Nuance multispectral system may reach up to 0.2  $\mu\text{m}$  [136], whereas that of MALDI-MSI is limited to about 1-5  $\mu\text{m}$  (albeit at very low speed) [163], let alone the 50- $\mu\text{m}$ -resolution of 'Visium' [164]. To solve this issue, Guilliams *et al.* measured single-cell CITE-seq, snRNA-Seq, spatial transcriptomics, and spatial proteomics separately from mice liver, and transferred the phenotypic information that were obtained from scRNA-Seq to spatial

transcriptomics and proteomics, for differentiation and localization of macrophages in liver sections[165]. However, this does not enable identification of multi-omics profiles at the same location of the tissue. Some other complex and expensive technologies, such as DBiT-seq [163] and Nanostring CosMx [166] make capturing or inferring the spatial multi-omics profile simultaneously possible, but they can only measure the proteome and transcriptome simultaneously. As metabolites play pivotal roles in the progress of CMDs, the applications of these approaches in CMDs are restricted. Therefore, integrating metabolomics data with other omics data at the single-cell resolution remains a daunting challenge.

### **From single time point to time series: studying Molecular trajectories in a spatial context**

Steady-state omics analysis at a single time point does not allow to study how biological pathways and genetic interactions change over time during the disease course. As the cost of high-throughput technology is steadily decreasing, also temporal omics analysis comes within reach enabling to track temporal biological processes [167].

The goal of the time-series analysis is to identify the characteristics or feature groups (e.g., genes, proteins, metabolites) that vary over time. A common solution is to identify features that are significantly different between time points (also called time-series differential expression analysis) [168]. For example, Nasias *et al.* applied maSigPro [169], a two-step regression approach, to select differential expressed genes in murine adipose tissue at 4, 8 and 12 weeks of high-fat feeding, to monitor changes in the transcriptome of the epididymal white adipose tissue during the development of metabolic syndrome. However, the majority of time-series tools are designed for transcriptomics [168], and some of them are only suited for data with a negative binomial distribution (i.e., RNA-Seq), which reduces their universality [170, 171].

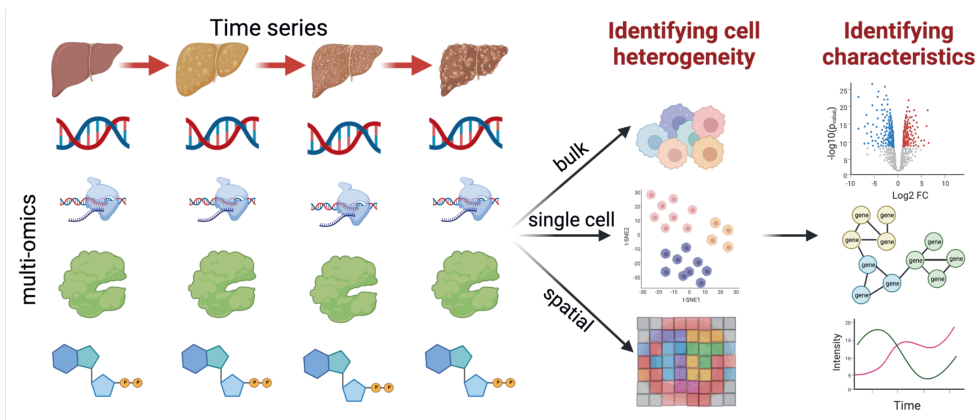
In addition, clustering genes with comparable time-series expression patterns allows for the identification of time-related gene groups and the interactions between genes[172]. Compared to other omics analyses, the application of time-series clustering in CMD-related studies is relatively lagging behind. Most longitudinal studies on atherosclerosis still use general-purpose heuristic algorithms such as k-means, hierarchical clustering and fuzzy c-means [173, 174]. For instance, Hou *et al.* identified 12 temporal gene expression patterns during the Goto-Kakizaki diabetes progression of rats, by clustering different expression genes and proteins using k-means [175]. Deshpande *et al.* found that the gene modules that are related to ER stress, inflammation and neurotransmission were differentially expressed at the early developmental stage, by constructing numerous WGCNA networks based on the time-series gene expressions from

Apobtm2Sgy and Ldlrtm1Her double knockout mice [176]. However, these traditional clustering methods do not suit time-series data because they ignore temporal dependencies and nonlinear time sampling. Recently, several models based on Dirichlet process (DP) and Gaussian Process (GP) were proposed for time-series gene expression clustering [177–179]. Compared to the clustering methods based on distance, GP can model time dependence, and DP is able to address the problem of uncertainty in the number of clusters [179].

In **chapter 3**, I apply time-series clustering on MALDI-MSI data from the livers of *Ldlr*<sup>-/-</sup> mice fed with 0-to-3-week high-fat diet, to identify the disease-process-related metabolic clusters. As mentioned previously, a liver has a markedly heterogeneous spatial structure that may be involved in different metabolic functions [64]. Therefore, analysis of metabolic changes within different liver compartments will provide a more accurate understanding of metabolic disfunction during NAFLD, yet this issue has been little studied. Given the above, how to apply time-series algorithms on MALDI-MSI data, especially to exploit both temporal and spatial dependencies to dissect the spatio-temporal heterogeneity of cells during progression of NAFLD, is the key question to be investigated in **chapter 3**.

## Scope of the thesis

To summarise the above, the rapid development of high-throughput technologies opens up unlimited possibilities for exploring the pathology of complex diseases. Specifically multi-omics data describe samples from different views, bridging the incompleteness of information between layers [153]. Single-cell strategies distinguish cell phenotypes and differentiation trajectories that cannot be recognised using bulk analysis. The addition of spatial information can depict the spatial relationships between cells in heterogeneous tissues, offering the possibility to explore cellular interactions. The study of heterogeneous samples, cells, or pixels allows us to find the most representative features, and the relationship between those features. Finally, the dynamic changes of these biosignatures during a disease progression can be depicted using time series omics profiles (**Fig 5**). It is foreseen that the applications of these advanced technologies in cardiometabolic research will lead to major breakthroughs in the analysis of underlying molecular mechanisms in disease progression and clinical manifestations.



**Figure 5: A framework of bulk, single-cell, spatial, and time-series multi-omics applications in CMD research.** The figure was drawn using BioRender (<https://biorender.com/>)

However, due to the technical limitations described above, the application of these methods in CMD remains limited. Therefore, in this thesis, I design and deploy computational pipelines to dissect the heterogeneity and microenvironment of macrophages in atherosclerotic plaques, as well as the changes of metabolites with NAFLD progression in different liver compartments. In addition, I apply a range of transcriptome analysis approaches to infer the transcriptional and functional profiles of monocytes and macrophages associated with CVD, which are central players in these diseases. I will briefly describe the main points of each chapter in the subsequent paragraphs.

Macrophages can adapt to stimuli in their immediate environment, adopting a functional phenotype profile throughout the course of atherosclerosis, suggesting

significant heterogeneity of plaque macrophages. Therefore, there is an urgent need to develop a sophisticated method to profile cell heterogeneity and microenvironment. Although traditional single-cell omics techniques have been used to identify cell heterogeneity, they are unable to preserve the location of cells. Some imaging-based spatial proteomics techniques provide great resolution for the localization of each cell, yet are constrained by limited biomarkers to identify cell types or by expensive experimental instrumentation. Furthermore, considering the important role of metabolites played in atherosclerosis, linking metabolomic features from MALDI-MSI to analyse the molecular context of cells is also a challenge. Therefore, in **chapter 2**, I design an extensible and flexible pipeline to integrate *in situ* multiplex immunofluorescent imaging and MALDI-MSI at single-cell resolution, to map tissue myeloid heterogeneity in its molecular and cellular context, and apply this on murine atherosclerotic plaque samples.

Cells in distinct liver compartments serve different roles [64, 180]; however, it is unknown if these differences are reflected level of metabolism in the development of NAFLD. As a result, in **chapter 3**, we analyse temporal patterns of metabolites in the context of spatial heterogeneity in liver tissues from NAFLD mice. To that goal, we deploy a computational pipeline capable of distinguishing disease-stage-independent liver compartments based on MALDI-MSI data and identifying disease stage-specific metabolites in various compartments during the progression of NAFLD.

In **chapter 4** and **5**, we aim to investigate the impact of CVD risk factors on monocyte phenotypes by applying transcriptomic analysis approaches on a cohort including 460 microarrays from monocytes of CVD and healthy subjects, collected by CTMM, as well as biological and lifestyle factors, that can enhance CVD risk. Specifically, **chapter 4** compares the gene expression profile of monocytes from male and female CVD patients to identify sex specific signalling pathways contributing to the disease. **Chapter 5** constructs a WGCNA network and a GRN to dissect the influence of CVD risk factors on monocytes after trauma and infection response.

**Chapter 6** discusses the findings presented this thesis as well as the advantages and limitations of the proposed methods. **Chapter 7** summarizes the main findings and future prospects of this thesis. Finally, **chapter 8** describes the potential scientific and societal impacts of this thesis.

## References

1. de Waard A-KM, Hollander M, Korevaar JC, Nielen MMJ, Carlsson AC, Lionis C, et al. Selective prevention of cardiometabolic diseases: activities and attitudes of general practitioners across Europe. *Eur J Public Health*. 2019;29:88–93.
2. Cardiovascular diseases (CVDs). 2021. [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).
3. Springer SC, Silverstein J, Copeland K, Moore KR, Prazar GE, Raymer T, et al. Management of Type 2 Diabetes Mellitus in Children and Adolescents. *Pediatrics*. 2013;131:e648–64.
4. Han JC, Lawlor DA, Kimm SY. Childhood obesity. *The Lancet*. 2010;375:1737–48.
5. Malik VS, Willett WC, Hu FB. Global obesity: trends, risk factors and policy implications. *Nat Rev Endocrinol*. 2013;9:13–27.
6. Kuklina EV, Tong X, George MG, Bansil P. Epidemiology and prevention of stroke: a worldwide perspective. *Expert Rev Neurother*. 2012;12:199–208.
7. Cardiovascular Disease 2020-2030: Trends, Technologies & Outlook. 2019.
8. Interpretation of Annual Report on Cardiovascular Health and Diseases in China 2019. *Cardiol Discov*. 2021;1:269–84.
10. Chatterjee A, Harris SB, Leiter LA, Fitchett DH, Teoh H, Bhattacharyya OK. Managing cardiometabolic risk in primary care. *Can Fam Physician*. 2012;58:389–93.
11. Rodgers JL, Jones J, Bolleddu SI, Vanthenapalli S, Rodgers LE, Shah K, et al. Cardiovascular Risks Associated with Gender and Aging. *J Cardiovasc Dev Dis*. 2019;6:19.
12. Dhingra R, Vasan RS. Age as a Cardiovascular Risk Factor. *Med Clin North Am*. 2012;96:87–91.
13. Maas AHEM, Appelman YEA. Gender differences in coronary heart disease. *Neth Heart J*. 2010;18:598–602.
14. Aksentijevich M, Lateef SS, Anzenberg P, Dey AK, Mehta NN. Chronic Inflammation, Cardiometabolic Diseases and Effects of Treatment: Psoriasis as a Human Model. *Trends Cardiovasc Med*. 2020;30:472–8.
15. Insull W. The Pathology of Atherosclerosis: Plaque Development and Plaque Responses to Medical Treatment. *Am J Med*. 2009;122:S3–14.
16. Sprague AH, Khalil RA. Inflammatory Cytokines in Vascular Dysfunction and Vascular Disease. *Biochem Pharmacol*. 2009;78:539–52.

17. Shao Y, Saredy J, Yang WY, Sun Y, Lu Y, Saaoud F, et al. Vascular Endothelial Cells and Innate Immunity. *Arterioscler Thromb Vasc Biol.* 2020;40:e138–52.
18. Arango Duque G, Descoteaux A. Macrophage Cytokines: Involvement in Immunity and Infectious Diseases. *Front Immunol.* 2014;5.
19. Grundtman C, Kreutmayer SB, Almanzar G, Wick MC, Wick G. Heat Shock Protein 60 and Immune Inflammatory Responses in Atherosclerosis. *Arterioscler Thromb Vasc Biol.* 2011;31:960–8.
20. Wick G, Jakic B, Buszko M, Wick MC, Grundtman C. The role of heat shock proteins in atherosclerosis. *Nat Rev Cardiol.* 2014;11:516–29.
21. Soto Y, Conde H, Aroche R, Brito V, Luaces P, Nasiff A, et al. Autoantibodies to oxidized low density lipoprotein in relation with coronary artery disease. *Hum Antibodies.* 2009;18:109–17.
22. Puurunen M, Mänttari M, Manninen V, Tenkanen L, Alfthan G, Ehnholm C, et al. Antibody Against Oxidized Low-Density Lipoprotein Predicting Myocardial Infarction. *Arch Intern Med.* 1994;154:2605–9.
23. Alavi A, Werner TJ, Raynor W, Høilund-Carlsen PF, Revheim M-E. Critical review of PET imaging for detection and characterization of the atherosclerotic plaques with emphasis on limitations of FDG-PET compared to NaF-PET in this setting. *Am J Nucl Med Mol Imaging.* 2021;11:337–51.
24. Moore KJ, Sheedy FJ, Fisher EA. Macrophages in atherosclerosis: a dynamic balance. *Nat Rev Immunol.* 2013;13:709–21.
25. Flynn MC, Pernes G, Lee MKS, Nagareddy PR, Murphy AJ. Monocytes, Macrophages, and Metabolic Disease in Atherosclerosis. *Front Pharmacol.* 2019;10.
26. *Frontiers | Human Monocyte Subsets and Phenotypes in Major Chronic Inflammatory Diseases.* <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02035/full>. Accessed 12 Jul 2022.
27. Ziegler-Heitbrock L, Ancuta P, Crowe S, Dalod M, Grau V, Hart DN, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood.* 2010;116:e74–80.
28. Ancuta P, Wang J, Gabuzda D. CD16+ monocytes produce IL-6, CCL2, and matrix metalloproteinase-9 upon interaction with CX3CL1-expressing endothelial cells. *J Leukoc Biol.* 2006;80:1156–64.
29. Ziegler-Heitbrock L. The CD14+ CD16+ blood monocytes: their role in infection and inflammation. *J Leukoc Biol.* 2007;81:584–92.



30. Cros J, Cagnard N, Woollard K, Patey N, Zhang S-Y, Senechal B, et al. Human CD14<sup>dim</sup> Monocytes Patrol and Sense Nucleic Acids and Viruses via TLR7 and TLR8 Receptors. *Immunity*. 2010;33:375–86.
31. Zhou J, Feng G, Beeson J, Hogarth PM, Rogerson SJ, Yan Y, et al. CD14<sup>hi</sup>CD16<sup>+</sup> monocytes phagocytose antibody-opsonised *Plasmodium falciparum* infected erythrocytes more efficiently than other monocyte subsets, and require CD16 and complement to do so. *BMC Med*. 2015;13:154.
32. Chilunda V, Martinez-Aguado P, Xia LC, Cheney L, Murphy A, Veksler V, et al. Transcriptional Changes in CD16<sup>+</sup> Monocytes May Contribute to the Pathogenesis of COVID-19. *Front Immunol*. 2021;12.
33. Hoefler IE, Sels JW, Jukema JW, Bergheanu S, Biessen E, McClellan E, et al. Circulating cells as predictors of secondary manifestations of cardiovascular disease: design of the CIRCULATING CELLS study. *Clin Res Cardiol*. 2013;102:847–56.
34. Hansson GK, Libby P. The immune response in atherosclerosis: a double-edged sword. *Nat Rev Immunol*. 2006;6:508–19.
35. Hilgendorf I, Swirski FK, Robbins CS. Monocyte Fate in Atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2015;35:272–9.
36. Barrett TJ. Macrophages in Atherosclerosis Regression. *Arterioscler Thromb Vasc Biol*. 2020;40:20–33.
37. Sica A, Mantovani A. Macrophage plasticity and polarization: in vivo veritas. *J Clin Invest*. 2012;122:787–95.
38. Mills C. M1 and M2 Macrophages: Oracles of Health and Disease. *Crit Rev Immunol*. 2012;32.
39. Orihuela R, McPherson CA, Harry GJ. Microglial M1/M2 polarization and metabolic states. *Br J Pharmacol*. 2016;173:649–65.
40. Mantovani A, Sica A, Sozzani S, Allavena P, Vecchi A, Locati M. The chemokine system in diverse forms of macrophage activation and polarization. *Trends Immunol*. 2004;25:677–86.
41. Murray PJ, Allen JE, Biswas SK, Fisher EA, Gilroy DW, Goerdts S, et al. Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity*. 2014;41:14–20.
42. Stein M, Keshav S, Harris N, Gordon S. Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. *J Exp Med*. 1992;176:287–92.

43. Goerdt S, Orfanos CE. Other Functions, Other Genes: Alternative Activation of Antigen-Presenting Cells. *Immunity*. 1999;10:137–42.
44. Mosser DM, Edwards JP. Exploring the full spectrum of macrophage activation. *Nat Rev Immunol*. 2008;8:958–69.
45. Mantovani A, Sozzani S, Locati M, Allavena P, Sica A. Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends Immunol*. 2002;23:549–55.
46. Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity*. 2014;40:274–88.
47. Kadl A, Meher AK, Sharma PR, Lee MY, Doran AC, Johnstone SR, et al. Identification of a novel macrophage phenotype that develops in response to atherogenic phospholipids via Nrf2. *Circ Res*. 2010;107:737–46.
48. Serbulea V, Upchurch CM, Ahern KW, Bories G, Voigt P, DeWeese DE, et al. Macrophages sensing oxidized DAMPs reprogram their metabolism to support redox homeostasis and inflammation through a TLR2-Syk-ceramide dependent mechanism. *Mol Metab*. 2018;7:23–34.
49. Colin S, Chinetti-Gbaguidi G, Staels B. Macrophage phenotypes in atherosclerosis. *Immunol Rev*. 2014;262:153–66.
50. Nagenborg J, Goossens P, Biessen EAL, Donners MMPC. Heterogeneity of atherosclerotic plaque macrophage origin, phenotype and functions: Implications for treatment. *Eur J Pharmacol*. 2017;816:14–24.
51. Iyer S, Upadhyay PK, Majumdar SS, Nagarajan P. Animal Models Correlating Immune Cells for the Development of NAFLD/NASH. *J Clin Exp Hepatol*. 2015;5:239–45.
52. Paschos P, Paletas K. Non alcoholic fatty liver disease two-hit process: multifactorial character of the second hit. *Hippokratia*. 2009;13:128.
53. Fernando DH, Forbes JM, Angus PW, Herath CB. The Receptor for Advanced Glycation End Products (RAGE): A Potential Target for Therapeutic Intervention in the Progression of Non-Alcoholic Fatty Liver Disease (NAFLD). In: *Prime Archives in Molecular Biology*. Vide Leaf, Hyderabad; 2020.
54. Bessone F, Razori MV, Roma MG. Molecular pathways of nonalcoholic fatty liver disease development and progression. *Cell Mol Life Sci*. 2019;76:99–128.

55. Ikeda N, Torii R. When does atherosclerosis become irreversible? Chronological change from an early to an advanced atherosclerotic lesion observed by angioscopy. *Angiology*. 2005;56:361–70.
57. Samuel VT, Shulman GI. Nonalcoholic Fatty Liver Disease as a Nexus of Metabolic and Hepatic Diseases. *Cell Metab*. 2018;27:22–41.
58. Lu Q, Tian X, Wu H, Huang J, Li M, Mei Z, et al. Metabolic Changes of Hepatocytes in NAFLD. *Front Physiol*. 2021;12.
59. Herman MA, Samuel VT. The Sweet Path to Metabolic Demise: Fructose and Lipid Synthesis. *Trends Endocrinol Metab*. 2016;27:719–30.
60. Alves-Bezerra M, Cohen DE. Triglyceride metabolism in the liver. *Compr Physiol*. 2017;8:1–8.
61. Santos-Baez LS, Ginsberg HN. Non-alcohol fatty liver disease: Balancing supply and utilization of triglycerides. *Curr Opin Lipidol*. 2021;32:200–6.
62. Jungermann K, Katz N. Functional hepatocellular heterogeneity. *Hepatology*. 1982;2:385–95.
63. Alpini G, Phillips JO, Vroman B, Larusso NF. Recent advances in the isolation of liver cells. *Hepatology*. 1994;20:494–514.
64. Bouwens L, De Bleser P, Vanderkerken K, Geerts B, Wisse E. Liver cell heterogeneity: functions of non-parenchymal cells. *Enzyme*. 1992;46:155–68.
65. Levy SE, Boone BE. Next-Generation Sequencing Strategies. *Cold Spring Harb Perspect Med*. 2019;9:a025791.
66. Li L, Han J, Wang Z, Liu J, Wei J, Xiong S, et al. Mass Spectrometry Methodology in Lipid Analysis. *Int J Mol Sci*. 2014;15:10492–507.
67. Alexandrov T. Spatial Metabolomics and Imaging Mass Spectrometry in the Age of Artificial Intelligence. *Annu Rev Biomed Data Sci*. 2020;3:61–87.
68. Sohag MMH, Raqib SM, Akhmad SA. OMICS approaches in cardiovascular diseases: a mini review. *Genomics Inform*. 2021;19.
69. Wirka RC, Pjanic M, Quertermous T. Advances in Transcriptomics. *Circ Res*. 2018;122:1200–20.
70. Ahmed Z, Zeeshan S, Liang BT. RNA-seq driven expression and enrichment analysis to investigate CVD genes with associated phenotypes among high-risk heart failure patients. *Hum Genomics*. 2021;15:67.

71. Zhang L, Liu M, Li Q, Shen B, Hu C, Fu R, et al. Identification of differential gene expression in endothelial cells from young and aged mice using RNA-Seq technique. *Am J Transl Res.* 2019;11:6553–60.
72. Hegenbarth J-C, Lezsoche G, De Windt LJ, Stoll M. Perspectives on Bulk-Tissue RNA Sequencing and Single-Cell RNA Sequencing for Cardiac Transcriptomics. *Front Mol Med.* 2022;2.
73. Bumgarner R. Overview of DNA microarrays: types, applications, and their future. *Curr Protoc Mol Biol.* 2013;Chapter 22:Unit 22.1.
74. Archacki SR, Wang QK. Microarray analysis of cardiovascular diseases. *Methods Mol Med.* 2006;129:1–13.
75. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS ONE.* 2014;9:e78644.
76. Kaliyappan K, Palanisamy M, Govindarajan R, Duraiyan J. Microarray and its applications. *J Pharm Bioallied Sci.* 2012;4:310.
77. Eilertsen IA, Moosavi SH, Strømme JM, Nesbakken A, Johannessen B, Lothe RA, et al. Technical differences between sequencing and microarray platforms impact transcriptomic subtyping of colorectal cancer. *Cancer Lett.* 2020;469:246–55.
78. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 2016;17:13.
79. Pedrotty DM, Morley MP, Cappola TP. Transcriptomic Biomarkers of Cardiovascular Disease. *Prog Cardiovasc Dis.* 2012;55:64–9.
80. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10:57–63.
81. Rai MF, Tycksen ED, Sandell LJ, Brophy RH. Advantages of RNA-seq Compared to RNA Microarrays for Transcriptome Profiling of Anterior Cruciate Ligament Tears. *J Orthop Res Off Publ Orthop Res Soc.* 2018;36:484–97.
82. List of RNA-Seq bioinformatics tools. Wikipedia. 2022.
83. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nat Rev Methods Primer.* 2021;1:59.
84. Tarca AL, Romero R, Draghici S. Analysis of microarray experiments of gene expression profiling. *Am J Obstet Gynecol.* 2006;195:373–88.

85. Koschmieder A, Zimmermann K, Trißl S, Stoltmann T, Leser U. Tools for managing and analyzing microarray data. *Brief Bioinform.* 2012;13:46–60.
86. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
87. van Dam S, Vösa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief Bioinform.* 2018;19:575–92.
88. Russo A, Biselli-Chicote PM, Kawasaki-Oyama RS, Castanhole-Nunes MMU, Maniglia JV, Santi Neto D de, et al. Differential Expression of Prostaglandin I2 Synthase Associated with Arachidonic Acid Pathway in the Oral Squamous Cell Carcinoma. *J Oncol.* 2018;2018:e6301980.
89. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
90. Rau C, Wisniewski N, Orozco L, Bennett B, Weiss J, Lusi A. Maximal information component analysis: a novel non-linear network analysis method. *Front Genet.* 2013;4.
91. Delgado FM, Gómez-Vela F. Computational methods for Gene Regulatory Networks reconstruction and analysis: A review. *Artif Intell Med.* 2019;95:133–45.
92. Meyer PE, Kontos K, Lafitte F, Bontempi G. Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J Bioinforma Syst Biol.* 2007;:79879.
93. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, et al. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics.* 2006;7:S7.
94. Maetschke SR, Madhamshettiwar PB, Davis MJ, Ragan MA. Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Brief Bioinform.* 2014;15:195–211.
95. Maleki F, Ovens K, Hogan DJ, Kusalik AJ. Gene Set Analysis: Challenges, Opportunities, and Future Research. *Front Genet.* 2020;11.
96. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci.* 2005;102:15545–50.

97. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171:1437-1452.e17.
98. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res*. 2019;29:1363–75.
99. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1–13.
100. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet*. 2016;48:838–47.
101. Jiang P, Zhang Y, Ru B, Yang Y, Vu T, Paul R, et al. Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat Methods*. 2021;18:1181–91.
102. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25–9.
103. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44:D457–62.
104. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al. WikiPathways: connecting communities. *Nucleic Acids Res*. 2021;49:D613–21.
105. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*. 2021;2:100141.
106. Holland CH, Tanevski J, Perales-Patón J, Gleixner J, Kumar MP, Mereu E, et al. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol*. 2020;21:36.
107. Ma CZ, Brent MR. Inferring TF activities and activity regulators from gene expression data with constraints from TF perturbation data. *Bioinformatics*. 2021;37:1234–45.
108. Kulakovskiy IV, Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res*. 2018;46:D252–9.

109. Jin H, Mees BME, Biessen EAL, Sluimer JC. Transcriptional Sex Dimorphism in Human Atherosclerosis Relates to Plaque Type. *Circ Res.* 2021;129:1175–7.
110. Vignoli A, Tenori L, Luchinat C, Saccenti E. Age and Sex Effects on Plasma Metabolite Association Networks in Healthy Subjects. *J Proteome Res.* 2018;17:97–107.
111. Nath M, Romaine SPR, Koekemoer A, Hamby S, Webb TR, Nelson CP, et al. Whole blood transcriptomic profiling identifies molecular pathways related to cardiovascular mortality in heart failure. *Eur J Heart Fail.* 2022;24:1009–19.
112. Jaakkola MK, Elo LL. Computational deconvolution to estimate cell type-specific gene expression from bulk data. *NAR Genomics Bioinforma.* 2021;3:lqaa110.
113. Li Y, Ma L, Wu D, Chen G. Advances in bulk and single-cell multi-omics approaches for systems biology and precision medicine. *Brief Bioinform.* 2021;22:bbab024.
114. Hill CA, Fernandez DM, Giannarelli C. Single cell analyses to understand the immune continuum in atherosclerosis. *Atherosclerosis.* 2021;330:85–94.
115. Roth R, Kim S, Kim J, Rhee S. Single-cell and spatial transcriptomics approaches of cardiovascular development and disease. *BMB Rep.* 2020;53:393–9.
116. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med.* 2018;50:1–14.
117. McKinnon KM. Flow Cytometry: An Overview. *Curr Protoc Immunol.* 2018;120.
118. Spitzer MH, Nolan GP. Mass Cytometry: Single Cells, Many Features. *Cell.* 2016;165:780–91.
119. Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat Rev Genet.* 2019;20:273–82.
120. Wang B, Zhu J, Pierson E, Ramazzotti D, Batzoglou S. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat Methods.* 2017;14:414–6.
121. Guo M, Wang H, Potter SS, Whitsett JA, Xu Y. SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis. *PLoS Comput Biol.* 2015;11:e1004575.
122. Lin P, Troup M, Ho JWK. CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol.* 2017;18:59.

123. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp.* 2008;2008:P10008.
124. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19:15.
125. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184:3573-3587.e29.
126. Jarvis RA, Patrick EA. Clustering Using a Similarity Measure Based on Shared Near Neighbors. *IEEE Trans Comput.* 1973;22:1025–34.
127. Maaten LJP van der, Postma EO, Herik HJ van den. Dimensionality Reduction: A Comparative Review. 2008.
128. Principal Component Analysis. New York: Springer-Verlag; 2002.
129. Maaten L van der, Hinton G. Visualizing Data using t-SNE. *J Mach Learn Res.* 2008;9:2579–605.
130. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2018. <https://doi.org/10.48550/ARXIV.1802.03426>.
131. Zernecke A, Winkels H, Cochain C, Williams JW, Wolf D, Soehnlein O, et al. Meta-Analysis of Leukocyte Diversity in Atherosclerotic Mouse Aortas. *Circ Res.* 2020;127:402–26.
132. Cochain C, Vafadarnejad E, Arampatzi P, Pelisek J, Winkels H, Ley K, et al. Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis. *Circ Res.* 2018;122:1661–74.
133. Winkels H, Ehinger E, Vassallo M, Buscher K, Dinh HQ, Kobiyama K, et al. Atlas of the Immune Cell Repertoire in Mouse Atherosclerosis Defined by Single-Cell RNA-Sequencing and Mass Cytometry. *Circ Res.* 2018;122:1675–88.
134. Guilliams M, Thierry GR, Bonnardel J, Bajenoff M. Establishment and Maintenance of the Macrophage Niche. *Immunity.* 2020;52:434–51.
135. Ramos-Vara JA, Miller MA. When tissue antigens and antibodies get along: revisiting the technical aspects of immunohistochemistry--the red, brown, and blue technique. *Vet Pathol.* 2014;51:42–87.
136. Nuance. Research. <https://research.pathology.wisc.edu/nuance/>. Accessed 15 Jul 2022.



137. Chang Q, Ornatsky OI, Siddiqui I, Loboda A, Baranov VI, Hedley DW. Imaging Mass Cytometry. *Cytometry A*. 2017;91:160–9.

138. McDonnell LA, Heeren RMA. Imaging mass spectrometry. *Mass Spectrom Rev*. 2007;26:606–43.

139. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. 2016;353:78–82.

140. Villacampa EG, Larsson L, Kvastad L, Andersson A, Carlson J, Lundeberg J. Genome-wide Spatial Expression Profiling in FFPE Tissues. 2020;:2020.07.24.219758.

141. Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball patterned arrays. 2021;:2021.01.17.427004.

142. Zhang W, Claesen M, Moerman T, Groseclose MR, Waelkens E, De Moor B, et al. Spatially aware clustering of ion images in mass spectrometry imaging data using deep learning. *Anal Bioanal Chem*. 2021;413:2803–19.

143. Bemis KD, Eberlin L, Ferreira C, Cooks RG, Vitek O. Spatial segmentation and feature selection for desi imaging mass spectrometry data with spatially-aware sparse clustering. *BMC Bioinformatics*. 2012;13:A8.

144. Alexandrov T, Kobarg JH. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics*. 2011;27:i230–8.

145. Bemis KD, Harry A, Eberlin LS, Ferreira CR, Ven SM van de, Mallick P, et al. Probabilistic Segmentation of Mass Spectrometry (MS) Images Helps Select Important Ions and Characterize Confidence in the Resulting Segments \*. *Mol Cell Proteomics*. 2016;15:1761–72.

146. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, et al. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell*. 2018;174:968-981.e15.

147. Schüffler PJ, Schapiro D, Giesen C, Wang HAO, Bodenmiller B, Buhmann JM. Automatic single cell segmentation on highly multiplexed tissue images. *Cytometry A*. 2015;87:936–42.

148. Schapiro D, Jackson HW, Raghuraman S, Fischer JR, Zanutelli VRT, Schulz D, et al. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat Methods*. 2017;14:873–6.

149. Gu W, Ni Z, Tan Y-Q, Deng J, Zhang S-J, Lv Z-C, et al. Adventitial Cell Atlas of wt (Wild Type) and ApoE (Apolipoprotein E)-Deficient Mice Defined by Single-Cell RNA Sequencing. *Arterioscler Thromb Vasc Biol.* 2019;39:1055–71.
150. Alsaigh T, Evans D, Frankel D, Torkamani A. Decoding the transcriptome of atherosclerotic plaque at single-cell resolution. preprint. *Genomics*; 2020.
151. Wu M, Lee MYY, Bahl V, Traum D, Schug J, Kusmartseva I, et al. Single-cell analysis of the human pancreas in type 2 diabetes using multi-spectral imaging mass cytometry. *Cell Rep.* 2021;37:109919.
152. Saldarriaga OA, Freiberg B, Krishnan S, Rao A, Burks J, Booth AL, et al. Multispectral Imaging Enables Characterization of Intrahepatic Macrophages in Patients With Chronic Liver Disease. *Hepatol Commun.* 2020;4:708–23.
153. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinforma Biol Insights.* 2020;14:1177932219899051.
154. Santiago-Rodriguez TM, Hollister EB. Multi 'omic data integration: A review of concepts, considerations, and approaches. *Semin Perinatol.* 2021;45:151456.
155. Wörheide MA, Krumsiek J, Kastenmüller G, Arnold M. Multi-omics integration in biomedical research – A metabolomics-centric review. *Anal Chim Acta.* 2021;1141:144–62.
156. Fiehn O. Metabolomics — the link between genotypes and phenotypes. In: Town C, editor. *Functional Genomics.* Dordrecht: Springer Netherlands; 2002. p. 155–71.
157. Huang S, Chaudhary K, Garmire LX. More Is Better: Recent Progress in Multi-Omics Data Integration Methods. *Front Genet.* 2017;8.
158. Inouye M, Silander K, Hamalainen E, Salomaa V, Harald K, Jousilahti P, et al. An Immune Response Network Associated with Blood Lipid Levels. *PLOS Genet.* 2010;6:e1001113.
159. Jin H, Goossens P, Juhasz P, Eijgelaar W, Manca M, Karel JMH, et al. Integrative multiomics analysis of human atherosclerosis reveals a serum response factor-driven network associated with intraplaque hemorrhage. *Clin Transl Med.* 2021;11:e458.
160. Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinforma Oxf Engl.* 2019;35:3055–62.

161. Handley E. Next generation sequencing in tissues: Towards in situ multiomics. Open Access Government. 2022. <https://www.openaccessgovernment.org/next-generation-sequencing-tissues-spatial-transcriptomics/131639/>. Accessed 15 Jul 2022.
162. Ogbeide S, Giannese F, Mincarelli L, Macaulay IC. Into the multiverse: advances in single-cell multiomic profiling. *Trends Genet.* 2022;38:831–43.
163. Liu Y, Yang M, Deng Y, Su G, Enniful A, Guo CC, et al. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell.* 2020;183:1665–1681.e18.
164. What is the spatial resolution and configuration of the capture area of the Visium Gene Expression Slide? 10X Genomics. <https://kb.10xgenomics.com/hc/en-us/articles/360035487572-What-is-the-spatial-resolution-and-configuration-of-the-capture-area-of-the-Visium-Gene-Expression-Slide->. Accessed 15 Jul 2022.
165. Guilliams M, Bonnardel J, Haest B, Vanderborgh B, Wagner C, Remmerie A, et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell.* 2022;185:379–396.e38.
166. Spatial Multiomics at Any Scale. NanoString. <https://nanosttring.com/research-focus/spatial-multiomics/>. Accessed 23 Aug 2022.
167. Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. *Nat Rev Cardiol.* 2021;18:313–30.
168. Spies D, Renz PF, Beyer TA, Ciaudo C. Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Brief Bioinform.* 2019;20:288–98.
169. Conesa A, Nueda MJ, Ferrer A, Talón M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinforma Oxf Engl.* 2006;22:1096–102.
170. Äijö T, Butty V, Chen Z, Salo V, Tripathi S, Burge CB, et al. Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation. *Bioinformatics.* 2014;30:i113–20.
171. Leng N, Li Y, McIntosh BE, Nguyen BK, Duffin B, Tian S, et al. EBSeq-HMM: a Bayesian approach for identifying gene-expression changes in ordered RNA-seq experiments. *Bioinformatics.* 2015;31:2614–22.
172. Androulakis IP, Yang E, Almon RR. Analysis of Time-Series Gene Expression Data: Methods, Challenges, and Opportunities. *Annu Rev Biomed Eng.* 2007;9:205–28.

173. Magni P, Ferrazzi F, Sacchi L, Bellazzi R. TimeClust: a clustering tool for gene expression time series. *Bioinformatics*. 2008;24:430–2.
174. Kumar L, E. Futschik M. Mfuzz: A software package for soft clustering of microarray data. *Bioinformatics*. 2007;2:5–7.
175. Hou J, Li Z, Zhong W, Hao Q, Lei L, Wang L, et al. Temporal Transcriptomic and Proteomic Landscapes of Deteriorating Pancreatic Islets in Type 2 Diabetic Rats. *Diabetes*. 2017;66:2188–200.
176. Deshpande V, Sharma A, Mukhopadhyay R, Thota LNR, Ghatge M, Vangala RK, et al. Understanding the progression of atherosclerosis through gene profiling and co-expression network analysis in Apobtm2SgyLdlrtm1Her double knockout mice. *Genomics*. 2016;107:239–47.
177. Rasmussen CE, de la Cruz BJ, Ghahramani Z, Wild DL. Modeling and visualizing uncertainty in gene expression clusters using dirichlet process mixtures. *IEEE/ACM Trans Comput Biol Bioinform*. 2009;6:615–28.
178. Hensman J, Rattray M, Lawrence ND. Fast Nonparametric Clustering of Structured Time-Series. *IEEE Trans Pattern Anal Mach Intell*. 2015;37:383–93.
179. McDowell IC, Manandhar D, Vockley CM, Schmid AK, Reddy TE, Engelhardt BE. Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLOS Comput Biol*. 2018;14:e1005896.



# Chapter 2

Integrating multiplex immunofluorescent and mass spectrometry imaging to map tissue myeloid heterogeneity in its metabolic and cellular context

Pieter Goossen<sup>†</sup>, **Chang Lu**<sup>†</sup>, Jianhua Cao, Marion Gijbels, Joël M.H. Karel, Erwin Wijnands, Britt S.R. Clae, Gregorio E. Fazzi, Tim F.E. Hendriks, Kristiaan Wouters, Evgueni Smirnov, Marc J.M. van Zandvoort, Benjamin Balluff, Eva Cuypers, Marjo M.P.C. Donners, Ron M.A. Heeren, Erik A.L. Biessen

*Cell Metabolism. 2022 Aug 2;34(8):1214-1225.e6*

<sup>†</sup> Authors contributed equally



## Abstract

Cells often adopt different phenotypes, dictated by tissue-specific or local signals such as cell-cell and cell-matrix contacts or molecular micro-environment. This holds in extremis for macrophages with their high phenotypic plasticity. Their broad range of functions, some even opposing, reflects their heterogeneity, and a multitude of subsets has been described in different tissues and diseases. Such micro-environmental imprint cannot be adequately studied by single-cell applications as cells are detached from their context, while histology-based assessment lacks the phenotypic depth due to limitations in marker combination. Here, we present a novel, integrative approach in which 15-color multispectral imaging allows comprehensive cell classification based on multi-marker expression patterns, followed by downstream analysis pipelines to link their phenotypes to contextual, micro-environmental cues such as their cellular (“community”) and metabolic (“local lipidome”) niches in complex tissue. The power of this approach is illustrated for myeloid subsets and associated lipid signatures in murine atherosclerotic plaque.



## Introduction

Mammals contain over 200 cell types, but even these can vary strongly in shape, activation state or function, depending on tissue or disease context. Probably the best studied example of this heterogeneity is the macrophage [1], a myeloid cell type with ubiquitous presence throughout the body whose phenotype not only differs between but even within tissues. Recent findings have demonstrated that, regardless of their ontogeny [2], these highly plastic cells adopt distinct phenotypes dictated and fostered by micro-environmental cues and stromal niches [3]. Studying such cross-interactions in tissue by traditional methods is however hampered by technological limitations that complicate the parallel monitoring of multiple cell types or phenotypes and their context.

Cytometric approaches and, increasingly, single-cell transcriptomics have greatly advanced our understanding of myeloid cell heterogeneity and have been instrumental in mapping subsets in several healthy or diseased tissues. Though powerful, they require prior tissue dissociation. This leads to underrepresentation of fragile, large or adherent cells and, even worse, to complete loss of spatial and contextual information. This is not trivial as tissues, particularly when diseased, rarely display an even cellular or molecular composition. Therefore, these approaches do neither allow assessment of the spatial distribution of subsets nor identification of cell-cell and -niche interactions or micro-environmental cues that dictate the macrophages' phenotype. Immunohistochemistry (IHC) or immunofluorescent microscopy (IF) do not suffer from this setback but fail to provide the required phenotypic resolution due to the limited number of markers that can simultaneously be imaged. Several multiplex imaging approaches have been developed over the past decade, each with their specific merits and flaws [4]. Manual and automated sequential IF staining approaches [5-7] provide illustrative images but suffer from lengthy acquisition and the risk of gradual signal loss with each cycle, disqualifying them for extensive screening purposes. Alternatively, one-shot multiplex imaging approaches, including but not limited to imaging mass cytometry [8], stimulated Raman scattering microscopy [9] or multiplexed ion beam imaging [10], produce impressive and figurative images but the high cost in terms of custom consumables and/or infrastructure limits widespread application. Other novel high-dimensional solutions such as SeqFISH+ [11] or Slide-Seq [12] will only cover limited amounts of cells.

To meet this need, we here propose an integrated pipeline combining a novel, multispectral IF imaging approach for rapid, affordable, comprehensive phenotypic and spatial mapping of cellular heterogeneity in tissue, combined with parallel mass spectrometry imaging (MSI [13]) to define the cells' molecular context.

## Methods

### Experimental models and subject details

All imaging was performed on a spleen of an adult male wild-type C57Bl/6 mouse (Charles River Laboratories, <https://www.criver.com/products-services/find-model/c57bl6-mouse>), and on aortic roots of eleven adult *Ldlr*<sup>-/-</sup> mice (Jackson Laboratory, <https://www.jax.org/strain/002207>) fed a Western-type diet (“diet W”, SDS Diets) for 10 weeks to induce atherogenesis.

Animal experiments were approved by a local Committee for Animal Welfare (IvD Maastricht University) and performed under the Maastricht University animal facility’s standard conditions, in accordance with Dutch national laws and regulations.

### Method details

#### Tissue preparation and staining

A spleen of a male *wild type* C57Bl/6 mouse was embedded in OCT compound and cryosectioned with a thickness of 7µm on standard adhesive microscopy slides (KliniPath). Eleven female *Ldlr*<sup>-/-</sup> mice were fed a Western-type diet (“diet W”, SDS Diets) for 10 weeks to induce atherogenesis and upon sacrifice, their aortic roots were snap-frozen in OCT compound. Serial cryosections of 7µm thickness were cut from the aortic valves up to the transverse aortic arch, as described earlier [14], and sections were mounted on a series of 23 standard adhesive microscopy slides for IHC and IF stainings and one indium tin oxide (ITO) coated slide (Delta Technologies) for mass spectrometry imaging, leaving 168µm between every section on one slide. All sections were dried in a desiccator for 5 hours at room temperature before storing them at -80°C.

Cryosections were thawed and air-dried, then fixed for five minutes with dry acetone and immersed in blocking buffer (PBS + 4% FCS) for overnight exposure to 400nm UV light in a cold room (4°C) (**Fig 1G-H**), prior to the staining procedure. All antibodies were mixed just prior to staining in PBS + 4% FCS and applied for four hours in the dark at 4°C on PAP Pen (ThermoFisher) lined tissue sections, followed by three washing steps (PBS) and mounting with the anti-fading Prolong Gold (ThermoFisher).

#### Multispectral imaging

Multiplex spectral imaging was performed using a Nuance FX camera (Perkin Elmer) mounted on an upright fluorescence microscope (Leica DM4000) equipped with a white light source, four different narrow band pass excitation filter cubes and an HC PLAN APO 20x/0.70 PH 2 (Air) objective (all Leica). Combined with the accompanying Nuance software, this setup allows to record the emitted

fluorescence intensity for each pixel in one field of view (FOV) in sequential pictures spanning the full visible electromagnetic range (420-720nm) while maintaining a constant, narrow-band filtered excitation wavelength. Filter cube properties and the Nuance software settings for spectral imaging are indicated in **Table S1**. In short, excitation with filter I3 was followed by N2.1 and the composite image was unmixed in its individual spectra, including the different dyes used in the antibody panel as well as separate background and autofluorescence spectra. Next, the exact same field of view was recorded sequentially with the A and S Blue Aqua filter excitation and again unmixed into separate spectra (**Fig 1A-F**). The total recording time for one 0.36mm<sup>2</sup> FOV corresponds to the addition of exposure times over the full spectra covered by all four filters (indicated in **Table S1**) and approaches six minutes.

The tissue's background, autofluorescence and the dye-specific spectra were first recorded in unstained and single label-stained sections of the same tissue type (murine spleen and aortic root, respectively) before combining all antibodies into one panel. Antibody concentrations were adjusted to obtain similar maximal fluorescence intensities for each of the markers and dye bleeding into overlapping spectra was assessed using fluorescence-minus-one (FMO) stained sections.

Five WT spleen cryosections, stained with the antibody panel in **Table S2**, were imaged, using the exact same procedure as used when defining the individual spectra on single-stained sections, and unmixed into their individual spectra. For visualization, brightness and contrast of each picture were manually rescaled using Fiji [15] to normalize signal intensities and remove subthreshold signal. A selection of the markers was combined into a representative merged picture.

A total of 148 FOVs of atherosclerotic plaques stained with the antibody panel in **Table S3**, spread evenly over the different aortic roots and ranging from proximal to distal from the valves, were imaged. Following spectral imaging, the slides were submerged upside down in PBS at 37°C at an angle that allowed the cover slip to spontaneously detach without damaging the underlying tissue. After additional washing steps, a haematoxylin/eosin (H&E) staining was performed on these slides and areas corresponding to the fluorescent fields of view (FOV) were imaged by brightfield microscopy (DM3000 LED with DFC320 camera and HC PL FLUOTAR 10x/0.30 (Air) objective, all Leica).

### Mass Spectrometry imaging

For MSI, tissue cryosections were captured on ITO glass slides that were cleaned by subsequent 10 minutes sonication steps in ethanol and hexane. The sections were dried under a vacuum desiccator for 10 minutes and followed by norharmane matrix deposition. Briefly, 15 layers of norharmane (7 mg/mL) in 2:1 chloroform:methanol (v:v) matrix solution were homogeneously coated onto the slides at 30°C

using a TM-sprayer (HTX Technologies, Chapel Hill, NC, USA) with a fixed flow rate of 120  $\mu\text{L}/\text{min}$ . Subsequently, samples were measured using a rapifleX MALDI Tissuetyper™ (Bruker Daltonik, Bremen, Germany) within a mass mass-to-charge ratio ( $m/z$ ) of 400–2000 in negative ion mode. Pixel size was  $15 \times 15 \mu\text{m}^2$  using 200 shots per pixel. The instrument was calibrated prior to each measurement using red phosphorus.

Next, the matrix was removed by dipping the slides in ethanol and a subsequent H&E staining was performed. Full slides were scanned using a digital slide scanner (Mirax Desk, Zeiss, Jena, Germany) and coregistered with the MALDI-MSI data in FlexImaging (Bruker Daltonics, Version 5.0) using 3 pre-marked reference points per slide.

### Image registration procedure

Three registration strategies were designed for linking every single cell segment's fluorescent imaging intensity and MSI intensity (**Fig S1**). Linear registration was used for aligning images from the same section, i.e. the alignment of multispectral images and their corresponding H&E images (hereinafter referred to as "fluorescent imaging layer", **Fig S1A**), and the alignment of H&E images with different resolutions from the sections analysed by MSI (hereinafter referred to as "MSI layer", **Fig S1C**). A combination of linear and nonlinear registration was applied to H&E images from adjacent sections (**Fig S1B**). All algorithms used for image registration are from Image Processing Toolbox (MathWorks[16]).

#### *Alignment of the H&E and multispectral fluorescent images*

The H&E image from the fluorescent imaging layer was aligned with the fluorescent images by enhancing and extracting the positional information of nuclei on both the H&E and the 7-AAD image (**Fig S1A**, (1) and (2)). In detail, for both the contrast was enhanced by transforming the values using contrast-limited adaptive histogram equalization (CLAHE) [17]. Wiener filter [18] and morphological operators were then used for removing the nuclei-unrelated noise. Next, binary cell nuclei images were obtained using a locally adaptive threshold [19] (**Fig S1A** (3) and (4)). The final registration transformation was obtained by aligning these two processed images using a Mattes Mutual Information (MMI) based registration algorithm[20] (**Fig S1A** (5)).

#### *Alignment of the MS image layer and the corresponding H&E section*

The low-resolution full-slide images of the H&E staining performed on the MSI tissues (Mirax Micro digital slide scanner), that were manually aligned with the MSI data in the FlexImaging software package based on landmarks (as described above), were now aligned with high-resolution images (VENTANA iScan HT slide scanner) (**Fig S1C** (1) and (2)). Features of both H&E images were detected and extracted using Oriented FAST and Rotated BRIEF (ORB) algorithm [21]. Corresponding

features were matched based on the pair-wise Sum of Squared Difference (SSD) between feature vectors from two H&E images (**Fig S1C** (3)). The final registration image was obtained by estimating geometric transforms from coordinate pairs of matching features (**Fig S1C** (4)) [22].

### *Coregistration between adjacent MS and multispectral imaging layer*

As MSI induced some tissue damage, the subsequently obtained H&E images were often of lower quality. To guarantee high-quality registration with the adjacent slides, only 25 out of 65 sections with no/minor damage were singled out for alignment. Coregistration between H&E-stained images from adjacent sections involved a two-step process. Rough registration was achieved by automated rigid registration based on Mattes Mutual Information (**Fig S1B** (1)) [20]. Subsequent precise alignment was performed by B-spline based elastic deformation (Fiji plugin, BUnwarpj) [23]. Hereto, landmarks were selected manually on both layers to improve and quantify registration precision (**Fig S1B** (2) to (4)). The average distances of landmarks before and after registration on each image are between 2.329 to 10.820 $\mu\text{m}$  (5.786 on average), hence lower than the resolution of MSI, illustrating the high coregistration precision.

## Computational procedures for identification and characterization of phenotypes

### *Fluorescence intensity matrix building*

A fluorescence intensity matrix was built to identify phenotypes by cell clustering. In brief, all cells were detected on H&E images and a matrix mapping each marker's fluorescent intensity per cell segment was constructed. For high-quality estimate of each cell segment's marker expression, this matrix was cleaned by spillover compensation and outlier removal (**Fig S2A**).

### *Cell segmentation and initial matrix building*

Cells were segmented on the H&E image from the fluorescent imaging layer, using a watershed cell detection tool from QuPath [24] (**Fig S2B** (1)). Nuclei were identified and expanded by 5 pixels. The thus created image mask containing all coordinates of the defined cell segments could be projected to the corresponding unmixed multispectral images (**Fig S2B** (2)).

Assuming  $n$  segmented cells were identified in a single H&E staining image, we first built a cell-biomarker matrix  $X \in \mathbb{R}^{n \times d}$  harboring the average intensity vectors for all fluorescence markers of each cell segment (**Fig S2B** (3) and (4)), where  $d$  is the number of biomarkers used for clustering analysis ( $d = 12$  in this study).

### Spillover compensation

Although cell segments identified through the watershed detection were kept relatively small, overlap with a neighboring cell could not be excluded. To correct for contamination of biomarker fluorescence from neighboring cells, a spillover compensation was applied on the cell-biomarker matrix, using a previously published protocol [5] with minor modifications. Firstly, neighbors of each cell were confirmed by a Delaunay triangulation graph (**Fig S2C (1)**). However, unlike Goltsev *et al.*, who constructed a spill coefficient matrix by calculating the percentage shared edge per cell segment [5], compensation was here based on the estimated overlap in area of the two cells (as illustrated in **Fig S2C (3)**). For each cell we defined its circumscribed ellipse, and calculated the overlapping area of a cell and its neighbor's ellipse (**Fig S2C (2)** and (4)). An adjacency matrix  $A \in \mathbb{R}^{n \times n}$  was defined as follows:

$$A_{ij} = \frac{S_{C_{ij}}}{s_j} \quad (1)$$

where  $s_i$  is the area of cell  $i$ , and  $S_{C_{ij}}$  is the common area of cell  $i$  and  $j$ , and  $A_{ij} = 1$  when  $i = j$ . In Goltsev *et al.*, the adjacency matrix was introduced as follows:

$$A_{ij} = \frac{C_{ij}}{p_j} \quad (2)$$

where  $C_{ij}$  is the length of common edge of cell  $i$  and  $j$ ;  $p_j$  is the perimeter of cell  $j$ .

This new approach lowers the risk of overcompensation, intrinsic to common edge length correction as illustrated in **Fig S2D**. Compensation by equation (1) effectively reduced the number of cells with intensity 0, as compared to the original method (equation 2), thus obtaining a compensated intensity matrix  $\hat{X} \in \mathbb{R}^{n \times d}$ :

$$\hat{X} = X^T A^{-1} \quad (3)$$

### Outlier removal

As shown in **Fig S2E (1)**, (2) and (3), intensities for all markers in  $\hat{X}$  were rescaled between 0 to 100 and cell segments with low intensities (<3) for all biomarkers or small size (< 50 pixels) were removed for denoising, preserving 102,159 cells from the 178,145 detected segments.

### Clustering analysis

The fluorescence intensity matrices of 148 images, described above, were integrated as a whole containing a total of 102,159 12-dimensional single-cell phenotypic marker intensities. Cell segments were subject to several clustering algorithms (k-medoids [25], Self-organized Map (SOM) [26], X-Shift [27], Hierarchical clustering, and Spectral clustering [28]) with different distance metrics. Except for X-shift, the number of clusters  $k$  was fixed at 70. Since in X-Shift, the

number of clusters  $k$  depends on the number of nearest neighbors of a cell  $c$ ,  $c$  was determined by selecting the clustering result whose  $k$  was closest to 70, as to compare these algorithms as fair as possible ( $k=62$  using Euclidean distance;  $k=69$  using Angular distance). Clustering performance was assessed by Silhouette [29], a measure of the similarity (range -1 to 1) of a cluster member to its own cluster versus to other clusters (Fig S3), where a high coefficient indicates that the member is well matched to its own cluster but not to neighboring clusters.

### *Visualization of clustering performance*

A minimal spanning tree (MST) [30] and polar histograms were constructed based on the median fluorescence intensity of each cell cluster (Fig 4A-B), and the t-SNE (t-Distributed Stochastic Neighbor Embedding) [18] plot (Fig S4A) was generated using the `tsne` function provided by MathWorks (<https://nl.mathworks.com/help/stats/t-sne.html>). Cluster members were also overlaid on H&E images for pathological analysis (Fig 5).

### *Correlation between superclusters and scRNA-Seq dataset*

Log transformation and standardization were performed for each biomarker (from Nuance cell-biomarker matrix) and individual gene (from Zernecke RNA-Seq data) and across all cells. We selected 10 single-cell RNA-Seq (scRNA-Seq) genes (*Cd44*, *Cd47*, *Cd68*, *Mrc1*, *Ly6c1*, *H2-Ab1*, *Itgam*, *Lamp2*, *Adgre1*, and *Plin2*) corresponding to Nuance biomarkers (CD44, CD47, CD68, CD206, Ly6C, MHCII, CD11b, CD107b, F480, and Perilipin2, respectively; Dectin1 expression was not covered in the scRNA-Seq dataset and Ly6G<sup>+</sup> neutrophils were not included in the myeloid meta-analysis). Subsequently, the Pearson correlation coefficient between transcriptomics-based clusters, as annotated by Zernecke et al., and the imaging superclusters was calculated based on the mean of expression or intensity per cell cluster, respectively. After deleting the correlations lower than 0.2, the top 2 connections with the highest correlations were listed in Table S5 and visualized in a bipartite graph (Fig 6A).

### *I-niche analysis for preferred cell communities*

I-niche analysis was performed as described by Goltsev *et al.* [5] with some modifications. Instead of defining a cell's  $i$ -niche as a ring including a set of first-order Delaunay neighbors (Fig S5A (1)), we preferred to account for both the first-order Delaunay neighbors and the Euclidean distance between cells on H&E images, in view of the uneven cell density distribution in atherosclerotic plaques. Therefore, only neighbor cells were included in the  $i$ -niche ring if their Euclidean distance was lower than a given threshold (50 pixels) (Fig S5A (2)).

Assume  $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N]^T$ ,  $\mathbf{B} \in \mathbb{R}^{N \times m}$ , where  $N$  is the number of cell segments in the whole dataset ( $N = 102,159$ ) and  $m$  is the number of supercluster ( $m =$

17), is an i-niches matrix for recording the fraction of the number of a supercluster in an i-niche (Fig S5A, (3) and (4)). For the  $i$ -th cell and  $j$ -th supercluster, let  $D_i$  be a set of cell segments in the i-niche, and  $H_j$  be a set of cell segments belonging to supercluster  $j$ . Then  $b_{ij}$  was defined as follows:

$$b_{ij} = \frac{|D_i \cap H_j|}{|D_i|} \quad (4)$$

$\mathbf{B}$  was clustered into 60 i-niche groups by k-means (Fig S5B). Assume the elements in vector  $\mathbf{q} \in \mathbb{R}^{1 \times m}$  represent the numbers of i-niches for  $m$  central cell types. Co-occurrence matrix of superclusters (cell communities)  $\bar{\mathbf{B}} = [\bar{\mathbf{b}}_1, \bar{\mathbf{b}}_2, \dots, \bar{\mathbf{b}}_m]^T$ ,  $\bar{\mathbf{B}} \in \mathbb{R}^{m \times m}$  was depicted in a heatmap, where average fraction of superclusters in an i-niche ring circumventing central cell type  $l$ , was calculated as follows:

$$\bar{\mathbf{b}}_l = \frac{\sum_{i=l} \mathbf{b}_i}{q_l} \quad (5)$$

## Functional analysis of superclusters

### *Nuclear size and intensity detection*

Nuclei were characterized based on their hematoxylin intensities in post-multispectral imaging H&E staining. Color deconvolution extracted nuclei-only images [22], which were overlaid with the above-described cell segments-defining masks. Next, a series of morphological operations was used to extract the nuclei area, specifically including gamma correction ( $\gamma=4$ ) for contrast adjustment, a dark channel prior dehazing method [23] for image haze reduction, and the top-hat filtering for uneven illumination correction where the images' contrast was adjusted again to obtain 1% saturation in the intensity values. The final nuclei areas were identified by binarizing the processed image and removing single-pixel noise. The size of a nucleus was defined as the number of pixels in a nucleus region, and the maximum intensity of color deconvoluted image per nucleus was used to represent the nuclear haematoxylin staining intensity.

### *Ki-67<sup>+</sup> cell identification*

For each supercluster separately, we calculated the intensity of Ki-67 staining in all cells as a measure of their proliferation rate. To reduce the impact caused by artefact Ki-67 signal and improve the accuracy of Ki-67<sup>+</sup> cell identification, cells were considered Ki-67<sup>+</sup> only if the cell pixel harboring the Ki-67 intensity peak (the local maximum) coincided with the nucleus pixel of that cell, and if the peak intensity was higher than 5.

## Computational procedures for characterizing the cells' micro-environment

The pipeline of cell's micro-environment characterization includes three steps (Fig S6). Firstly,  $m/z$  peaks were picked from mass spectrums via a series of MSI



processing procedures for denoising and reducing the data dimensionality (**Fig S6** (1) and (2)). An MSI pixel was then assigned one of 70 cell clusters by transforming coordinates of cells from the multispectral imaging layer to the MSI layer (**Fig S6** (3) and (4)). Finally, characteristic  $m/z$  values were selected from regenerated samples using Adaboost based feature selection with 10-fold cross-validation [31, 32].

### *Mass spectrometry imaging preprocessing*

A total of 25 sections from MSI layer with no/minor tissue damage were used for lipidomic analysis, as previously described [33]. First, a linear alignment was performed on all mass spectra (intensity vs.  $m/z$  plots of MSI pixels) using FlexAnalysis v3.4 (Bruker Daltonik), choosing  $m/z$  885.6 as calibrant with a 1000 ppm peak assignment tolerance. The mass spectra were then resampled to 40,000 peaks within the 400 to 2000  $m/z$  range using the “msresample” function from Bioinformatics Toolbox (MathWork[34]) with default parameters to share the same separation-unit range. The remaining steps of preprocessing (including baseline correction, denoising, and pick peaking) were followed by a peak detection method using the UDWT on the mean spectrum (MUDWT) [35].  $S/N$  threshold  $\phi$  was set at 6. As a result, 664  $m/z$  peaks were picked for subsequent analyses.

### *Assignment of MSI pixels to phenotypic supercluster*

MSI images were linked to the multispectral images using the above-described 2-step coregistration procedure (**Fig S1B**), transforming the coordinates of all cell segments on the multispectral imaging layer into the MSI layer image. Projected cell segments were then checked by histological examination of the corresponding H&E image to correct for tissue artefacts or imperfect coregistration of nonidentical image layers. Incorrect projections were discarded. For cell segments which are spread over more than one MS pixel, the MS pixel was assigned based on the coordinates of the centroid of that segment. A pixel- $m/z$  matrix was eventually constructed by integration of preprocessed mass spectra from 25 sections. This matrix had a total of 11,995 pixels with 17 phenotypic supercluster labels and 664  $m/z$  peaks (**Fig S6**).

### *Supercluster-specific lipid feature selection*

Cell segment projection onto MSI images was not always flawless, despite manual inspection of co-registration quality, as this was based on plaque textures and not on the locations of cell segments. Thus, each supercluster included outlier MSI pixels. Therefore, for each supercluster, outliers of MSI pixels were smoothed by a bootstrapping strategy involving iterative artificial sample generation (the number of iterations was 50), based on the average MSI intensities from the pixel- $m/z$  matrix of half of randomly selected pixels within that supercluster. 50 samples with 664  $m/z$  peaks were then generated per supercluster. For superclusters E, F and G, or superclusters K, N and O, MSI peak importance was then estimated by summing

these estimates over all weak learners (decision tree) in Adaptive Boosting (AdaBoost) [32] (using the “predictorImportance” function from Statistics and Machine Learning Toolbox, MathWorks[31]). The comparison of classification accuracies in 10-fold cross validation between AdaBoost and other ensemble learning methods (Random Forest [36], Random Undersampling Boosting (RUSBoost) [37], and Linear Programming Boosting (LPBoost) [38]) was shown in **Fig S7A**. The top  $r$  important peaks were picked using two strategies based on the changing curve of the 10-fold cross-validation accuracies with the number of top peaks we used in the model (**Fig S7B**): 1) the number of peaks with the highest accuracy ( $r = 60$  in supercluster E, F and G;  $r = 46$  in supercluster K, N, and O), and 2) the number of features whose accuracy is not significantly different ( $p > 0.05$ ) from the highest accuracy ( $r = 15$  in supercluster E, F and G;  $r = 5$  in supercluster K, N, and O). These  $r$  peaks’ intensities were presented on heatmaps (from R package: ComplexHeatmap [39], **Fig 7B and E**; **Fig S7C and E**), and their classification qualities were visualized by reducing dimensions from  $k$  to 2 using Uniform manifold approximation and projection (UMAP) [40] (**Fig 7C and F**; **Fig S7D and F**).

### *Lipid identification*

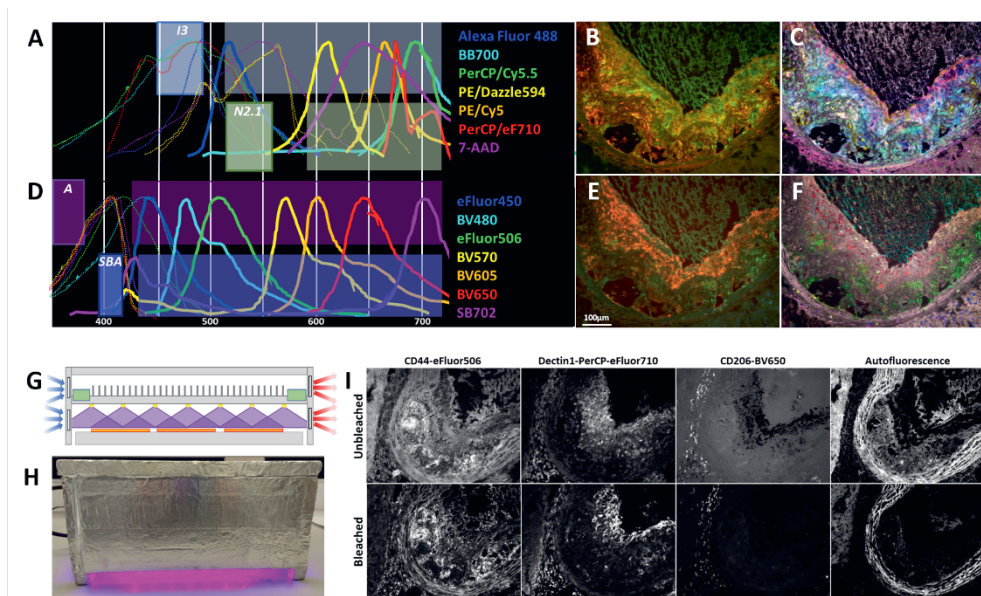
MS/MS was performed in negative polarity using data-dependent acquisition (DDA) on an Orbitrap Elite (Thermo Fisher Scientific, Bremen, Germany) coupled to a MALDI/ESI injector (Spectrograph LLC, Kennewick, WA, USA) [41], or via manual MS/MS acquisition on a timsTOF fleX (Bruker Daltonics, Bremen, Germany). For the DDA, an isolation window of 1 Da, activation  $Q$  of 0.25, and normalized collision energy of 38 (manufacturer units) were used. The full MS was acquired over  $m/z$  400-2000 at 240,000 mass resolution (FWHM at  $m/z$  400). The injection time was set to 250 ms and source pressure was 8.0 Torr. LipostarMSI 1.1.0b28 was used for analysis and lipid identification [42]. Manual MS/MS was performed on the timsTOF fleX using an isolation window of 1 Da and a collision energy of 35-60 eV, optimized per lipid. A total of 1,000 shots were summed, fired at 10 kHz frequency. All lipids were identified by accurate mass from Orbitrap acquisition and by manual assignment of the fragments using Alex123 Lipid Calculator [43].

## Results

### Multispectral immunofluorescent imaging of murine spleen myeloid heterogeneity

2

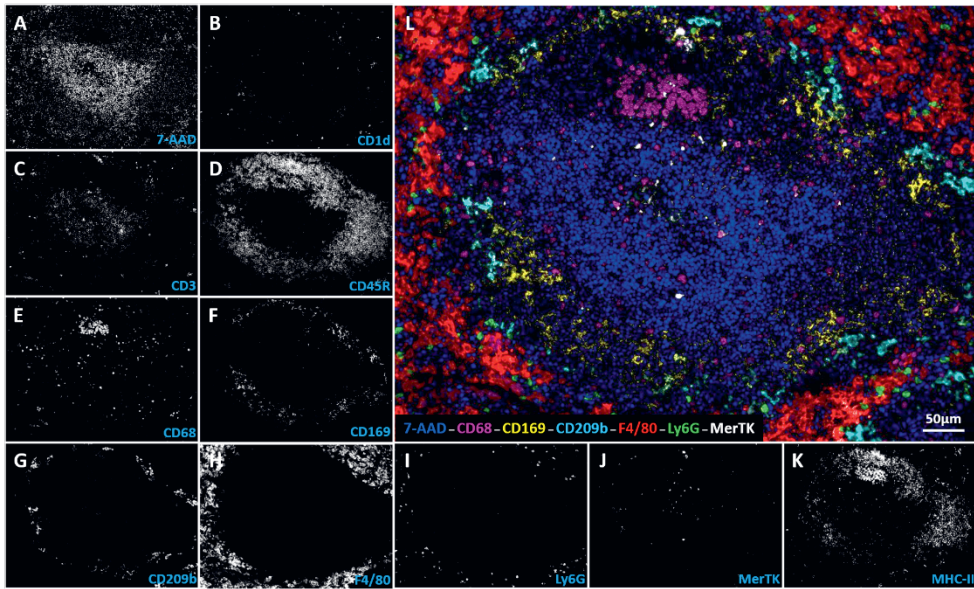
Proof of concept for spatial phenotypic heterogeneity mapping by multispectral imaging was obtained in murine spleen tissue, with its well-documented architecture and microanatomic locations of the different myeloid subsets [44]. Although characteristic markers for spleen leukocyte subsets had been established before [45] and proven successful in distinguishing them in cytometry [46] or immunofluorescence [47], combining multiple markers to reveal relative subset locations remained challenging and required sequential staining approaches [5].



**Figure 1: Multispectral immunofluorescent imaging and bleaching of tissue autofluorescence.** Excitation (dotted) and emission (full) spectra of the antibody-linked dyes used in this study are drawn against a background that indicates the excitation and detection range of the I3 and N2.1 filter cubes (A) or the A and S Blue Aqua filter cubes (D). Illustrative composite pictures adding all imaged wavelengths for each filter cube combination (B and E, respectively) could be unmixed into their individual spectra, of which pseudo-colored merged pictures are shown (C and F, respectively). A schematic overview (G) and picture (H) of the custom-designed bleaching device, featuring a plateau that fits up to 12 tissue slides (orange) simultaneously, evenly distributed near-UV (400nm) light (purple) by 4 rows of 7 high-power (3W) LEDs (yellow), two power sources (green) and both active (fans) and passive (fins) cooling to prevent the LEDs and sections from heating. (I): Examples of murine atherosclerotic plaques stained and imaged without or with prior bleaching, illustrating the reduction in tissue autofluorescence, the improved signal-to-noise ratio for some of the markers and the persistent autofluorescence in the vessel's media layer.

In multispectral IF imaging, stepwise detection of fluorescence emission intensity over the full visible range at fixed excitation wavelength, a so-called *lambda* scan,

allows to recreate a complete emission spectrum for each individual pixel. By comparing this to a pre-defined library of spectra, measured in unstained and single-fluorochrome stained tissue sections for each dye (**Fig 1A&D**), one can computationally deconvolute multiplex stained images (**Fig 1B&E**) into their individual components (**Fig 1C&F**), even despite spectral overlap. This so-called unmixing thus creates a greyscale image of the same field of view (FOV) for each individual fluorochrome (**Fig 2A-K**).



**Figure 2: Murine spleen multispectral imaging of cellular heterogeneity and distribution.** (A-K) Representative images of an 11-marker staining of a follicle in 7µm-thick sections of murine spleen. The captured spectra were unmixed into the individual dyes representing the 11 markers. (L) Since a combination of 11 colors is difficultly distinguishable by eye, a selection of 7 markers was pseudo-colored and combined into a merged image using Fiji [15] to highlight the respective tissue locations of red pulp macrophages (F4/80<sup>+</sup>), marginal zone macrophages (CD209b<sup>+</sup>), metallophilic macrophages (CD169<sup>+</sup>), white pulp macrophages (MerTK<sup>+</sup>), tingible body macrophages (CD68<sup>+</sup>) and granulocytes (Ly6G<sup>+</sup>), in and circumventing a B- and T-cell rich follicle.

Typically, polymer or tandem dyes with similar optimal excitation wavelengths but distinct emission spectra are combined. To increase the number of compatible fluorochromes, we combined dyes that are excited by blue (I3 filter) or green (N2.1 filter) light (**Fig 1A**). Then, to even further expand the number of simultaneously detectable markers on the same tissue section, a second image is made of the exact same FOV, now imaging dyes excited by UV (A filter) and violet (S blue aqua filter) light (**Fig 1D**, **Table S1**). Single-stained tissues were used to define the individual spectra, and the extent of bleed-through between channels was monitored by including the appropriate fluorescence-minus-one (FMO) controls during protocol setup.

We validated the capacity of multiplex spectral imaging to simultaneously discriminate and localize the splenic resident myeloid subsets by staining cryosections with the nuclear marker 7-aminoactinomycin-D (7-AAD) and antibodies against CD1d, CD3, CD45R, CD68, CD169, CD209b, F4/80, Ly6G, MerTK and MHC-II, all coupled to inter-compatible dyes, in an 11-plex panel (**Table S2**). Unmixed images demonstrated efficient and clear separation of these markers into individual greyscale images (**Fig 2A-K**), from which a selection was pseudo-colored and merged to show the respective myeloid phenotype localizations (**Fig 2L**).

### Murine atherosclerotic plaque myeloid heterogeneity

Next, we aimed to map the poorer defined myeloid subsets in murine atherosclerotic plaque, a vascular tissue notorious for its stage- and location-dependent complexity, harboring highly diverse microenvironments and thus myeloid heterogeneity.

Macrophages play a central role throughout the pathogenesis of atherosclerosis. Already at steady state, arteries are populated by myeloid cells that play an essential role in vascular tone control, amongst others [48]. Upon hypercholesterolemia, resident macrophages can clonally expand [49] and additional, monocyte-derived macrophages are recruited to the intimal layer [50], where they become foam cells by ingesting lipids and produce proteolytic enzymes, growth factors and inflammatory mediators, ultimately determining plaque stability.

We now appreciate that all these functions are not executed simultaneously by one single phenotype. Rather, macrophages adopt specialized phenotypes, driven by the highly heterogeneous micro-environment. Plaque, with its plethora of micromilieus, ranging from quiescent, inflammatory, necrotic, smooth muscle cell (SMC)-rich, fibrotic, hypoxic to calcified and neovascularized areas, harbors highly diverse cellular and molecular environments.

Accordingly, conventional histology [51] and, more recently, CyTOF and single-cell RNA-Seq [52-58] have already revealed that plaques harbor several distinct macrophage subsets, yet spatial information on the identified subsets and their molecular context is still lacking.

### Reduction of tissue autofluorescence

Atherosclerotic plaques notoriously display high, polychromatic autofluorescence, which could cause false-positive signals and cannot be quenched by Sudan Black or commercially available kits, as these also quench the often dim signals from directly-labeled, non-amplified antibodies [59]. We therefore opted for overnight photobleaching [59-61], prior to antibody-staining, using a custom-designed bleaching device with high-power near-UV LEDs (**Fig 1G-H**), which strongly reduced

background fluorescence and improved the markers' signal-to-noise ratio, while leaving tissue texture and epitope recognition unaffected. Interestingly, the strong monochromatic autofluorescence of the elastin-rich media layer and cap is maintained (**Fig 1I**).

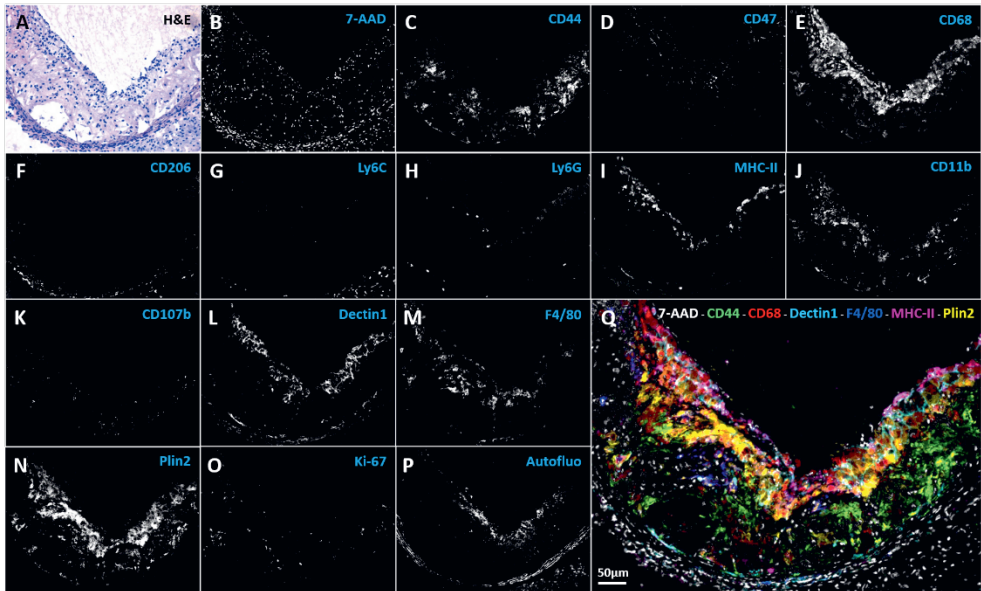
### Atherosclerotic plaque multispectral imaging

After testing the discriminative power and compatibility of different marker combinations and benchmarking each marker with single (and published) IHC staining patterns, we selected a 14-marker panel (supplemented with remaining autofluorescence) for multispectral imaging. It consisted of 12 phenotypic markers known from own experience and relevant literature, a nuclear staining (7-aminoactinomycin D, 7-AAD) and a marker for proliferation (Ki-67) to map the subset's mitogenic capacity (**Table S3**) to apply on a cohort of aortic root atherosclerotic plaques from high fat diet fed LDL receptor deficient mice.

Plaque cryosections were stained with this panel and 148 multispectral images were recorded and each deconvoluted into 15 individual marker components (**Fig 3B-Q**). As this approach leaves the tissue intact, the same sections were subsequently stained with hematoxylin and eosin (H&E) (**Fig 3A**) and aligned to the fluorescent images, regarding dimension, resolution and orientation (**Fig S1A**), enabling pathological validation of our findings as well as characterization of cell morphology.

### Identification and characterization of plaque myeloid phenotypes

Contrary to splenic myeloid heterogeneity, where every marker corresponded to a specific, known cell type or subset, plaque macrophages often featured more than one of the selected markers and were therefore discriminated based on their overall marker expression patterns. Cell segments, identified on the corresponding H&E images, were overlaid on the individual greyscale images to record all cells' relative intensities for each of the markers (**Fig S2B**). These were corrected for spillover from neighboring cell fluorescence, proportional to the estimated overlap of cell masks (**Fig S2C-D**). Eventually, a total of 102,159 cells were found to be positive for at least one of the twelve phenotypic markers (**Fig S2E**) and were subjected to subsequent clustering analyses to define myeloid heterogeneity in plaque but also media, adventitia, aortic valves, and cardiac tissue.



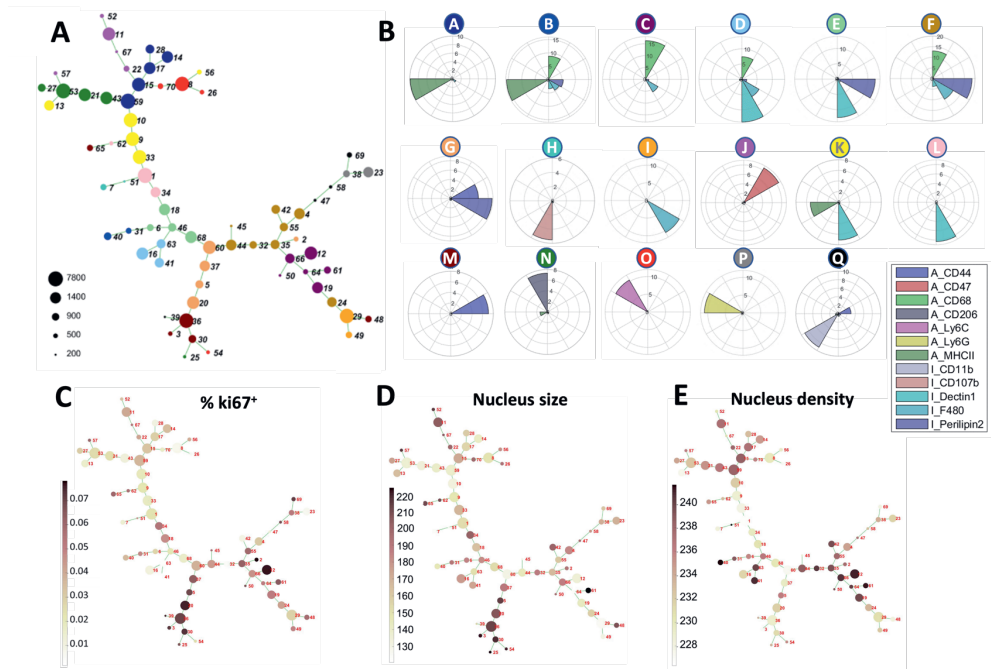
**Figure 3: Murine atherosclerotic plaque multispectral imaging of myeloid heterogeneity.** Multispectral imaging of a representative murine aortic root atherosclerotic plaque. (A) An H&E image, cropped to the same size, orientation and resolution as the fluorescent images. (B-P) Unmixed greyscale images for the individual spectra and thus markers. (Q) A selection of these images was pseudo-colored and merged using Fiji [15] to illustrate the uneven distribution of these markers over the tissue. The signal-to-noise ratio of these images was optimized manually using Fiji for esthetic purposes.

Cells were intentionally over-clustered (number of clusters or  $k=70$ ) to ensure proper segregation of even the rare subsets. Several clustering methods were compared for their performance, and as most cells are only partly entrapped in 7  $\mu\text{m}$  sections, we included both clustering based on absolute, integrated marker intensities and the more appropriate relative marker intensities from cosine, rather than Euclidian, distance. K-medoid clustering based on cosine distance performed best, as judged from silhouette analysis [29] (Fig S3A) and empirical histopathological examination of cluster-assigned cells in at least 7 sections from different mice. It was therefore selected as method of choice.

The 70 resulting clusters were presented as a tSNE plot [62] (Fig S4A) or minimal spanning tree (MST [30]) to visualize mutual similarity of clusters as well as their respective cell numbers (Fig 4A). The marker expression profile for each cluster is shown in radar plots (Fig S3B). Next, these 70 clusters were assigned manually to 17 superclusters (named A–Q, color coded in Fig 4A and Fig S4A), based on inter-cluster distance in the MST, similarity in marker expression pattern and analysis of their predominant tissue localization. Their mean marker expression was visualized in Fig 4B and S4D, and their relative abundance was shown in Fig S4B.



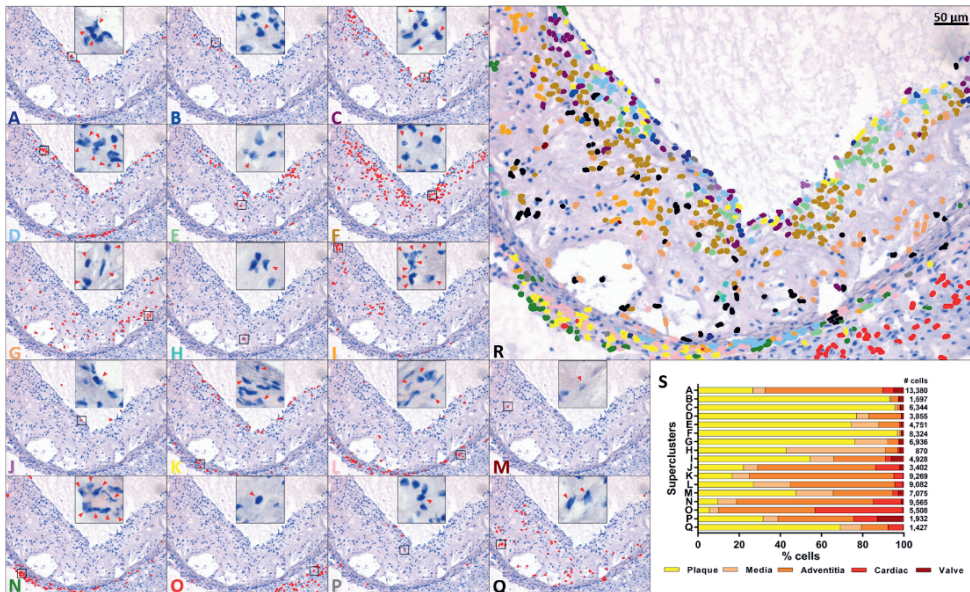
Next to phenotypic marker expression, adding functional reporters to the staining panel can further characterize these superclusters. As an example, the percentage of cells positive for Ki-67, a proliferation marker, was projected on the MST (**Fig 4C**), showing almost 10-fold differences in clonal expansion capacity between plaque macrophage clusters [63]. Similarly, H&E-based heterogeneity in nuclear size and density reflected the sizeable differences between superclusters (**Fig 4D-E** and **Fig S4C**).



**Figure 4: Murine atherosclerotic plaque myeloid cell phenotypic clustering based on multispectral imaging.** (A) In 148 images from 65 sections of aortic root atherosclerotic plaques from 11 WD-fed *Ldlr*<sup>-/-</sup> mice, a total of 102,159 cells were identified as positive for at least one of the phenotypic markers. They were divided over 70 clusters and represented in a minimal spanning tree, where the color code groups the clusters that were combined into 17 superclusters and the circle sizes reflect the number of cells in each cluster; (B) Radar plots indicating the mean expression patterns for 12 phenotypic markers in each of these superclusters; The percentage of Ki-67<sup>+</sup> cells (C), mean nuclear size (D) and mean nuclear density (E) for each cluster were overlaid on the MST.

As every of these cells' tissue location can be traced back, the distribution of the superclusters could be projected on the corresponding H&E image (**Fig 5A-R**). This not only allowed for histopathological examination of their morphology by an experienced pathologist, which further confirmed the validity of the assigned superclusters, but also revealed their preferential localization in five different manually annotated tissue compartments (**Fig 5S**).

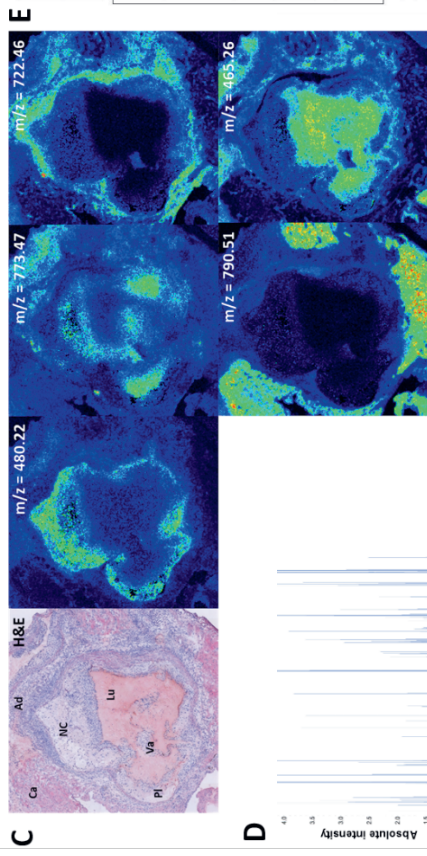
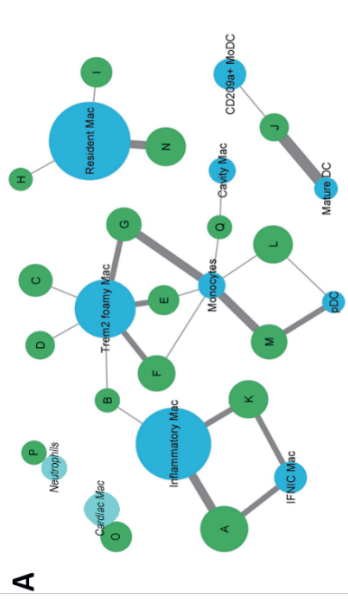
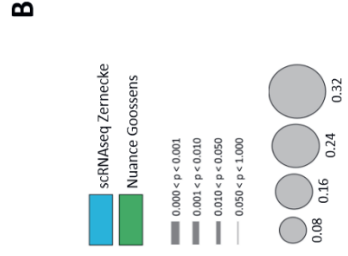
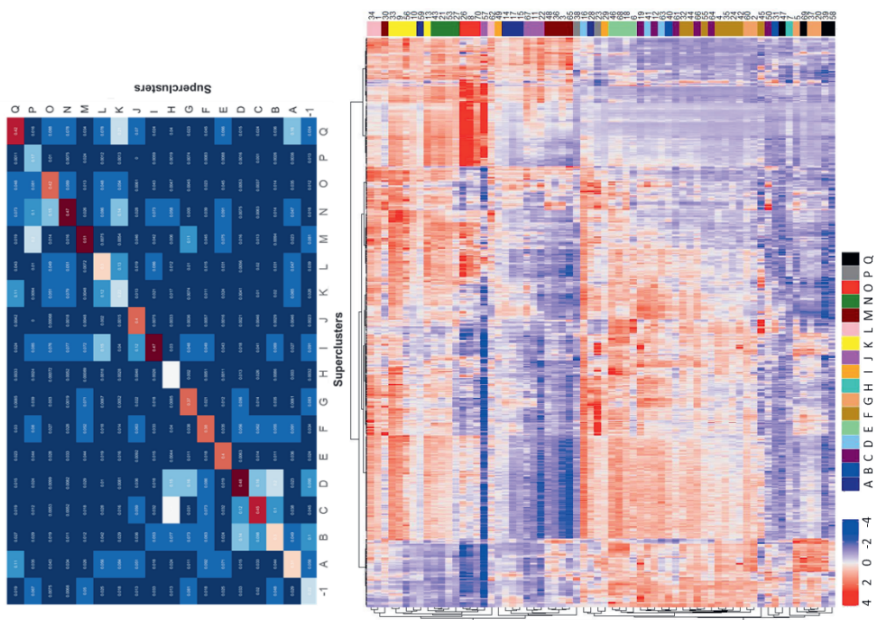




**Figure 5: Tissue distribution of murine atherosclerotic plaque myeloid phenotypes.** A representative H&E image of a murine aortic root atherosclerotic plaque was overlaid with the localizations of the cells that were featured in the 17 superclusters separately (A-Q) or, pseudo-colored, together (R) to illustrate the histological localization of the superclusters. For each supercluster, an insert with a high-magnification image of the area corresponding to the square illustrates representative cells' (marked with red arrowheads) histopathological phenotypes. The superclusters' distributions over five different tissue components (plaque, medial layer, adventitia, cardiac tissue and aortic valves) were compared (S).

## Supercluster identification and characterization of the cells' cellular and molecular context

Single-cell RNA-Seq has been a useful tool not only to illustrate the atherosclerotic plaque's cellular and even myeloid heterogeneity but also to further identify each subset [52, 55, 56]. We illustrate its complementarity with our multiplex imaging approach by comparing marker expression patterns of the imaging-based superclusters to the corresponding genes' expressions in myeloid cell types identified through meta-analysis of existing transcriptomic datasets [58], and calculated the similarities between populations defined by both approaches (Table S4). Despite the limited number of features, the comparison of gene expression with staining intensity or the different cellular composition in aortic arch versus aortic root plaques, we found strong correlations that shed a light on the phenotypic identities of the superclusters, as discussed below (Fig 6A).



**Figure 6: Characterization of murine atherosclerotic plaque myeloid phenotypes' cellular and lipidomic contexts.** (A) A bipartite graph showing the correlation between the 17 superclusters (green) and 9 myeloid cell clusters from Zernecke *et al.*[58] (blue) based on 10 phenotypic markers. Two populations not included in the myeloid scRNA-Seq dataset were added manually (turquoise). Node sizes indicate the clusters' cell count proportion of the respective dataset's total cell number, edge thickness marks the significance level of the correlation; (B) Co-localization of superclusters, represented in a heatmap, illustrates that subsets tend to aggregate rather than intermingle with other phenotypes; (C) Illustrative examples of MSI m/z peaks that are highly associated with tissue compartments such as lumen (Lu), plaque (Pl), aortic valves (Va), adventitia (Ad), cardiomyocytes (Ca) or the plaque's necrotic core (NC); (D) A representative example of a single pixel's MSI-derived lipidomic spectrum, ranging from m/z 400 to 2000. The height of the peaks represents the signal intensity and therefore the local corresponding lipid concentration; (E) A heatmap of the 664 m/z peak intensities assigned to the 70 clusters or 17 (color-coded) superclusters identifies subset-associated lipidomic environments. While the top half of the heatmap contains mostly plaque-associated clusters, the bottom half is highly enriched in adventitial clusters.

Since macrophage phenotypes are orchestrated by their cellular and molecular micro-environment, we explored approaches to address both. First, we demonstrated the high degree of supercluster compartmentalization, when identifying each cell's nearest neighbors (**Fig 6B**). This finding was subsequently confirmed and refined in cell community mapping by defining first-order Delaunay neighbors for each index cell in an i-niche matrix [5] (**Fig S5A-B**). Both approaches showed that most niches are surprisingly homogenous, underpinning the importance of context-dependent instructions.

For complementary characterization of the molecular context associated with the superclusters, we performed high-resolution matrix-assisted laser desorption/ionization (MALDI) mass spectrometry imaging (MSI) on adjacent sections, identifying molecular constituents such as peptides, metabolites or lipids. Considering lipids' central role in atherogenesis, we focused on differences in lipidomic context, thereby revealing a combination of potentially phenotype-directing and subset-derived lipids.

High-spatial resolution MALDI-MSI for lipids in negative ionization mode (**Fig 6C-D**) was performed on 65 aortic root sections that were thaw-mounted onto indium-tin oxide (ITO) coated slides, flanking the multispectral imaged sections, and followed by an H&E staining on the same sections, as previously described [33] (**Fig S1C**). Due to MSI-intrinsic tissue damage, only 25 of these were selected for alignment with their adjacent sections. We designed a two-step coregistration protocol to correct for differences in tissue shape, orientation and image resolution. This allowed us to align the two data layers at subcellular precision (5.8 +/- 2.0  $\mu\text{m}$  deviation, based on 40 image alignments in 25 sections) (**Fig S1B**). Next, data layers were spatially linked and the previously determined IF-based cell segment coordinates were projected to the adjacent MSI sections, assigning the lipid signatures in the 15x15  $\mu\text{m}$  MSI pixels to the different superclusters. The measured mass spectra were analysed using a preprocessing pipeline, as previously

described [28, 33], including baseline correction, denoising, and peak picking. This compressed the lipid list from 40,000 to 664 m/z peaks or features (**Fig S6**).

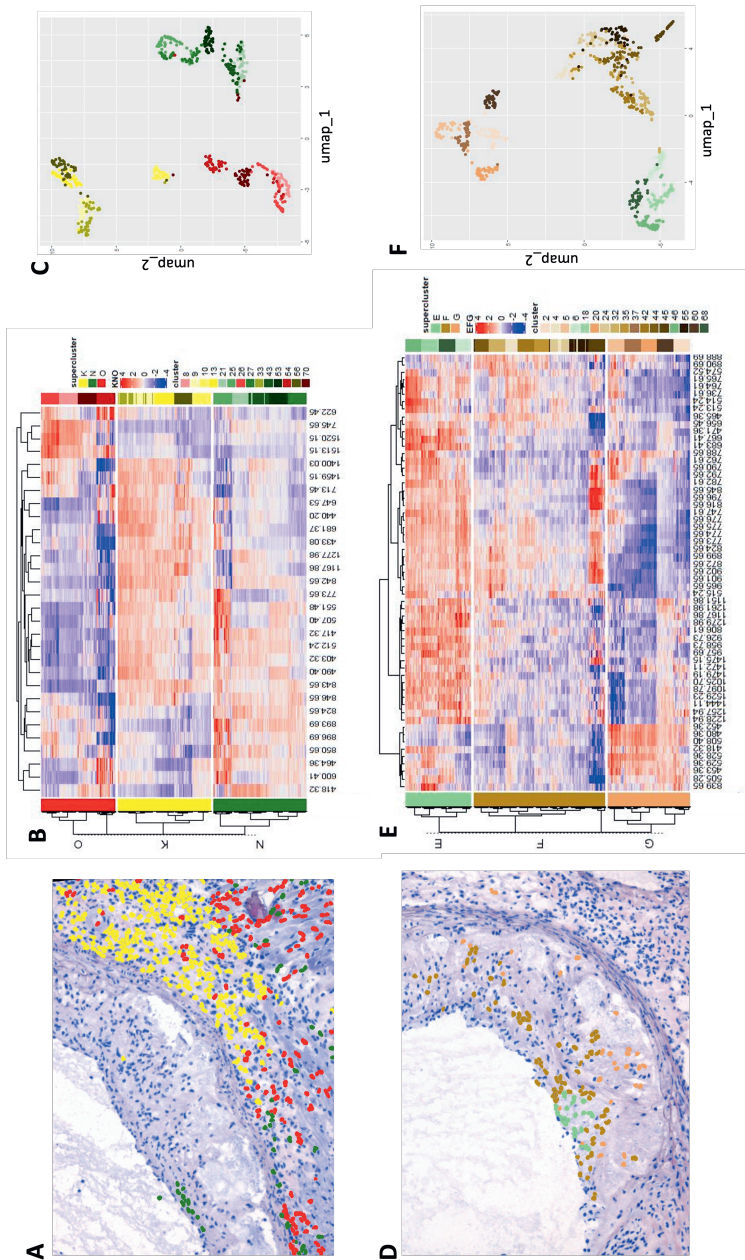
As illustrated by the heatmap that integrates the final lipid set with the phenotypic clusters (**Fig 6E**), not only did we note striking differences in lipidomic environment between plaque and adventitial cells but also within these compartments several outspoken signatures exist. Furthermore, cell segment organization based on lipid context showed close resemblance to the empirical definition of the superclusters based on multispectral imaging.

### Characterization of adventitial and plaque myeloid phenotypes

We highlighted two examples, comparing three distinct adventitial subsets and three foam cell phenotypes, respectively, to further illustrate the value and relevance of these measurements in the definition and characterization of superclusters and their micro-environment.

Within the adventitial compartment of the aortic roots, three predominant myeloid phenotypes were identified (**Fig 7A**). With high CD206 expression, supercluster N corresponds to the prototypical (*Lyve1*<sup>+</sup>) vessel wall resident phenotype [48, 64]. A second, Ly6C<sup>HI</sup> population (supercluster O), appeared to reside mainly in activated cardiac tissue [65] but was also found in the adjacent inflamed aortic adventitia. Thirdly, a population of small, non-mitogenic Dectin1<sup>+</sup> MHC-II<sup>+</sup> macrophages with faint nuclei (supercluster K) represented an inflammatory phenotype in adventitia that highly resembled a population with similar characteristics but lower MHC-II expression that is divided over the adventitia and media layer of some plaques (supercluster L), urging further research on whether these populations would be able to cross this physical barrier towards the plaque.

Lipid signatures corresponding with these three subsets were picked using tree-based adaptive boosting (AdaBoost [32]) guided exhaustive selection (**Fig S7A-B**). Maximal and optimal performance was provided by 46 or 5 m/z signature sets, respectively, which allowed robust segregation of the adventitial subsets solely based on this lipidomic profile, as illustrated in heatmaps (**Fig 7B & S7C**) and UMAP plots (**Fig 7C & S7D**).



**Figure 7: Comparison of lipidomic contexts of myeloid phenotypes in adventitia and plaque.** (A, D) The histological localization of three adventitial (K in yellow, N in green, O in red) and three foam cell phenotypes (E in lime, F in gold, G in orange), respectively, was overlaid on an H&E staining of illustrative plaques; (B, E) Heatmaps reveal the local abundance of the respective 46 and 60 lipids that best associated with the different subsets in both tissue compartments; (C, F) UMAPs visualize classification performance based on these best correlating lipid features, as compared to the color-coded classification based on multispectral imaging.



Foam cell superclusters were analysed in a similar manner. A prototypical plaque macrophage phenotype, foam cells are typified by marked intracellular lipid accumulation but are often missed or underrepresented in single-cell approaches due to their fragility during tissue processing [58]. In our data, three distinct subsets markedly expressing the lipid droplet-associated protein Perilipin2 (Plin2, or Adipophilin) could be distinguished (**Fig 6D**). Inferred from their location proximal to the aortic lumen and histopathologically defined limited lipid content, supercluster E likely represents an early foam cell phenotype. Supercluster F, deeper in plaque, involves highly lipid-laden cells with fading membrane integrity as an early sign of necrotic death, thus representing *bona fide* foam cells. Supercluster G's Plin2 expression [66] and lack of further myeloid markers, together with an often spindle-shaped morphology, suggest a smooth muscle cell origin of these foam cells [56, 66-68]. Although residing in deeper, necrotic plaque areas, they are mitogenic (**Fig 4C**) rather than necrotic. They share their proliferative capacity and CD44 expression with Plin2-negative, non-foamy cells (supercluster M), which morphologically correspond to smooth muscle cells.

In search of lipid signatures, characteristic of these individual superclusters, we identified 60-lipid and 15-lipid sets to provide maximal (**Fig 7E-F**) and optimal (**Fig S7E-F**) accuracy, respectively. These correlations highlight the strong, mutual interactions between myeloid phenotypes and their micro-environment.

Despite relatively low peak intensities of most m/z values determined by high-resolution MS imaging, we managed to assign 26 of those featured in **Fig 7B&E** to their corresponding lipid identities through tandem mass spectrometry (MS/MS). They are listed in Table S5 and their correspondence to the above-highlighted superclusters is illustrated in a heatmap (**Fig S7G**).

## Discussion

We present a powerful, new pipeline for integrated spatial analysis of cell heterogeneity and their immediate molecular context in complex tissue. The 15-color multispectral imaging approach is rapid and affordable, does not compromise the tissue's integrity and provides high phenotypic resolution, as illustrated here for myeloid cells where we were able to distinguish an unprecedented number of myeloid subsets and their preferred habitat in murine atherosclerotic plaques. Embedded in the above-mentioned pipeline it unfolds its true power, as it offers a unique opportunity to define the cellular and molecular micro-environment that associates with a cell's phenotype, providing key information on cells' crosstalk with their direct context.

**Multispectral imaging versus IHC and single cell approaches:** Multispectral imaging outperforms conventional IHC and IF microscopy approaches on the number of markers that can be simultaneously detected, revealing subtle

phenotypic differences or cell type co-localizations that were previously overlooked or difficult to image. Despite its more limited marker panel, it offers similar phenotypic resolution as modern single-cell cytometric [55, 69], CyTOF [57] or transcriptomic tissue heterogeneity assessment [52, 55, 56]. This is partly due to the higher number of cells analysed: phenotypic clustering was based on well over 100,000 cells in our study, which is 10 times higher than the cumulative cell number in the myeloid meta-analysis of several murine atherosclerosis scRNA-Seq datasets (10,551) [58]. Combining the latter dataset with phenotypic and contextual information on our superclusters led to new insights in plaque myeloid cell diversity.

Supercluster A corresponds to the inflammatory macrophage subset and represents newly-recruited monocyte-like cells lining the plaque's luminal border. The very similar supercluster B bridges the gap to macrophages with histopathologically confirmed lipid accumulation. While RNA-Seq studies identified one single foamy macrophage subset ( $Trem2^{hi}$ ), multispectral imaging revealed five distinct foam cell subsets (superclusters C, D, E, F & G), of which the first four display increasing levels of lipid accumulation as well as of Perilipin2 expression. The last one however highly correlates, just like the Perilipin2-negative supercluster M, with Zernecke's monocyte population. Their localization in the more necrotic plaque areas, their relatively high proliferative capacity and their often spindle-like morphology however suggests these might rather represent transdifferentiated smooth muscle cells, thought to make up for up to half of the plaque's foam cell content [56] but not included [52, 55] or indistinguishable from macrophage-derived foam cells [56] in scRNA-Seq-based studies.

Similarly, in the adventitia we identified 3 major subsets. A population of inflammatory macrophages that resembles the above-described newly-recruited inflammatory macrophages but predominantly lines the plaque's adventitial instead of luminal borders (supercluster K),  $CD206^+$  macrophages that correspond to the  $Lyve-1^+/PF4^+$  population of resident macrophages (supercluster N) [48, 52, 64], and  $Ly6C^+$  cardiac resident macrophages that infiltrate the inflamed adventitia (supercluster O) [65]. The latter was not included in the scRNA-Seq meta-analysis since it covers aortic arch plaques while our plaques in the aortic root were in close contact with cardiac tissue. Also neutrophils (supercluster P) were missing from the transcriptomic dataset as these cells were excluded prior to meta-analysis of the myeloid populations.

Projecting Ki-67 positivity on the clusters highlighted the surprisingly high divergence in mitogenic capacity across myeloid phenotypes, some subsets (such as G) showing an almost 10-fold higher proportion of  $Ki-67^+$  proliferating cells than others (such as D). This may partly underly the discrepant results of studies

suggesting the plaque macrophage pool to mainly originate from resident macrophage expansion [63] versus monocyte influx [49, 70].

**Phenotype-associated context:** Linking phenotypic heterogeneity to cellular and molecular context is the second key advance in this study. Both the i-niche's indication that communities were surprisingly homogenous and the shared lipid signature of supercluster members are supportive of a cell's local context as its main phenotype driver, although it remains to be determined whether signature-contained lipids are causal in or epiphenomena of their phenotype. The pursued strategy allows more direct assessment of tissue- or environment-specific differentiation signals than conventional transcription factor-centered approaches used so far for peritoneal macrophages (GATA6/RAR $\beta$ , retinoic acid) [71], microglia cells (SMAD/TGF $\beta$ /Lrrc33) [72], or Kupffer cells (DLL4/TGF $\beta$ /LXR $\alpha$ ) [73].

Several MSI studies by us and others have underpinned the potential of MS imaging for mapping lipid patterns in plaque, identifying lysophosphatidylcholine, diacyl glycerol, and cholesteryl species to be strongly associated with atherosclerotic plaque [33, 74, 75]. Attempts to integrate MSI to histological features [76] or cells [77] are very scarce and multi-isotope MSI for thymidine and glucose has been deployed to map metabolic and mitogenic activity of plaque cells, respectively [78]. Here we have taken one step further, defining the lipid context for each identified subset at cellular level. The physiological and metabolic mechanisms involved in the accumulation of these context lipids and functional implications of these phenotype-lipid correlations remain as yet unclear and require more lipid identities to be precisely defined through MALDI-MS/MS.

**Broader applicability of the pipeline:** Next to its use for validating or complementing single-cell study outcomes, our versatile multispectral imaging pipeline lends itself very well for application on other plastic cell lineages (e.g., mesenchymal cells, endothelial cells), disease settings, tissues and/or species. In that case only the antibody clones, not the dye combination, need to be adapted. With the ever-expanding collection of commercially available antibody probes developed for cytometry, this will not pose a major hurdle. A key step in the setup of a robust multispectral imaging pipeline involved the autofluorescence bleaching. Depending on the tissue or environment, autofluorescence can strongly interfere with the non-amplified signals. While others have used broad-spectrum photobleaching [59], we found that overnight bleaching with a narrow spectrum near-UV light was sufficient to almost completely abrogate all auto-fluorescent signal to manageable levels, while the persistent signal from the fibrous media layer and cap could be utilized to label-free image these structures.

**Limitations of Study:** Despite some advantages over other multiplexing technologies [4], the here described multispectral imaging comes with its own caveats. The need for directly labelled antibodies is a limitation of this multispectral



approach when compared to conventional IHC or IF, as it reduces commercially available options, precludes signal amplification and often necessitates longer exposure times with increased risk of photobleaching. Brightness of each labelled antibody cannot be adjusted separately when imaging with fixed exposure time and light intensity. Therefore, each antibody's concentration must first be titrated prior to incorporation in the multiplex staining panel, to achieve compatible signal strengths and avoid bleed-through from overly bright dyes into the neighboring spectra. Moreover, a multi-marker panel also renders staining of formalin-fixed paraffin-embedded (FFPE) tissue sections more challenging as this may require incompatible antigen retrieval steps for the different antibodies.

A further limitation involves the linkage of subsets to cell populations defined by scRNA-Seq analyses [58]. Due to the limited size of the myeloid marker panel and lack of correspondence between protein and gene expression in general, congruity based on expression of the 12 myeloid markers used in our study and the scRNA-Seq signatures of myeloid populations is risky, and especially so when considering the incomplete cell recovery in single-cell approaches. Integration of the phenotypes with MS imaging-based peptidomic measurements, similar to the here-presented lipidomic profiles, may partly bypass this issue.

A final limitation involves the spatial versus mass resolution of MSI, which at present does not allow to discern concomitant assessment of the exact chemical identity of lipid features at subcellular resolution. This complicates identification of micro-environmental cues that are driving the phenotype. Also, limited signal-over-noise levels of the peaks measured by high-resolution MSI hampers their identification by tandem mass spectrometry. However, MSI technology is developing at such a high pace, that these challenges are likely to be addressed in the near future.

In conclusion, we introduce a novel pipeline for integrated spatial analysis of cell heterogeneity and their immediate molecular context in complex tissues. The presented multispectral imaging approach is rapid, affordable and provides high phenotypic resolution, as illustrated in myeloid cells. It not only reveals the cells' histological localization but the analysis pipeline equally allows subsequent histopathological assessment as well as to associate the cells with process markers or their cellular and molecular micro-environment, thereby allowing further characterization of the subsets and the identification of their phenotype-driving niches.

## Acknowledgements

The authors would like to thank Wouter Van Der Poten for his help in constructing the custom autofluorescence photobleaching tool, and Helma Kuijpers and Kèvin Knoops for their technical support.

This work has been supported by the Marie Skłodowska-Curie Actions (H2020 MSCA-IF) [661099 AtheroMphProliferation to P.G.], the European Research Area Network Joint Transnational Call for Cardiovascular Disease (ERA-CVD and the Dutch Heart Foundation) [JTC-2017t100 AtheroMacHete to P.G. and E.A.L.B.], the Dutch Heart Foundation [Dekker 2020T042 to P.G.] and the China Scholarship Council (CSC) [No. 201706990018 to C.L. and No. 201706040068 to J.H.C.]. J.H.C. acknowledges the financial support by the Dutch Province of Limburg through the LINK program.

## Author contributions

P.G., E.W., M.J.M.vZ., R.M.A.H. and E.A.L.B. contributed to the study design and developed the experimental approach. P.G., J.H.C. and G.E.F. performed the experiments. C.L., J.H.C. and B.S.R.C. processed and integrated the data. C.L., P.G., J.M.H.K., E.S. and E.A.L.B. developed the data integration approach. P.G., M.G. and E.A.L.B. interpreted the data. T.F.E.H., K.W., M.J.M.vZ., B.B., E.C. and R.M.A.H. provided essential expertise or tools. P.G., C.L., J.H.C., M.P.C.D. and E.A.L.B. wrote the paper.

## Declaration of interests

The authors declare no competing interests.

## Supplemental files

2

**Table S1: An overview of the four filter cubes combined in the multispectral imaging approach; Related to Fig 1A & D.**

This overview indicates the range and resolution in which the emission spectra were recorded as well as the exposure time per step.

Filter cube	Band pass	Excitation		Emission recorded		
		Dichromatic mirror	Long pass	Range	Steps	Time / step
<b>I3</b>	450 - 490 nm	510 nm	515 nm	510 - 720 nm	5 nm	1100 ms
<b>N2.1</b>	515 - 560 nm	580 nm	590 nm	580 - 720 nm	5 nm	700 ms
<b>A</b>	340 - 380 nm	400 nm	425 nm	420 - 720 nm	5 nm	3500 ms
<b>SBA</b>	395 - 415 nm	425 nm	430 nm	420 - 720 nm	10 nm	1100 ms

**Table S2: The dye-conjugated antibodies and their dilutions in the panel used for multispectral imaging in WT spleens; Related to Fig 2A-K.**

As an antibody against CD209b was not commercially available labelled with a compatible dye, an Alexa488-coupled secondary antibody that did not cross-react with the other antibodies in the panel was used.

Antigen	Fluorochrome	Manufacturer	Clone	Dilution
-	<b>7-AAD</b>	BD Biosciences	-	1:500
<b>CD1d</b>	PE/Dazzle594	BioLegend	1B1	1:50
<b>CD3</b>	eFluor 506	ThermoFisher	17A2	1:60
<b>CD45R</b>	PerCP/Cy5.5	ThermoFisher	RA3-6B2	1:100
<b>CD68</b>	VioBlue	Miltenyi Biotec	FA-11	1:30
<b>CD169</b>	Brilliant Violet 605	BioLegend	3D6, 112	1:30
<b>CD209b</b>	Unconjugated & Anti-armenian hamster Alexa488	ThermoFisher & abcam	22D1	1:40 & 1:500
<b>F4/80</b>	PE/Cy5	BioLegend	BM8	1:100
<b>Ly6G</b>	Brilliant Violet 570	BioLegend	1A8	1:20
<b>MerTK</b>	PerCP/eF710	ThermoFisher	DS5MMER	1:60
<b>MHC-II (I-A/I-E)</b>	Super Bright 702	ThermoFisher	M5/114.15.2	1:40

**Table S3: The dye-conjugated antibodies and their dilutions in the panel used for multispectral imaging in murine aortic root plaques; Related to Fig 3B-P.**

Antigen	Fluorochrome	Manufacturer	Clone	Dilution
-	7-AAD	BD Biosciences	-	1:230
CD11b	Brilliant Blue 700	BD Biosciences	M1/70	1:40
CD44	eFluor 506	ThermoFisher	IM7	1:75
CD47	Brilliant Violet 480	BD Biosciences	miap301	1:15
CD68	Brilliant Violet 605	BioLegend	FA-11	1:100
CD107b (Mac3)	PerCP/Cy5.5	BD Biosciences	M3/84	1:20
CD206 (MMR)	Brilliant Violet 650	BioLegend	C068C2	1:20
Dectin-1 (Clec7a)	PerCP/eFluor 710	ThermoFisher	bg1fpj	1:30
F4/80	PE/Cy5	BioLegend	BM8	1:100
Ki-67	PE/Dazzle594	BioLegend	16A8	1:35
Ly6C	eFluor 450	ThermoFisher	HK1.4	1:20
Ly6G	Brilliant Violet 570	BioLegend	1A8	1:20
MHC-II (I-A/I-E)	Super Bright 702	ThermoFisher	M5/114.15.2	1:35
Perilipin2	Alexa488	Novus Biologicals	(polyconal)	1:20
Elastin	Autofluorescence	-	-	-

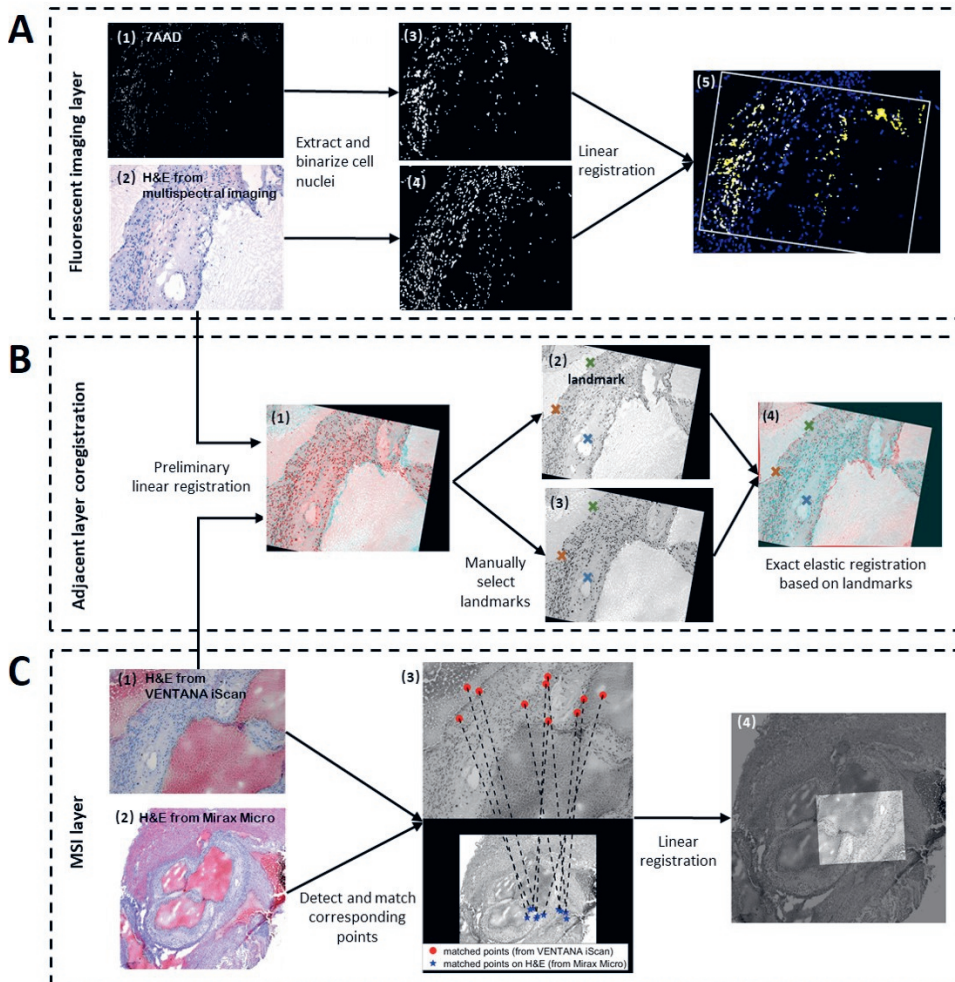
**Table S4: Pearson correlation between superclusters identified through multispectral imaging and populations annotated in the meta-analysis of myeloid cells from murine atherosclerotic plaque scRNA-Seq datasets by Zernecke *et al.*; Related to Fig 6A.**

For each supercluster, up to two transcriptome-defined populations with a correlation higher than 0.2 were assigned. Significant correlations ( $p < 0.05$ ) were displayed in bold.

Supercluster	Zernecke <i>et al.</i> populations	Correlation	p value
A	Inflammatory Mac	0.813	<b>0.004</b>
A	IFNIC Mac	0.687	<b>0.028</b>
B	Inflammatory Mac	0.418	0.229
B	Trem2 foamy Mac	0.271	0.450
C	Trem2 foamy Mac	0.515	0.128
D	Trem2 foamy Mac	0.523	0.121
E	Trem2 foamy Mac	0.639	<b>0.047</b>
E	Monocytes	0.543	0.105
F	Trem2 foamy Mac	0.707	<b>0.022</b>
F	Monocytes	0.294	0.409
G	Monocytes	0.817	<b>0.004</b>
G	Trem2 foamy Mac	0.706	<b>0.022</b>
H	Resident Mac	0.223	0.535
I	Resident Mac	0.476	0.164
J	Mature DC	0.884	<b>0.001</b>
J	CD209a+ MoDC	0.563	0.090
K	IFNIC Mac	0.711	<b>0.021</b>
K	Inflammatory Mac	0.670	<b>0.034</b>
L	pDC	0.574	0.083
L	Monocytes	0.491	0.150
M	Monocytes	0.781	<b>0.008</b>
M	pDC	0.752	<b>0.012</b>
N	Resident Mac	0.837	<b>0.003</b>
Q	Cavity Mac	0.454	0.187
Q	Monocytes	0.380	0.279

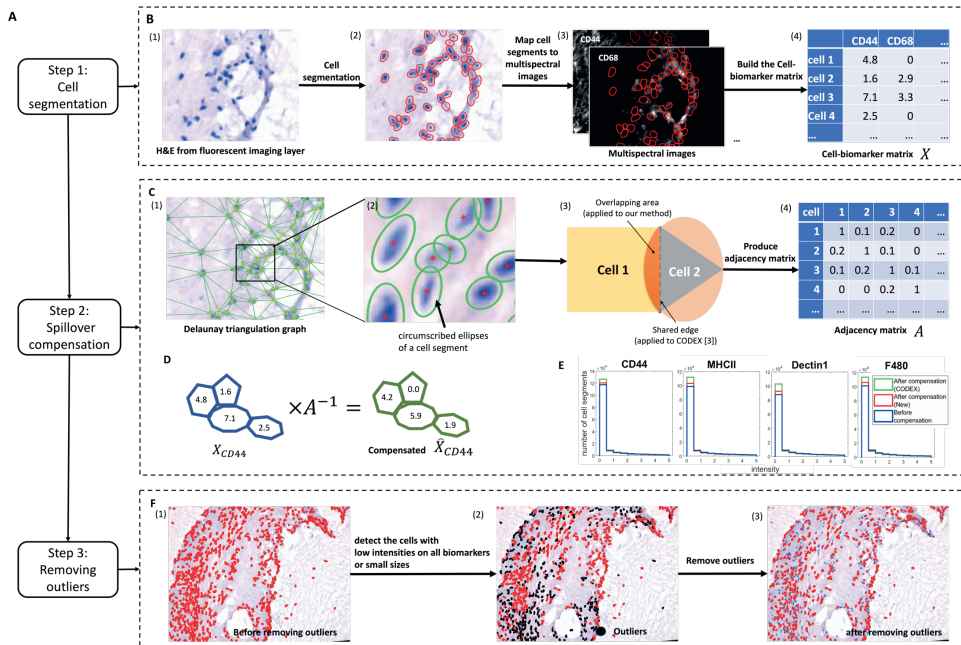
**Table S5: Lipid identities corresponding to some of the most prominent MS imaging m/z peaks, as identified by MS/MS; Related to Fig 7.**

m/z detected	Lipid ID
465.36	Cholesterol sulfate
480.36	LPE 18:0
480.36	LDMPE 16:0
508.40	LDMPE 18:0
599.3	LPI 18:0
619.3	LPI 20:4
673.5	PA 16:0_18:1
697.5	PA 18:1_18:2
699.5	PA 18:1_18:1
701.5	PA 18:0_18:1
722.51	PE O-36:5
723.5	PA 18:0_20:4
742.62	PE 16:0_20:2
742.62	PE 18:0_18:2
747.61	PA 18:0_22:6
747.61	PG 16:0_18:1
762.5	PE 16:0_22:6
762.61	PE 16:0_22:6
764.61	PE 16:0_22:5
764.61	PE 18:1_20:4
766.5	PE 18:0_20:4
770.6	PE 18:0_20:2
773.65	PG 18:1_18:1
775.65	PG 18:0_18:1
788.65	PE 18:1_22:6
788.65	PS 18:0_18:1
790.5	PE 18:0_22:6
792.65	PE 18:0_22:5
792.65	PE 18:1_22:4
810.5	PS 16:0_22:4
821.64	PG 18:0_22:6
821.64	PG 18:1_22:5
834.5	PS 40:6
871.66	PI 17:0_20:4
883.5	PI 18:1_20:4
885.5	PI 18:0_20:4
887.67	PI 18:0_20:3



**Fig S1: Pipeline of the different image registration procedures; Related to Fig 3A & 6E.**

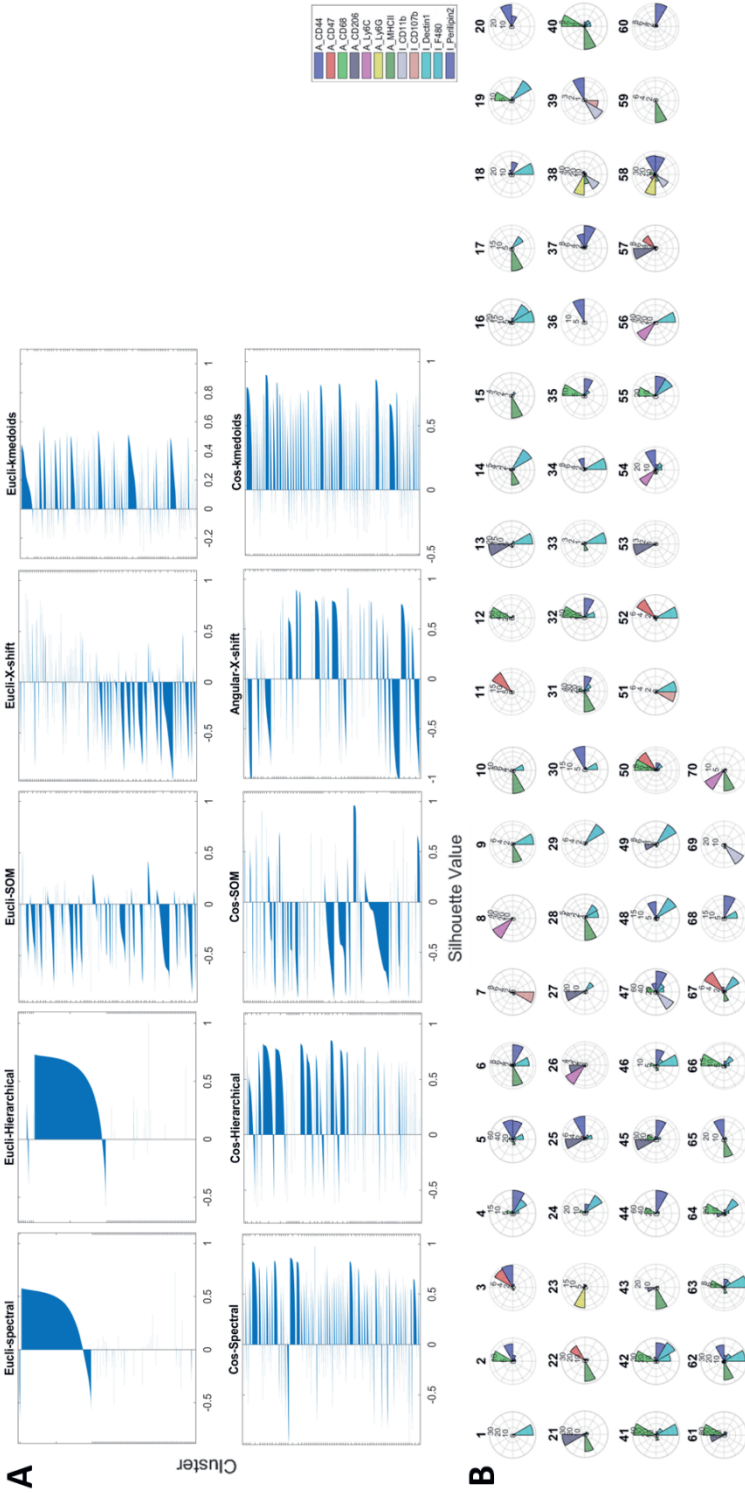
(A) Linear registration of the H&E and multispectral fluorescent images from the same section (Fluorescent imaging layer), starting from (1) a 7-AAD spectral image and (2) its corresponding H&E staining image, extracting and binarizing the nuclei-related features (3 & 4) and aligning with MMI metric (5). (B) Co-registration between the adjacent MS and multispectral imaging layers. H&E images from adjacent sections (A2 & C1) were co-registered by a two-step (MMI based linear, and landmark-based elastic (Fiji plugin, BUnwarpj)) registration procedure. (C) Alignment of the two H&E images of different resolutions from the same section (MSI layer). A linear transformation was performed based on the SSD of matched image features between the multi-resolution H&E images.



**Fig S2: Pipeline of the cell-biomarker matrix building (A); Related to Fig 4A.**

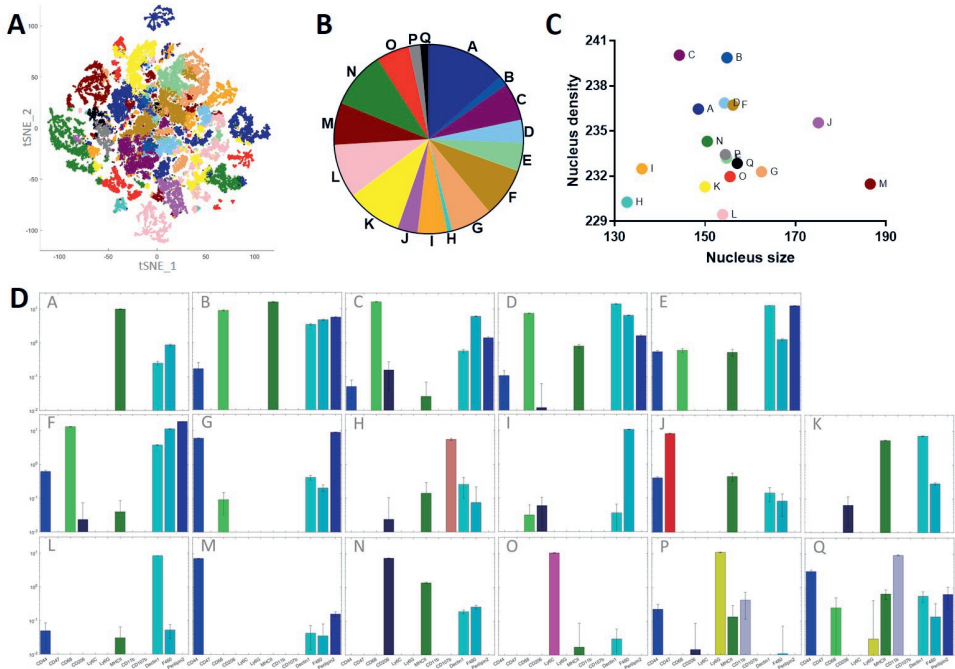
(B) Initial cell-biomarker matrix building. Cell segments (2) on an H&E image from fluorescent imaging layer (1), obtained from a watershed plugin of QuPath[24], were mapped to the corresponding multispectral images of the 12 phenotypic markers (3), and integrated into a cell-biomarker matrix  $X$  (4) based on the average intensity in individual cell segments. (C) Adjacency matrix of spillover compensation building. The adjacency matrix  $A$  (4) was constructed via 3 steps: for each cell, we (1) defined its neighbors based on Delaunay triangulation, (2) confirmed circumscribed ellipses of its neighboring cell segments, (3) calculated its adjacency coefficients based on the overlapping area between this cell and the ellipse of its neighbor. The improvement compared to the original CODEX approach (Goltsev et al., 2018) is illustrated in (3). For example, based on the improved compensation, the spillover coefficient of cell 1 (yellow square) from cell 2 (grey triangle) is the proportion of overlapping area (orange) to the area of cell 1, which is much smaller than the CODEX-defined coefficient  $\frac{1}{4}$ , defined by the shared edge (black dotted line) divided by the perimeter of cell 1. (D) An example (CD44) for explaining the algorithm for spillover compensation. The final corrected cell-biomarker matrix is the product of the original cell-biomarker matrix and the inverse of the adjacency matrix. (E) Comparison of distributions of cell intensities based on different compensation strategies. Compensation increases the number of cells with low intensities. However, compared to Goltsev *et al.*, our method is more conservative and milder, thereby reducing the risk of “overcompensation”. (F) Visualization of cell outlier removal.





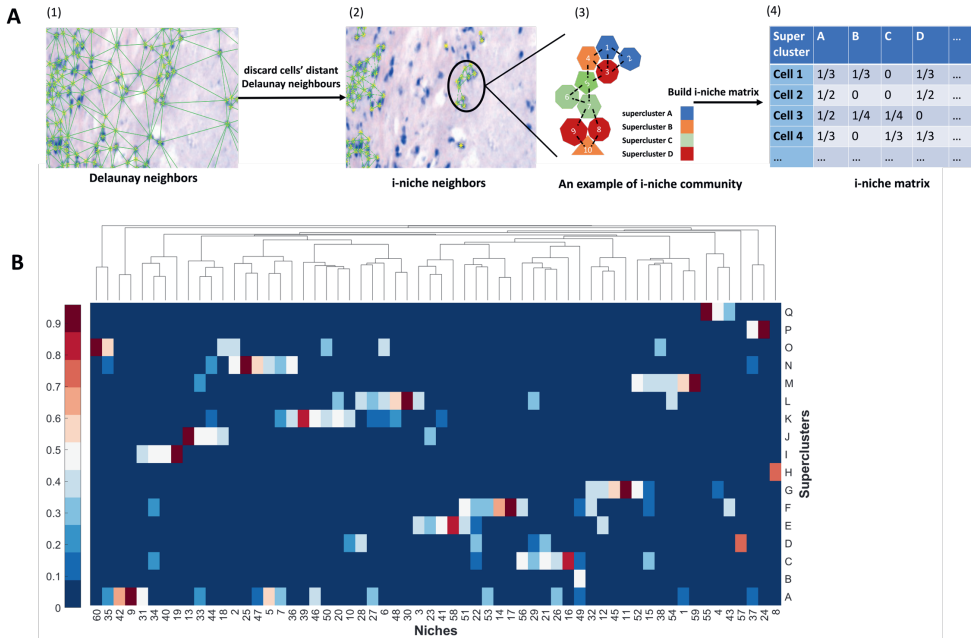
**Fig S3: Silhouette plots to evaluate the results of different clustering methods, and resulting radar plots to visualize marker expression patterns; Related to Fig 4A-B.**

(A) A silhouette plot splits cells into 70 clusters (y-axis) and ranks silhouette values of cells (x-axis) per cluster, showing how close each point in one cluster is to points in the neighbouring clusters, with a range from -1 to 1. The closer a cell's silhouette gets to +1, the further it is from neighbouring clusters. We selected the best performing approach, i.e., the method featuring the highest number of cell segments with positive silhouette values (cosine-based K-medoid clustering). (B) Radar plots indicate the mean expression patterns of 12 phenotypic markers for each of the 70 clusters identified in murine atherosclerotic plaques.



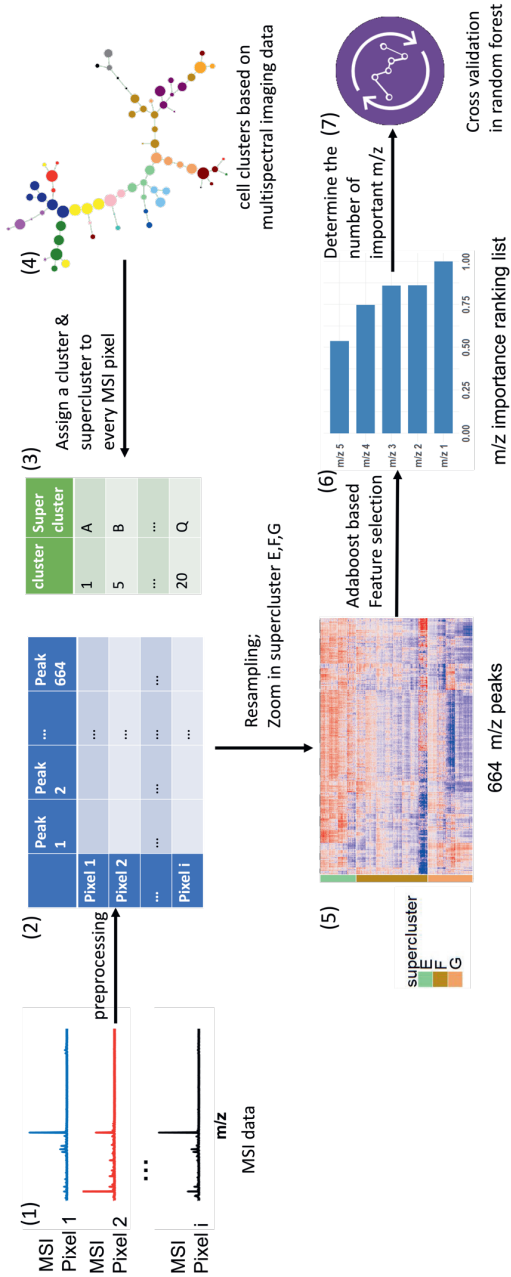
**Fig S4: Multispectral imaging-based myeloid heterogeneity and cell clustering; Related to Fig 4A-E.**

(A) The 17 superclusters are pseudo-colored in a tSNE plot. The proportions of cells assigned to each supercluster were visualized in (B) and their mean nuclear size and density, as defined in H&E stainings, were plotted in (C). (D) Mean expression patterns of all 12 phenotypic markers for the 17 superclusters were represented at a logarithmic scale to reveal dimmer marker expression not visible in the radar plots of Figure 4B. Error bars indicate SEM.



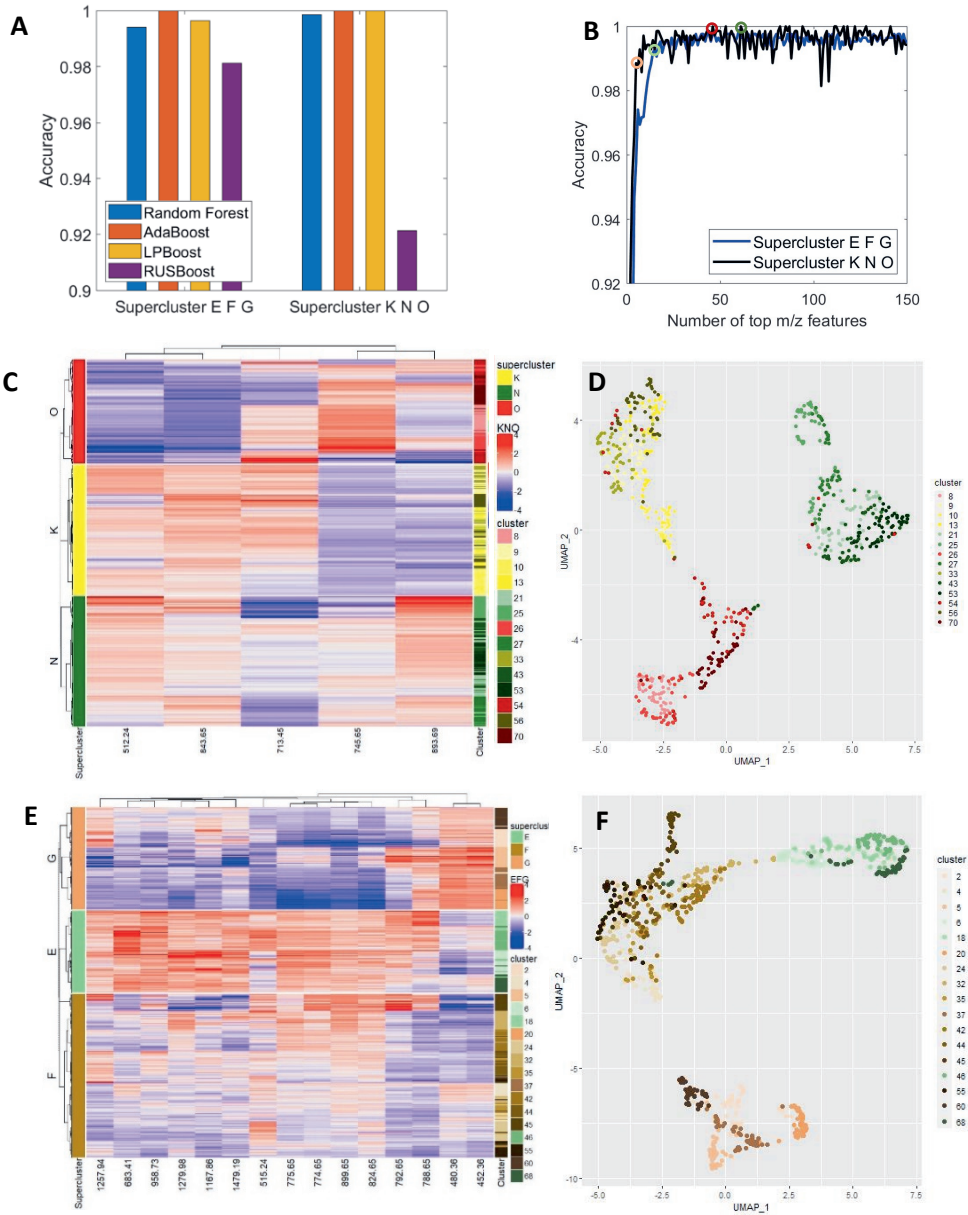
**Fig S5: i-niche cell community analysis; Related to Fig 6B.**

(A) i-niche matrix building. (1) & (2) An illustration of how i-niche neighbours of cell segments were defined in this study. (3) & (4) An example explaining the calculation of i-niche matrix. (B) Cell community identification through i-niche revealed 60 different niches that illustrate the significant co-localization of the 17 superclusters and cells belonging to the same tissue compartment.



**Fig S6: Illustration of m/z value selection procedure; Related to Fig 6E.**

(1) & (2) After preprocessing, 664 m/z peaks were selected from the raw MSI data, (3) & (4) Every pixel of MSI data was then assigned a certain cell cluster, as defined by multispectral imaging intensities in its adjacent section. (5), (6) & (7) Important m/z values were identified from several interesting superclusters (e.g., supercluster E, F, G) using Adaboost-based feature selection with 10-fold cross-validation.



**Fig S7: Supercluster-specific lipid feature selection; Related to Fig 7.**

(A) The accuracy of four ensemble learning methods in 10-fold cross-validation on supercluster E, F, G and supercluster K, N, O. (B) The 10-fold cross-validation accuracies of Adaboost with different numbers of selected top m/z features on supercluster E, F, G and supercluster K, N, O. The numbers of top lipid features with the highest accuracy are 46 (for K, N, O; dark red circle) and 60 (for E, F, G; dark green circle) separately. The smallest numbers of top lipid features with the performance which is not significantly different from the highest accuracy are 5 (for K, N, O; light red circle, represented in the heat map and UMAP (C-D)) and 15 (for E, F, G; light green circle (represented in

## Chapter 2

---

E-F)). (G) MS/MS identified 26 of the  $m/z$  values that differentially correlated with superclusters. Their relative intensities, measured over 11,995 pixels that colocalize with the six superclusters featured in figure 7, were displayed as a heatmap and intensities above a threshold of 0.7 were highlighted with a dot.

## References

1. Murray PJ, Allen JE, Biswas SK, Fisher EA, Gilroy DW, Goerdt S, Gordon S, Hamilton JA, Ivashkiv LB, Lawrence T et al: Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity* 2014, 41(1):14-20.
2. Guilliams M, De Kleer I, Henri S, Post S, Vanhoutte L, De Prijck S, Deswarte K, Malissen B, Hammad H, Lambrecht BN: Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. *J Exp Med* 2013, 210(10):1977-1992.
3. Guilliams M, Thierry GR, Bonnardel J, Bajenoff M: Establishment and Maintenance of the Macrophage Niche. *Immunity* 2020, 52(3):434-451.
4. Taube JM, Akturk G, Angelo M, Engle EL, Gnjatic S, Greenbaum S, Greenwald NF, Hedvat CV, Hollmann TJ, Juco J et al: The Society for Immunotherapy of Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation. *J Immunother Cancer* 2020, 8(1).
5. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, Black S, Nolan GP: Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* 2018, 174(4):968-981 e915.
6. Lin JR, Izar B, Wang S, Yapp C, Mei S, Shah PM, Santagata S, Sorger PK: Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *Elife* 2018, 7.
7. Radtke AJ, Kandov E, Lowekamp B, Speranza E, Chu CJ, Gola A, Thakur N, Shih R, Yao L, Yaniv ZR et al: IBEX: A versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues. *Proc Natl Acad Sci U S A* 2020, 117(52):33455-33465.
8. Giesen C, Wang HA, Schapiro D, Zivanovic N, Jacobs A, Hattendorf B, Schuffler PJ, Grolimund D, Buhmann JM, Brandt S et al: Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* 2014, 11(4):417-422.
9. Wei L, Chen Z, Shi L, Long R, Anzalone AV, Zhang L, Hu F, Yuste R, Cornish VW, Min W: Super-multiplex vibrational imaging. *Nature* 2017, 544(7651):465-470.
10. Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD, Levenson RM, Lowe JB, Liu SD, Zhao S et al: Multiplexed ion beam imaging of human breast tumors. *Nat Med* 2014, 20(4):436-442.
11. Eng CL, Lawson M, Zhu Q, Dries R, Kouloua N, Takei Y, Yun J, Cronin C, Karp C, Yuan GC et al: Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* 2019, 568(7751):235-239.



12. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ: Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019, 363(6434):1463-1467.
13. McDonnell LA, Heeren RM: Imaging mass spectrometry. *Mass Spectrom Rev* 2007, 26(4):606-643.
14. Goossens P, Gijbels MJ, Zerneck A, Eijgelaar W, Vergouwe MN, van der Made I, Vanderlocht J, Beckers L, Buurman WA, Daemen MJ et al: Myeloid type I interferon signaling promotes atherosclerosis by stimulating macrophage recruitment to lesions. *Cell Metab* 2010, 12(2):142-153.
15. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al: Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012, 9(7):676-682.
16. MATLAB and Image Processing Toolbox Release 2019a, The MathWorks, Inc., Natick, Massachusetts, United States.
17. Zuiderveld K: Contrast Limited Adaptive Histogram Equalization. *Graphic Gems IV* San Diego: Academic Press Professional 1994:474–485.
18. Lim JS: Two-Dimensional Signal and Image Processing. Englewood Cliffs NJ,: Prentice Hall; 1990.
19. Bradley DaR, G: Adaptive thresholding using the integral image. *Journal of graphics tools* 2007:12(12), 13-21.
20. Mattes D, Haynor DR, Vesselle H, Lewellyn TK, Eubank W: Nonrigid multimodality image registration. In: *Medical imaging 2001: image processing*: 2001. 1609–1620.
21. Rublee E, Rabaud V, Konolige K, Bradski G: ORB: An efficient alternative to SIFT or SURF. In: *2011 International conference on computer vision*: 2011. 2564–2571.
22. Hartley R, Zisserman A: *Multiple view geometry in computer vision*: Cambridge university press; 2003.
23. Arganda-Carreras I, Sorzano COS, Marabini R, Carazo JMa, Ortiz-de-Solorzano C, Kybic J: Consistent and elastic registration of histological sections using vector-spline regularization. In: *International Workshop on Computer Vision Approaches to Medical Image Analysis*: 2006. 85–95.
24. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG et al: QuPath: Open source software for digital pathology image analysis. *Scientific reports* 2017, 7:1–7.

25. Park H-S, Jun C-H: A simple and fast algorithm for K-medoids clustering. *Expert systems with applications* 2009, 36:3336–3341.
26. Vesanto J, Alhoniemi E: Clustering of the self-organizing map. *IEEE Transactions on neural networks* 2000, 11:586–600.
27. Samusik N, Good Z, Spitzer MH, Davis KL, Nolan GP: Automated mapping of phenotype space with single-cell data. *Nature methods* 2016, 13:493–496.
28. Ng AY, Jordan MI, Weiss Y: On spectral clustering: Analysis and an algorithm. In: *Advances in neural information processing systems: 2002*. 849–856.
29. Rousseeuw PJ: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 1987, 20:53–65.
30. Qiu P, Simonds EF, Bendall SC, Gibbs KD, Jr., Bruggner RV, Linderman MD, Sachs K, Nolan GP, Plevritis SK: Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol* 2011, 29(10):886–891.
31. MATLAB and Statistics and Machine Learning Toolbox Release 2019a, The MathWorks, Inc., Natick, Massachusetts, United States.
32. Freund Y, Schapire RE: A decision-theoretic generalization of on-line learning and an application to boosting. In: 1995; Berlin, Heidelberg. Springer Berlin Heidelberg: 23–37.
33. Cao J, Goossens P, Martin-Lorenzo M, Dewez F, Claes BSR, Biessen EAL, Heeren RMA, Balluff B: Atheroma-Specific Lipids in *Ildlr(-/-)* and *apoe(-/-)* Mice Using 2D and 3D Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging. *J Am Soc Mass Spectrom* 2020, 31(9):1825–1832.
34. MATLAB and Bioinformatics Toolbox Release 2019a, The MathWorks, Inc., Natick, Massachusetts, United States.
35. Morris JS, Coombes KR, Koomen J, Baggerly KA, Kobayashi R: Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics* 2005, 21:1764–1775.
36. Breiman L: Random Forests. *Machine Learning* 2001, 45:5–32.
37. Seiffert C, Khoshgoftaar TM, Hulse JV, Napolitano A: RUSBoost: Improving classification performance when training data is skewed. In: 2008 19th International Conference on Pattern Recognition: 8–11 Dec. 2008 2008. 1–4.
38. Warmuth MK, Liao J, Rätsch G: Totally corrective boosting algorithms that maximize the margin. In: *Proceedings of the 23rd international conference on Machine learning; Pittsburgh, Pennsylvania, USA*. Association for Computing Machinery 2006: 1001–1008.

39. Gu Z, Eils R, Schlesner M: Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016, 32(18):2847-2849.
40. McInnes L, Healy J, Saul N, Grossberger L: UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* 2018, 3:861.
41. Ellis SR, Paine MRL, Eijkel GB, Pauling JK, Husen P, Jervelund MW, Hermansson M, Ejsing CS, Heeren RMA: Automated, parallel mass spectrometry imaging and structural identification of lipids. *Nat Methods* 2018, 15(7):515-518.
42. Tortorella S, Tiberi P, Bowman AP, Claes BSR, Scupakova K, Heeren RMA, Ellis SR, Cruciani G: LipostarMSI: Comprehensive, Vendor-Neutral Software for Visualization, Data Analysis, and Automated Molecular Identification in Mass Spectrometry Imaging. *J Am Soc Mass Spectrom* 2020, 31(1):155-163.
43. Pauling JK, Hermansson M, Hartler J, Christiansen K, Gallego SF, Peng B, Ahrends R, Ejsing CS: Proposal for a common nomenclature for fragment ions in mass spectra of lipids. *PLoS One* 2017, 12(11):e0188394.
44. Lewis SM, Williams A, Eisenbarth SC: Structure and function of the immune system in the spleen. *Sci Immunol* 2019, 4(33).
45. Davies LC, Jenkins SJ, Allen JE, Taylor PR: Tissue-resident macrophages. *Nat Immunol* 2013, 14(10):986-995.
46. Hey YY, Tan JK, O'Neill HC: Redefining Myeloid Cell Subsets in Murine Spleen. *Front Immunol* 2015, 6:652.
47. Sorrelle N, Ganguly D, Dominguez ATA, Zhang Y, Huang H, Dahal LN, Burton N, Ziemys A, Brekken RA: Improved Multiplex Immunohistochemistry for Immune Microenvironment Evaluation of Mouse Formalin-Fixed, Paraffin-Embedded Tissues. *J Immunol* 2019, 202(1):292-299.
48. Lim HY, Lim SY, Tan CK, Thiam CH, Goh CC, Carbajo D, Chew SHS, See P, Chakarov S, Wang XN et al: Hyaluronan Receptor LYVE-1-Expressing Macrophages Maintain Arterial Tone through Hyaluronan-Mediated Regulation of Smooth Muscle Cell Collagen. *Immunity* 2018, 49(6):1191.
49. Williams JW, Zaitsev K, Kim KW, Ivanov S, Saunders BT, Schrank PR, Kim K, Elvington A, Kim SH, Tucker CG et al: Limited proliferation capacity of aortic intima resident macrophages requires monocyte recruitment for atherosclerotic plaque progression. *Nat Immunol* 2020.
50. Williams JW, Zaitsev K, Kim KW, Ivanov S, Saunders BT, Schrank PR, Kim K, Elvington A, Kim SH, Tucker CG et al: Limited proliferation capacity of aortic intima resident macrophages requires monocyte recruitment for atherosclerotic plaque progression. *Nat Immunol* 2020, 21(10):1194-1204.

51. Colin S, Chinetti-Gbaguidi G, Staels B: Macrophage phenotypes in atherosclerosis. *Immunol Rev* 2014, 262(1):153-166.
52. Cochain C, Vafadarnejad E, Arampatzi P, Pelisek J, Winkels H, Ley K, Wolf D, Saliba AE, Zernecke A: Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis. *Circ Res* 2018, 122(12):1661-1674.
53. Lin JD, Nishi H, Poles J, Niu X, McCauley C, Rahman K, Brown EJ, Yeung ST, Vozhilla N, Weinstock A et al: Single-cell analysis of fate-mapped macrophages reveals heterogeneity, including stem-like properties, during atherosclerosis progression and regression. *JCI Insight* 2019, 4(4).
54. van Kuijk K, Kuppe C, Betsholtz C, Vanlandewijck M, Kramann R, Sluimer JC: Heterogeneity and plasticity in healthy and atherosclerotic vasculature explored by single-cell sequencing. *Cardiovasc Res* 2019, 115(12):1705-1715.
55. Winkels H, Ehinger E, Vassallo M, Buscher K, Dinh HQ, Kobiyama K, Hamers AAJ, Cochain C, Vafadarnejad E, Saliba AE et al: Atlas of the Immune Cell Repertoire in Mouse Atherosclerosis Defined by Single-Cell RNA-Sequencing and Mass Cytometry. *Circ Res* 2018, 122(12):1675-1688.
56. Kim K, Shim D, Lee JS, Zaitsev K, Williams JW, Kim KW, Jang MY, Seok Jang H, Yun TJ, Lee SH et al: Transcriptome Analysis Reveals Nonfoamy Rather Than Foamy Plaque Macrophages Are Proinflammatory in Atherosclerotic Murine Models. *Circ Res* 2018, 123(10):1127-1142.
57. Cole JE, Park I, Ahern DJ, Kassiteridi C, Danso Abeam D, Goddard ME, Green P, Maffia P, Monaco C: Immune cell census in murine atherosclerosis: cytometry by time of flight illuminates vascular myeloid cell diversity. *Cardiovasc Res* 2018, 114(10):1360-1371.
58. Zernecke A, Winkels H, Cochain C, Williams JW, Wolf D, Soehnlein O, Robbins CS, Monaco C, Park I, McNamara CA et al: Meta-Analysis of Leukocyte Diversity in Atherosclerotic Mouse Aortas. *Circ Res* 2020, 127(3):402-426.
59. Duong H, Han M: A multispectral LED array for the reduction of background autofluorescence in brain tissue. *J Neurosci Methods* 2013, 220(1):46-54.
60. Kingsley K, Carroll K, Huff JL, Plopper GE: Photobleaching of arterial autofluorescence for immunofluorescence applications. *Biotechniques* 2001, 30(4):794-797.
61. Viegas MS, Martins TC, Seco F, do Carmo A: An improved and cost-effective methodology for the reduction of autofluorescence in direct immunofluorescence studies on formalin-fixed paraffin-embedded tissues. *Eur J Histochem* 2007, 51(1):59-66.

62. van der Maaten L, and Hinton, G: Visualizing Data using t-SNE. *Journal of Machine Learning Research* 2008, 9:2579--2605.
63. Robbins CS, Hilgendorf I, Weber GF, Theurl I, Iwamoto Y, Figueiredo JL, Gorbатов R, Sukhova GK, Gerhardt LM, Smyth D et al: Local proliferation dominates lesional macrophage accumulation in atherosclerosis. *Nat Med* 2013, 19(9):1166-1172.
64. Ensan S, Li A, Besla R, Degousee N, Cosme J, Roufaiel M, Shikatani EA, El-Maklizi M, Williams JW, Robins L et al: Self-renewing resident arterial macrophages arise from embryonic CX3CR1(+) precursors and circulating monocytes immediately after birth. *Nat Immunol* 2016, 17(2):159-168.
65. Epelman S, Lavine KJ, Beaudin AE, Sojka DK, Carrero JA, Calderon B, Brija T, Gautier EL, Ivanov S, Satpathy AT et al: Embryonic and adult-derived resident cardiac macrophages are maintained through distinct mechanisms at steady state and during inflammation. *Immunity* 2014, 40(1):91-104.
66. Giannotti KC, Weinert S, Viana MN, Leiguez E, Araujo TLS, Laurindo FRM, Lomonte B, Braun-Dullaeus R, Teixeira C: A Secreted Phospholipase A2 Induces Formation of Smooth Muscle Foam Cells Which Transdifferentiate to Macrophage-Like State. *Molecules* 2019, 24(18).
67. Feil S, Fehrenbacher B, Lukowski R, Essmann F, Schulze-Osthoff K, Schaller M, Feil R: Transdifferentiation of vascular smooth muscle cells to macrophage-like cells during atherogenesis. *Circ Res* 2014, 115(7):662-667.
68. Shankman LS, Gomez D, Cherepanova OA, Salmon M, Alencar GF, Haskins RM, Swiatlowska P, Newman AA, Greene ES, Straub AC et al: KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nat Med* 2015, 21(6):628-637.
69. Spitzer MH, Nolan GP: Mass Cytometry: Single Cells, Many Features. *Cell* 2016, 165(4):780-791.
70. Combadiere C, Potteaux S, Rodero M, Simon T, Pezard A, Esposito B, Merval R, Proudfoot A, Tedgui A, Mallat Z: Combined inhibition of CCL2, CX3CR1, and CCR5 abrogates Ly6C(hi) and Ly6C(lo) monocytosis and almost abolishes atherosclerosis in hypercholesterolemic mice. *Circulation* 2008, 117(13):1649-1657.
71. Okabe Y, Medzhitov R: Tissue-specific signals control reversible program of localization and functional polarization of macrophages. *Cell* 2014, 157(4):832-844.
72. Qin Y, Garrison BS, Ma W, Wang R, Jiang A, Li J, Mistry M, Bronson RT, Santoro D, Franco C et al: A Milieu Molecule for TGF-beta Required for Microglia Function in the Nervous System. *Cell* 2018, 174(1):156-171 e116.
73. Sakai M, Troutman TD, Seidman JS, Ouyang Z, Spann NJ, Abe Y, Ego KM, Bruni CM, Deng Z, Schlachetzki JCM et al: Liver-Derived Signals Sequentially Reprogram

---

Myeloid Enhancers to Initiate and Maintain Kupffer Cell Identity. *Immunity* 2019, 51(4):655-670 e658.

74. Visscher M, Moerman AM, Burgers PC, Van Beusekom HMM, Luider TM, Verhagen HJM, Van der Steen AFW, Van der Heiden K, Van Soest G: Data Processing Pipeline for Lipid Profiling of Carotid Atherosclerotic Plaque with Mass Spectrometry Imaging. *J Am Soc Mass Spectrom* 2019, 30(9):1790-1800.

75. Bot M, de Jager SC, MacAleese L, Lagrauw HM, van Berkel TJ, Quax PH, Kuiper J, Heeren RM, Biessen EA, Bot I: Lysophosphatidic acid triggers mast cell-driven atherosclerotic plaque destabilization by increasing vascular inflammation. *J Lipid Res* 2013, 54(5):1265-1274.

76. Sun X, Stewart DA, Sandhu R, Kirk EL, Pathmasiri WW, McRitchie SL, Clark RF, Troester MA, Sumner SJ: Correlated metabolomic, genomic, and histologic phenotypes in histologically normal breast tissue. *PLoS One* 2018, 13(4):e0193792.

77. Scupakova K, Dewez F, Walch AK, Heeren RMA, Balluff B: Morphometric Cell Classification for Single-Cell MALDI-Mass Spectrometry Imaging. *Angew Chem Int Ed Engl* 2020.

78. Guillermier C, Doherty SP, Whitney AG, Babaev VR, Linton MF, Steinhauser ML, Brown JD: Imaging mass spectrometry reveals heterogeneity of proliferation and metabolism in atherosclerosis. *JCI Insight* 2019, 4(11).



## Chapter 3

Spatio-temporal metabolomics reveals metabolic changes in the early stage of non-alcoholic fatty liver disease in mice

EMBARGOED

**Chang Lu**, Jianhua Cao, Kristiaan Wouters, Marion Gijbels, Marjo M.P.C. Donners, Benjamin Balluff, Evgueni Smirnov, Rachel Cavill, Joël M.H. Karel, Ron M.A. Heeren, Erik A.L. Biessen<sup>†</sup>, Pieter Goossen<sup>†</sup>

*In preparation*

<sup>†</sup> Authors contributed equally



# Chapter 4

Cardiovascular disease but not health is hallmarked by sex-specific cytokine signaling pathways in circulating monocytes

**Chang Lu<sup>†</sup>**, Marjo M.P.C. Donnerst<sup>†</sup>, Joël M.H. Karel, Hetty de Boer, Anton Jan van Zonneveld, Hester den Ruijter, Wouter Jukema, Adriaan Kraaijeveld, Johan Kuiper, Gerard Pasterkamp, Rachel Cavill, Javier Perales-Patón, Ele Ferrannini, Pieter Goossens, Erik A.L. Biessen

*Manuscript submitted*

<sup>†</sup> Authors contributed equally



## Abstract

While sex is a major risk factor in cardiovascular disease (CVD) and important determinant of its clinical manifestation, the underlying pathogenic mechanisms are still largely unknown. In this study, we comprehensively dissected sex-specific transcriptional differences in circulating monocytes in a CVD cohort to identify sex specific signalling pathways contributing to the disease.

We generated sex-biased gene expression signatures (GES) by comparing monocytes of a cohort of male versus female coronary artery disease (CAD) patients (n=450) from the Center for Translational Molecular Medicine–Circulating Cells Cohort. Gene set enrichment analysis demonstrated that monocytes from female CAD patients carry a stronger chemotaxis and migratory signature than those from male CAD patients. We then inferred cytokine signalling activities by ridge regression using the CytoSig database of 51 cytokine and growth factor regulation profiles. As compared to males, monocytes from females feature a higher level of activation of EGF, IFN1, VEGF, GM-CSF and CD40L pathways, whereas IL4, INS, HMGB signalling was preferentially activated in males. This gender-imbalance was not observed in healthy subjects, as shown for independent healthy subject monocyte cohort (GSE56034, n=485). More pronounced GM-CSF signalling in monocytes of female CAD patients was confirmed by the significant enrichment of GM-CSF-activated monocyte signature in the female gene signature. Surprisingly, the enhanced signalling activities could not be explained by increased plasma levels of the corresponding ligands, suggesting intrinsic sex differences in monocyte signalling regulation. Consistent with our findings, a regulatory network constructed to link the main (female activated) transcription factors to the identified sex-specific cytokine pathways, revealed jun-B as activator of all female-specific pathways but IFN1, while being suppressed by male-activated IL-4.

In conclusion, we observed overt CAD-specific sex differences in monocyte transcriptional profiles and cytokine-or growth factor induced responses, which provide insights into underlying mechanisms of sex differences in CVD.

## Introduction

Although ischemic cardiovascular diseases (CVD), including myocardial infarction and stroke, is the leading cause of death in both men and women [1], the sexual dimorphism in prevalence and presentation of ischemic heart disease (IHD) is well-known [2]. IHD incidence in young women was reported to be lower than in men, however with increasing age, particularly after menopause, this risk profile is reversed [3]. Estrogen is thought to be a major contributor to these sex differences, by improving arterial function and lipoprotein profile, amongst others. Indeed, several major CVD risk factors, such as blood pressure, diabetes, alcohol, markedly differ between sexes, especially before menopause. However, CVD risk factors were seen to be more impactful in females than in males (relative odds ratios of 1.5, 1.6 and 1.3 for hypertension, diabetes mellitus and smoking, respectively). This points to sexual dimorphism of the risk factor response [1, 4]. This may explain the higher prevalence of IHD for females than males with a similar risk factor profile [4, 5].

Culminating evidence points out that chronic inflammation, effected by macrophages and their precursors, monocytes, plays an important causal role in IHD; it contributes to atherosclerotic plaque formation and rupture amongst others [6–8]. Indeed, the CANTOS trial showed a significantly reduced secondary event rate after IL1 receptor intervention [9]. Macrophages and monocytes are major drivers of atherosclerosis [10, 11] and IHD injury repair. They are plastic cells that adapt to their environment [12], which may well differ between males and females. This has prompted several groups to study sex-specific differences in macrophages and monocytes [13–15], showing differences in monocyte numbers [16, 17] or inflammatory response [18, 19]. However, still little is known about sex differences in signalling and transcriptional profiles of monocytes in CAD patients.

This study has interrogated a monocyte cohort from the Center for Translational Molecular Medicine (CTMM) – Circulating Cells Cohort that were presented to the Maastricht University Medical Center, The Netherlands, for sex-related differences in signalling and transcriptional makeup. We first generated the gene expression signatures (GES) for male and female CAD patient monocytes. Gene set enrichment analysis (GSEA) was then used for interpreting sex-specific differential genes, and cytokine and transcription factor (TF) activities were inferred based on Cytosig [20] and Dorothea [21] databases respectively. Finally, CAD-specificity of pathways and TFs and sex-specific differences of plasma cytokines were validated in an independent healthy subject cohort and a CAD plasma proteomics cohort, respectively.

## Methods

### CTMM Cohort

#### Human Blood Samples from Center for Translational Molecular Medicine

A total of 460 CAD patients from the Center for Translational Molecular Medicine (CTMM) Circulating Cells Cohort were used for analysing sex differences in CAD. CAD patients were defined as subjects diagnosed with stable angina, unstable angina, Non-ST elevation Myocardial Infarction (NSTMI), or silent ischemia. All participants provided written informed consent prior to participation. This study was approved by the Institutional Medical Ethical Review Board of the University Medical Center Utrecht, The Netherlands.

### Cell Isolation

Upon inclusion, blood samples were obtained in ethylenediaminetetraacetic acid (EDTA) anti-coagulated vacuum tubes and processed as previously described [22]. Briefly, blood was centrifuged at 156x g for 15 minutes, plasma was withdrawn, and CD14<sup>+</sup> monocytes were isolated using Ficoll-Paque Plus (Sigma), magnetic bead isolation (BD Bioscience). Plasma and monocytes were stored at -80°C until needed.

### Flow Cytometry

To analyse and quantify monocyte subsets, 50 µl of EDTA anti-coagulated whole blood was analysed by flow cytometry (Beckman Coulter FC 500) as previously described [22] using the antibody panel shown in table 1. Lymphocytes and granulocytes were gated based on their scatter properties and confirmed using lymphocyte and granulocyte specific antibodies. Monocytes were identified based on their scatter properties and positive CD14 staining. Expression of surface markers was quantified by relative marker expression and mean fluorescence intensity (MFI).

**Table 1 : Antibodies for flow cytometry**

Antibody	Company	Dilution
CD14 PC7	BD Biosciences	1:10
CD16 PC5	BD Biosciences	1:10
CCR2 PE	BD Biosciences	1:5
CX <sub>3</sub> CR1 FITC	BD Biosciences	1:5

### RNA Isolation and Micro-Array Analysis

RNA isolation and micro-array data generation of monocyte samples were completed by AROS (Denmark). In brief, RNA was isolated using Illumina TotalPrep RNA Amplification Kit (Illumina, San Diego, CA, USA) and cDNA was produced. Next, labeled cRNA was prepared and used on the array for hybridization. Hybridized chips were scanned by Illumina BeadStation (Illumina, Inc., San Diego, CA, U.S.A.).

Raw image analysis and signal extraction was performed with Illumina Beadstudio Gene Expression software with default settings (no background subtraction) and no normalization.

### Data preprocessing

The CTMM microarray data were exported to R v3.6.3 after quality control using GenomeStudio software. Ten of 460 profiles were excluded due to their low number of detected genes (< 20% significantly detected genes) or large distance to the mean (>100) based on quality control files obtained from GenomeStudio. Log<sub>2</sub> transformation and Robust Spline Normalization (RSN) were performed using lumi R package [23]. For the resulting cohort of 450 profiles, to reduce the effect of background noise, the genes with low intensity (more than 90% of patients' log<sub>2</sub> expressions lower than 7.5) were filtered out. To inspect the preprocessing results, batch correction was performed using the sva R package before visualization [24]. Eventually, 450 microarrays from 316 men and 134 women with CAD, including 9,092 genes were used for differential gene expression analysis. All probes have been converted to official gene symbols. The schematic diagram of the cohort build-up and workflow of the study is shown in **Fig 1A**.

### Validation Cohorts

A monocyte cohort from Harvard Medical School, Boston, USA (GSE56034) was downloaded from GEO database for validation. CD14+CD16- monocytes were sorted from peripheral blood mononuclear cells (PBMCs) of 485 healthy individuals (272 women and 213 men). mRNAs of these monocytes were then profiled on Affymetrix GeneChip Human Gene ST 1.0 microarrays. Log<sub>2</sub> transformation and quantile normalization were used for the data pre-processing. After filtering out the genes with low intensity (more than 95% of patients' expressions lower than 100), 13,556 genes were left for differential gene expression analysis. The mean and standard deviation of age for healthy males and females are shown in **Table S1**.

Second, proteomics data from a cohort including 187 CAD patients and 341 healthy subjects [25] were used to study whether the augmented signalling activities were reflecting elevated levels of the corresponding ligands in plasma. The demographics of this cohort is shown in **Table S2**.

### Differential gene expression

The sex-biased genes were obtained from the pre-processed gene expression profile from 450 patients using limma package [26]. More specifically, for each gene *g*, a linear model was built between the gene expression values of *g* and genders of all patients. Significant levels were calculated based on the moderated t-statistics, which used empirical Bayes methods to obtain posterior variance estimators (for details see the limma tutorial). Ages and batches were added into the linear model as covariates for removing their effects. The male/female-biased genes in this study

are the genes that shows up/down regulation by comparing males and females. The GES here is the log<sub>2</sub> fold change of expression between male and female subjects. P-values were adjusted by false discovery rate (FDR).

### Gene Set enrichment analyses (GSEA)

GSEA of Gene Ontology (including Biological Process (BP), Molecular Function (MF), and Cellular Component (CC)), KEGG, WikiPathway, and Reactome for sex-related GES were performed using the ClusterProfiler R-package [27]. The p-values were adjusted for multiple comparison by FDR, and the cut-off was set to 0.05 to avoid presenting false discovery significant terms. Gene ratio stands for the percentage of genes associated with the given GO/Pathway term to the total number of genes in that term.

### Cytokine signalling activity inference

Cytokine signalling activities were calculated based on the database of target genes modulated by cytokines from Cytosig platform (<https://cytosig.ccr.cancer.gov/>). The Cytosig database including in total of 51 cytokines (and growth factors) and 6466 response genes. As suggested in [20], considering the signature collinearity of cytokine profiles, we utilized ridge regression (linearridge function in 'ridge' R-package), as the predictive model of cytokine signalling activities. Only 4,130 genes that were overlapped in Cytosig database and the sex-biased GES CTMM cohort were used in this model. Composite profiles of cytokine response were the explanatory variables, and the sex-biased GES was the response variable. The regression coefficients represent cytokine target activities. The ridge regression parameter is chosen automatically using 'ridge' R package proposed by [28]. The p-values are computed using the significance test of [29]. P-values were adjusted using FDR.

### TF activity inference

In this study, we predicted both activities of TFs that drive sex-biased GES and that drive cytokine-modulated gene profiles, based on the TF-target interaction network provided by DoRothEA [21]. Specifically, we first filtered out the TF–target interactions with low level of evidence (confidence score D and E). TF activities were then calculated using Weighted Connectivity Score (WTCS), a bi-directional version of weighted Kolmogorov-Smirnov enrichment statistic (ES) described in [30]. For a given TF, assuming  $q_{pos}$  and  $q_{neg}$  are the activated and inhibited target sets of TF  $q$  in DoRothEA, its activity  $w$  on a cytokine-modulated profile (or sex-biased GES)  $s$  is as follows:

$$w_{q,s} = \begin{cases} (NES_{pos} - NES_{neg})/2, & \text{if } \text{sgn}(NES_{pos}) \neq \text{sgn}(NES_{neg}) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $NES_{pos}$  is the normalized enrichment score of  $q_{pos}$  in  $s$  and  $NES_{neg}$  is the normalized enrichment score of  $q_{neg}$  in  $s$ . P-value is calculated based on 10000-time gene-wise permutation test. Finally, we selected female-activated TFs and sex-activated cytokines to construct a TF-cytokine bipartite graph (**Fig 4B**), where the edges between a TF and a cytokine represents the activity of this TF on its target genes under the modulation of this cytokine.

### Visualization

Cytoscape was applied to visualize the WTCS between cytokines and TFs. All other plots were generated using ggplot2 package in R.

### Validation on GM-CSF-induced RNA-Seq dataset

Human CD14+ peripheral blood monocytes were stimulated for 24 h with GM-CSF (X ng/ml; n=3) and expression profiles were assessed by RNA-Seq. The parental monocytes were used as controls. DEG analysis revealed 132 upregulated genes in GM-CSF stimulated vs unstimulated monocytes ( $\log_2FC > 0.5$ ;  $P_{adj} < 0.05$ ). This DEG geneset was used to perform GSEA to check if the up-regulated genes were enriched in the female-biased gene set from GES in the CTMM monocyte cohort.

### Statistical analysis

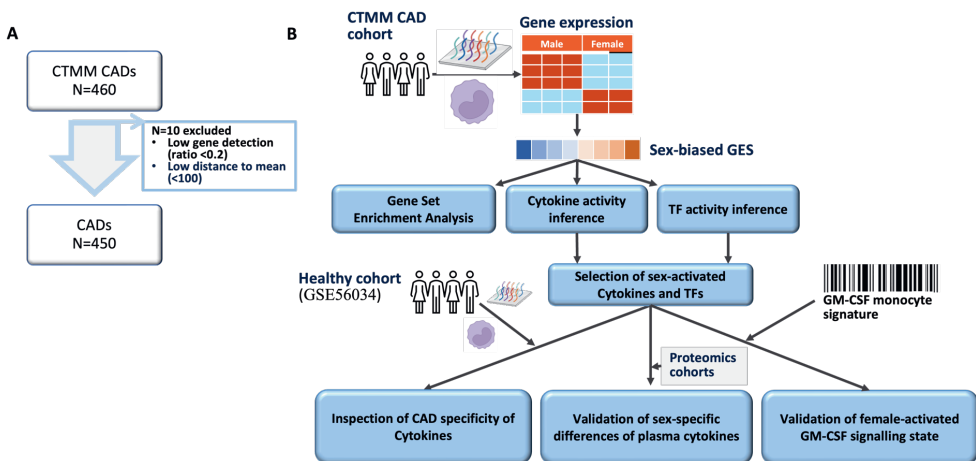
Unpaired two-sided t-tests with non-equal variance were applied to compare the gene expressions from females and males in **Fig 5B** and **Fig 3C**. Significant level is denoted by \*p-value < 0.05, \*\*p-value < 0.01, \*\*\*p-value < 0.001. All statistical analyses were performed in R (v3.6.3).



## Results

### Monocytes from female CAD patients show stronger migratory capacity than males

To identify sex differences in monocyte transcriptomes of CAD patients, we compared microarray data from monocytes of male (n=316) and female (n=134) CAD patients within the CTMM cohort. The workflow of this study is shown in Figure 1B. Clinical information of this cohort is listed in Table 2. Except for higher HDL and lower creatinine in females, there were no significant differences in other clinical characteristics between males and females. Similarly, flow cytometry analyses confirmed that the number of classical, non-classical, and intermediate monocytes were essentially similar between male and female CAD patients (**Fig 2A**). We observed 48 male- and 107 female-biased genes (**Fig 2B**), including several notorious X or Y chromosome-linked genes, such as PRKY, RPS4Y1, EIFAY and XIST.

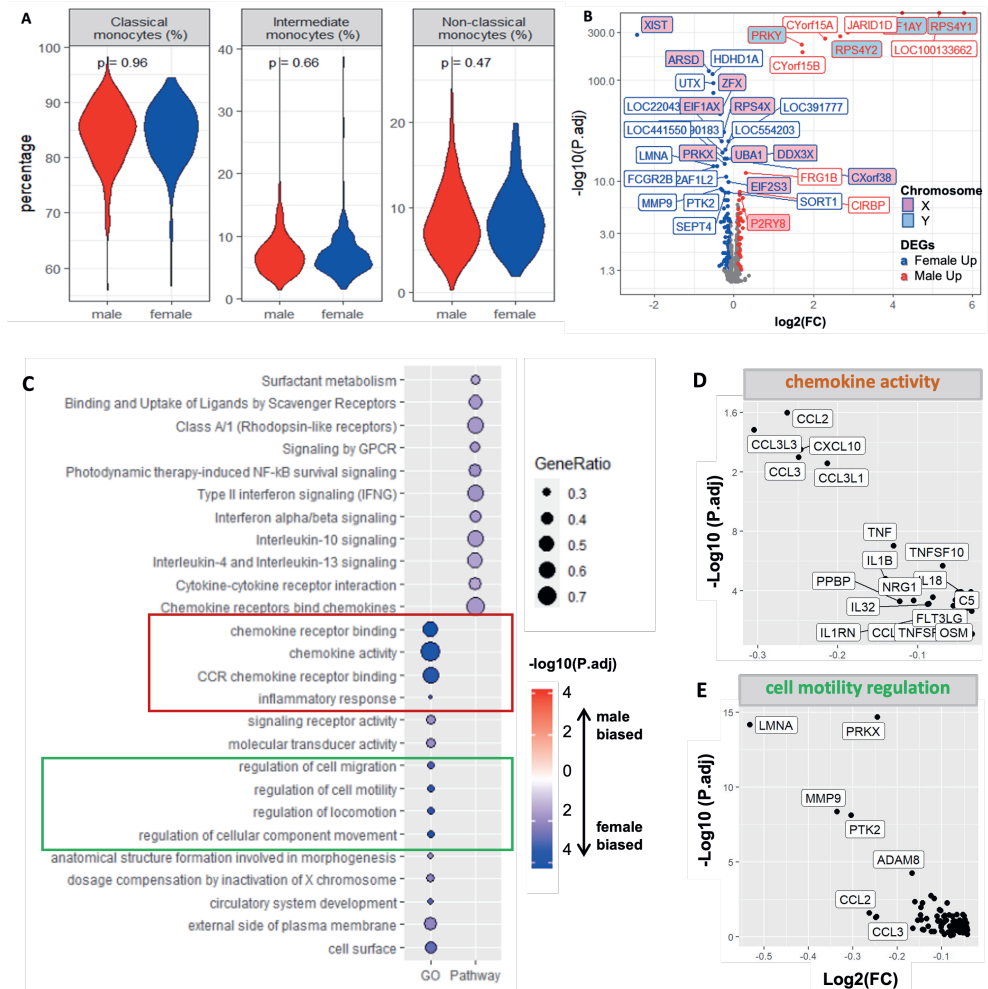


**Figure 1: CTMM cohort and analysis procedure.** (A) A schematic of cohort build-up; (B) The schematic diagram of the analysis.

We then performed GSEA based on sex-biased signatures to obtain a more detailed view of the functions of sex-biased genes (**Fig 2C**). While male-biased genes did not show clear, significantly enriched GO terms, female-biased genes were significantly enriched in pathways related to cellular movement and chemokine- or interleukin signaling. Analysis of the genes responsible for the observed enrichment of chemokine activity (**Fig 2D**) term revealed that all GO term chemokine activity genes (e.g. CCL2, CCL3, CXCL10) were upregulated in female CAD patients compared to males. Similarly, LMNA, PRKX, MMP9 and PTK2 are the top 4 differentially expressed ‘regulation of cell motility’ term members (**Fig 2E**), next to CCL2 and CCL3

**Table 2: Demographics of females and males who participated in the CTMM cohort expressed as mean  $\pm$  SD or frequencies (%)**

riskfactor	Male (n=316)	Female (n=134)	p-value
HDL (mmol/dL)	1.06+0.26	1.23+0.36	<b>9.48E-05</b>
Triglyceride (mmol/dL)	1.57+0.93	1.54+0.87	0.76
Systolic Blood Pressure (mmHg)	134.81+18.92	137.89+22.83	0.17
LDL (mmol/dL)	2.47+0.98	2.71+1.12	0.08
Smoking (Pack years)	25.41+22.28	26.85+24.33	0.56
BMI (kg/m <sup>2</sup> )	27.4+3.89	27.7+5.28	0.56
glucose	6.51+2.18	6.76+2.09	0.32
Age	62.48+10.02	62.92+10.66	0.69
Heart rate	67.49+13.53	69.42+12.44	0.14
Diastolic Blood Pressure (mmHg)	78.84+11.15	76.46+12.32	0.06
Creatinine (mmol/dL)	91.86+53.76	74.99+19.84	<b>2.24E-06</b>
Hypertension	192 (60.76%)	93 (69.4%)	0.08
Renal failure	6, (1.9%)	3, (2.24%)	0.81
Diabetes Mellitus	64, (20.25%)	34, (25.37%)	0.26
Current Smoker	66, (20.89%)	33, (24.63%)	0.38



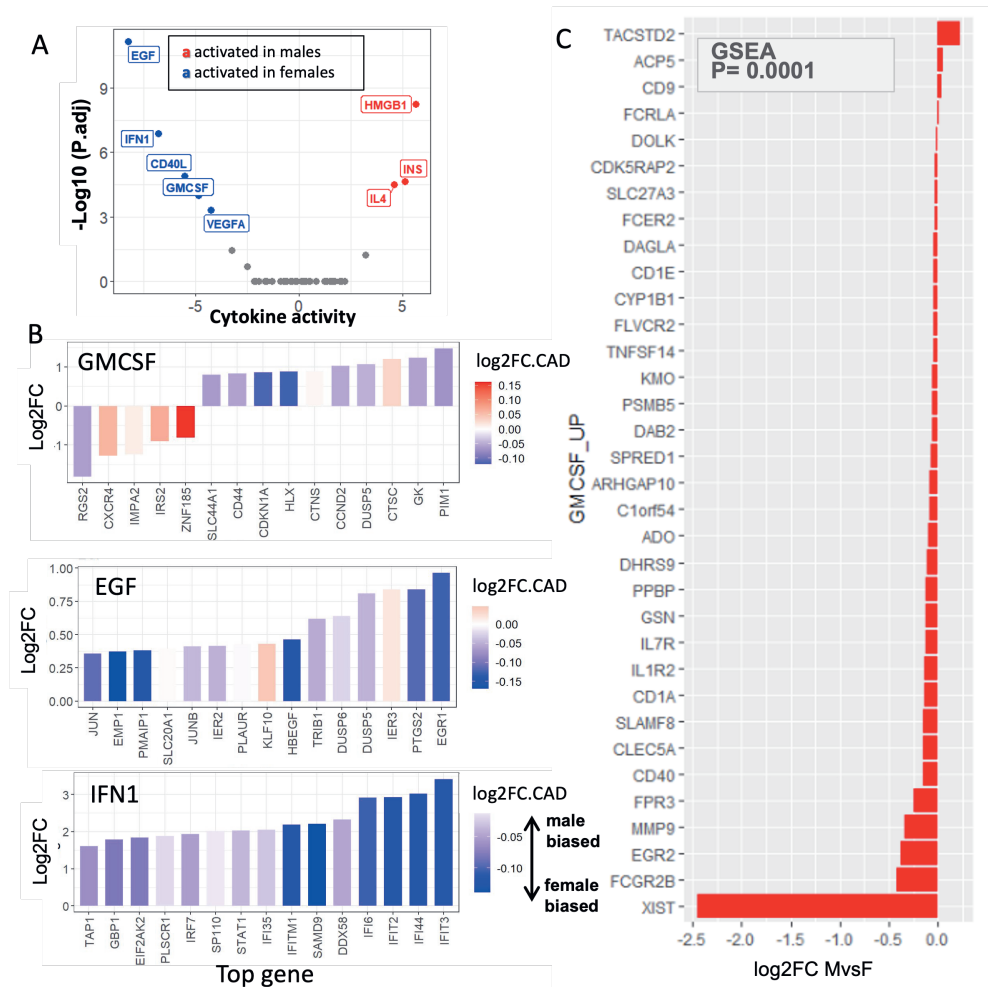
**Figure 2: Monocyte transcriptional signature of female CAD patients points to stronger migratory potential than that of men.** (A) Violin plots showing the percentages of the quantity of classical, non-classical, and intermediate monocytes from female and male patients. (B) Volcano plot showing sex-biased genes in CTMM cohort. significant female-biased ( $\log_2FC < -0.1$  and  $P\text{-adj} < 0.05$ ) and male-biased ( $\log_2FC > 0.1$  and  $P\text{-adj} < 0.05$ ) genes were color-coded with blue and red. (C) Dot plots visualized the significant levels of top enriched GO terms and pathways using GSEA. Male/female-biased categories are color-coded with red/blue. Significant levels are shown by using  $\log_{10}$ -transformed adjusted p-values. (D)&(E).  $\log_2FC$ s of genes in the enriched GO terms (chemokine activity and regulation of cell motility) have the top two highest significant levels.

## Identification of male- and female-biased cytokines

Since chemokine/interleukin signalling and cellular motility related pathways were significantly different between sexes, we inferred the cytokine, chemokine as well as growth factor pathway activities in male vs female CADs based on the intersection of genes from our CTMM cohort and Cytosig database [20] ( $n=4130$

genes). EGF, IFN1, GMCSF, VEGFA and CD40L signalling activities were significantly enriched in female patients, while INS, HMGB1 and IL4 pathways were enriched in male patients (**Fig 3A**). Indeed, for the female-enriched pathways, the majority of cytokine- and growth factor-induced genes are upregulated in female compared to male CAD patients (indicated by the blue-colored bars in **Fig 3B** and **Fig S1B**), while genes suppressed by the female biased ligands were enriched in males (red colors). Similarly, genes induced by IL4 and HMGB1 were mostly enriched in males (**Fig S1C**).

To determine whether the enhanced signalling activities were reflecting higher levels of the ligands in plasma, we interrogated an plasma proteomics dataset from the CAPIRE study cohort of CAD and healthy subjects [25]. In this plasma cohort (n =155 for CAD males; n=32 for CAD females, n=159 for healthy males, and n=182 for healthy females), we could only demonstrate a minor, yet significant, increase in VEGF levels in female plasma compared to males, but none of the other ligands showed even a trend towards significant differences between male and female CAD (**Fig S1D**), suggesting enhanced sensitivity for these ligands and/or intrinsic signalling differences within the monocytes. In line with the enrichment of chemokine activity term members in female monocytes (**Fig 2D**), we did find significantly increased plasma levels of CCL3 in female CAD patients (**Fig S1D**). Next, we performed GSEA of signature gene expression profiles of human CD14+ peripheral blood monocytes stimulated or not with GM-CSF (24h; 193 genes with  $\log_2FC > 0.5$ ,  $P_{adj} < 0.05$ ). The intersection of genes between CTMM GES and up-regulated genes induced by GM-CSF was visualized displayed in **Fig 3C**. GM-CSF induced signature indeed was significantly enriched in female compared to male CAD patient monocytes (GSEA,  $P = 0.0001$ ). Most GM-CSF induced genes, including MMP9, LMNA and CCL2 (all enriched GO term members, **Fig 2D-E**) were indeed seen to be significantly increased in female CAD patients. Interestingly, the gene transcript with the strongest upregulation in females (see figure 2B), i.e., the X-linked non-coding RNA XIST, also appeared to be induced by GM-CSF ( $\log_2FC=0.53$ ,  $P_{adj} = 3.54 * 10^{-10}$ ), next to Fc-gamma receptor 2B (FCGR2B), Early growth response gene-2 (EGR-2) and the motility transducer Formyl-peptide receptor 3 (FPR3). Moreover, we also found CD40, the receptor mediating CD40L signalling (activated in females) to be significantly induced by GM-CSF and enriched in female over male CAD patients.



**Figure 3: Inference of male- and female-biased cytokines using Cytosis.** (A) A volcano plot visualizing the Cytokine activities of CTMM CAD patients by comparing males versus females. By setting the cutoff of FDR adjusted p-value as 0.05, female- and male-biased cytokines were highlighted in blue and red. (B) Bar plots showing the top 15 genes (ranked by absolute  $\text{log}_2\text{FC}$ ) which are modulated by IL4, GMCSF and EGF from Cytosis platform. Bar colors indicate the  $\text{log}_2\text{FC}$  of these genes in CTMM cohort, comparing male vs female patients. (C) A bar graph showing the enrichment of GMCSF-induced upregulated genes from the RNA-Seq dataset over the sex-biased GES from the CTMM cohort. y-axis shows the overlapped genes between the GMCSF-induced upregulated genes from the RNA-Seq dataset and the GES from the CTMM cohort. X-axis indicates the  $\text{log}_2\text{FCs}$  (male versus female) of these genes in the CTMM cohort.

### Associations between sex-biased enriched cytokines and TFs

Our validation experiments point to cell-intrinsic differences in sex specific signalling pathways. Therefore, we investigated whether common regulatory mechanisms underly the sex-specific pathway activation. Hereto we first calculated TFs activities driving these sex-biased genes in CAD based on a TF-target interaction

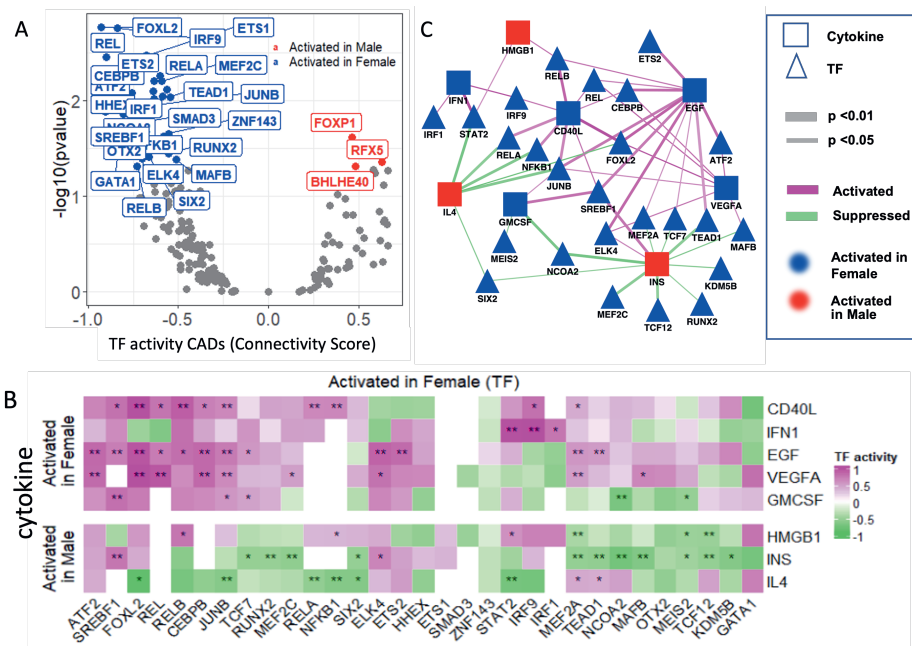
network from DoRothEA [21]. In line with our previous findings, we identified more activated TFs in women ( $n = 31$ ) than in men ( $n = 3$ ) (**Fig 4A**). We then inferred the activities of these sex-biased TFs to identify which TFs are linked to the 8 ligand-induced gene profiles from the Cytosig database (**Fig 4B**) and visualized the TFs that are significantly activated or suppressed by the ligands in a bipartite graph (**Fig 4C**). CD40L, EGF and VEGF were found to be linked to several of the TFs enriched in female CAD patients. As expected, we found STAT2, IRF1 and IRF9 to be significantly associated with IFN1, while NFKB1 and RELA were associated with CD40L. Both pathways were significantly enriched in females, while the male-enriched IL-4 cytokine activity was linked to suppressed STAT2, NFkB and RELA signalling. Most notably, JUNB was associated with 4 out of 5 female-enriched pathways, i.e., CD40L, EGF, VEGF and GM-CSF, and 1 of 2 male-enriched pathways (IL-4).

To identify the key target genes of TFs associated with these female-activated ligand pathways, we visualized the differential gene expressions between male/female CAD patients for selected TF's target genes and their dysregulation ( $\log_2FC$ ) induced by the respective cytokine/growth factors (**Fig S2**). Even though both the ligand- and TF pathways were highly significantly enriched in females, their target genes' expression showed only modest dysregulation in females compared to males. Interestingly, both GM-CSF and JUNB specifically induce the expression of LMNA, which is one of the genes with the strongest upregulation in female CAD patients compared to males (**Fig 2A** and **Fig S1A**), and a dominant gene involved in the female enriched GO term 'cell motility' (**Fig 2D**).

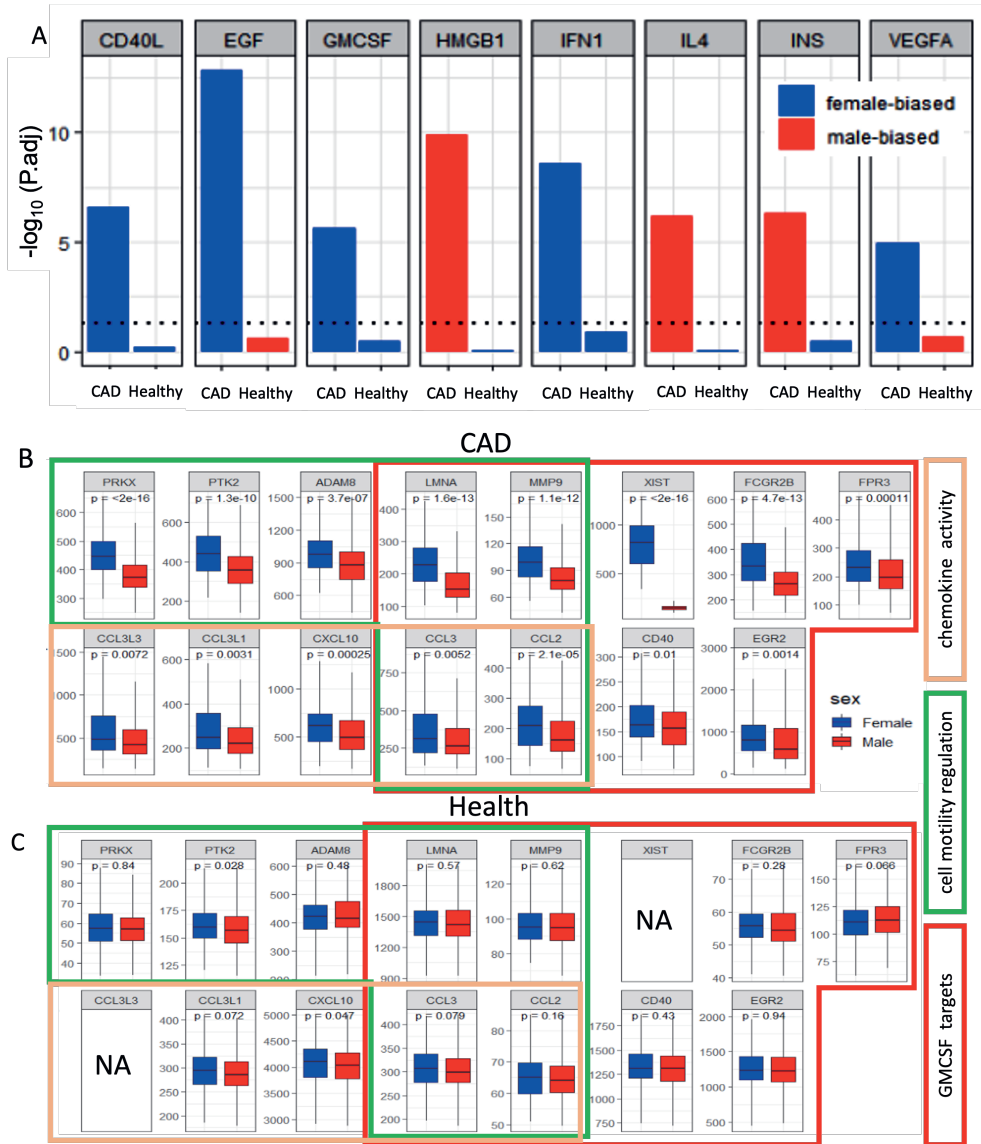
### Validation on monocytes from Healthy cohort

To validate if the sex differences we observed in the CTMM cohort are specific for CAD, we calculated pathway activities for the 5 ligands that are enriched in female CAD patients (i.e. CD40L, EGF, GMCSF, IFN1, VEGFA) and the 3 ligands that are enriched in male CAD patients (i.e. IL4, HMGB1, INS) in a monocyte cohort from 485 healthy males and females (GSE56034). **Fig 5A** shows that all these ligands did not show any significant activation or suppression scores in healthy men versus women, confirming the specificity of these sex differences for CAD. As the average age of the healthy cohort (i.e.  $32.9 \pm 10.9$  years for male;  $28.5 \pm 9.1$  years for females) is much lower than that of the CTMM cohort, we also analysed a subcohort of all subjects above 45 years old ( $n=43$  males, age=  $48.8 \pm 2.1$ ;  $n=29$  females; age=  $48.9 \pm 2.549$ ) (**Table S1**). As shown in **Fig S3**, also in this subcohort, GMCSF, CD40L, IFN1, IL-4 and HMGB1 signalling was not showing any differences in activation status in healthy individuals. Surprisingly, EGF, VEGF and Insulin activities were even found to be enriched in the opposite sexes in healthy subjects compared to CAD patients, pointing to sex and disease specific differences in ageing of these pathways.

We then visualized the expression of female-upregulated genes responsible for the female-biased GO term enrichment as well as the GM-CSF signature genes, i.e. XIST, FCGR2B, EGR2, MMP9, FPR3 and CD40. **Fig 5B** and **C** show that these genes were significantly upregulated in the monocytes of female compared to male CAD patients, but not in the healthy cohort. Only LMNA was already slightly elevated in females vs males, but only in the healthy >45 year sub-cohort suggestive of an age-dependent upregulation in women (**Fig 53B**). Of note, chemokine CCL2 expression was significantly higher in healthy middle-aged men vs women, whereas in CAD it was elevated in women. Altogether, these data strongly indicate the observed sex-biased signalling activities are specifically found in CAD and not in healthy subjects.



**Figure 4: Associations between sex-biased enriched cytokines and TFs. (A)** Volcano plot of inferred TF activities on CAD cohort by comparing male and female subjects. Nodes were color-coded based on the WTCS. A TF was defined as a female/male-activated TF if its absolute WTCS is higher than 0.4 and  $p$ -value  $< 0.05$  **(B)** Heatmap visualizing the activities of TFs that drive female-biased TFs and sex-biased cytokine in CAD. **(C)** Bipartite graph showing the connections between TFs that drive female-biased TFs and sex-biased cytokine in CAD. Only TFs that are significantly enriched in a cytokine-modulated gene profile are visualized ( $P < 0.05$  and absolute value of connectivity score  $> 0.5$ ).



**Figure 5: female-/male-biased genes and cytokines inferred from CTMM CAD cohort were validated on a healthy cohort (GSE56034). (A)** Activities of 8 CAD-related sex-biased cytokines on CAD and healthy cohort. Dashed lines indicate the significant threshold 0.05. **(B)** Expression levels in the CTMM cohort for 15 differentially expressed genes (male vs. female) that are involved in chemokine activity (yellow box), regulation of cell motility (green box) and GMCSF-driven target genes (red box). **(C)** Expression levels of the same 15 sex-biased genes as in panel B in a healthy cohort (GSE56034)

## Discussion

In this study, we examined sex-related differences in signalling pathways and their regulation in circulating monocytes of CAD patients. We found monocytes of



female CAD patients to have significantly activated CD40L, GM-CSF, EGF, VEGF and IFN1 signalling compared to men, whereas male CAD patients' monocytes displayed higher IL-4, Insulin and HMGB1 signalling activities compared to females. These sex effects were not observed in healthy subjects. The corresponding cytokine and growth factors activating these pathways did not differ between men and women in CAD, pointing to downstream differences in pathway regulation. Indeed, several of these pathways appeared to share a common transcriptional regulation, with JUNB central in 4 of 5 female and 1 of 2 male signalling pathways as most prominent sex specific TF.

Although several studies have shown increased plasma levels of the X-chromosome linked CD40L [31, 32], we could not confirm this in the plasma proteomics dataset of the CAPIRE study [25]. Except for significantly higher VEGF levels in female CAD patients, we did not observe any differences in plasma levels of female- or male-biased ligands. Type I IFN responses, mediated by the X-chromosome linked Toll-like receptor 7 (TLR7), are known to be enhanced in females [33], yet we did not find differences in IFN1 plasma levels between male and female CAD patients. Likewise, levels of HMGB1, reported to be higher in men than in women [34], were similar in men and women in the CAPIRE study. Therefore, we assume that sex biased differences in signalling activity are mainly pathway intrinsic and reflect sex differences in the monocyte's capacity to respond to a particular cytokine in CAD. These findings are in line with the observation of Gupta et al that female neutrophils displayed increased responsiveness to type I IFNs, independent of serum IFN1 levels [11].

We could confirm that the monocyte GM-CSF signalling signature was enriched in monocytes of female CAD patients. Moreover, several female-enriched GO term members, such as lamin A (LMNA), were induced by GM-CSF, which concurs with GMCSF's reported role in myeloid cell recruitment in inflammatory and pathological conditions [35].

While some studies reported a role for the nuclear membrane integrity stabiliser LMNA in NFkB-dependent inflammation [36], differentiation and migration of monocyte/macrophages [37, 38], sex differences in LMNA expression and the induction by GM-CSF are to our knowledge hitherto unknown.

Next to LMNA, GMCSF also induced CD40 which, at least partly, may explain the increased CD40L signalling activity in female monocytes. CD40-CD40L are co-stimulatory molecules well known for their role in inflammatory responses and atherosclerosis [39]. Interestingly, elevated expression of CD40L has been associated with cardiovascular diseases and risk for acute cardiovascular symptoms, also in women [40, 41]. While GM-CSF is mainly thought to act pro-inflammatory, it also can dampen excessive inflammation. For instance, GM-CSF

induced XIST, an X-chromosome linked long non-coding RNA and in fact, the RNA with the highest enrichment in females, was recently shown to accelerate wound healing in skin by promoting macrophage polarization towards an anti-inflammatory M2 phenotype [42]. Similar anti-inflammatory effects were reported for the GM-CSF induced female-enriched genes early-growth-response gene 2 (*egr-2*) [14, 43] and Fc-gamma receptor 2B (*FCGR2B*) [44]. Although it remains to be elucidated how monocyte pathway activities relate to sex differences in CAD development or manifestation, it is tempting to speculate that the abovementioned inflammation dampening and wound healing responses may account for the more stable, fibrocalcific plaque phenotype observed in females [45].

At premenopausal stages, most sex differences in CAD responses have been attributed to sex hormones. For instance, estrogen has been found to have anti-inflammatory effects on macrophages, by dampening NF $\kappa$ B signalling [18, 19, 46, 47], to reduce oxidative stress response in healthy murine peritoneal macrophages to bias macrophages toward the M2 phenotype [48] and to attenuate their migratory capacity [14]. However, sex specific disease mechanisms in post-menopausal women are less well studied, although CVD risk profiles of women increases profoundly even beyond that of men after menopause. Although we observed no differences in CVD risk factors between gender in our cohort, except for creatinine and HDL levels, it is well known that the relative risk for CVD conferred by such risk factors is greater in females than in males [4, 49]. This may also match with the seemingly paradoxical bias towards proinflammatory signalling in women with CAD in the CTMM cohort.

Our study has some limitations. While the CTMM contains a reasonable amount of 316 male and 134 female CAD patients, the cohort of the CAPIRE study is more limited and contains an imbalance in the number of male and female CAD patients in (**Table S2**, 32 female CADs and 155 male CADs). Therefore, Welch t test was used to compare the protein levels between males and females, to reduce type I error rates. Secondly, the average age of the healthy cohort subjects was considerably lower than that of the CAD cohorts (i.e. the CTMM patient cohort and the CAPIRE study cohort). While >50% of women in the latter cohorts were post-menopausal, this did not hold for the healthy cohort (average age men: 32.93 $\pm$ 10.96 years, women: 28.49 $\pm$ 9.11 years). However, also using this subcohort the main findings on sex-biased pathways remain valid giving confidence in our findings.

In conclusion, our study shows gender differences in signalling capacities of circulating monocytes in CAD patients. While male monocytes display enhanced signalling activities induced by HMGB1, insulin and IL-4, female monocytes are enriched in EGF, IFN1, CD40L, GM-CSF and VEGF signalling pathways. Although we did observe increases in plasma VEGF levels in female CAD patients compared to

males, these enhanced signalling responses mostly represent increased intrinsic signalling capacities. We showed increased GM-CSF signalling in female monocytes enhanced expression of various genes related to the female-enriched GO terms of chemokine activity and regulation of cell motility, as well as the expression of CD40, which may contribute to increased CD40L responsiveness.

## Funding and author contributions

This work was funded by the Center for Translational Molecular Medicine ([www.ctmm.nl](http://www.ctmm.nl)), project CIRCULATING CELLS (grant 01C-102) and the China Scholarship Council (CSC) (no. 201706990018 to C.L.).

C.L., H.d.B. performed analyses and experiments. M.M.P.C.D., J.K., R.C., J.P.-P., P.G., and E.A.L.B. supervised the project, and conceived and designed the analyses. C.L. and M.M.P.C.D. drafted the manuscript and all authors revised it critically for intellectual content. A.J.v.Z., J.W.J., A.K., J.K., G.P., and E.F. provided resources.

The data that support the findings of this study are openly available in the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo>), reference number GSE56034.

## Conflict of interest statement

Competing Interests: The authors have declared that no competing interests exist.

## Supplemental files

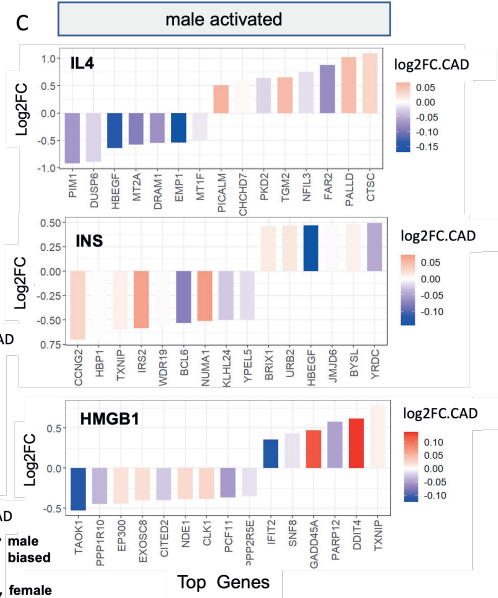
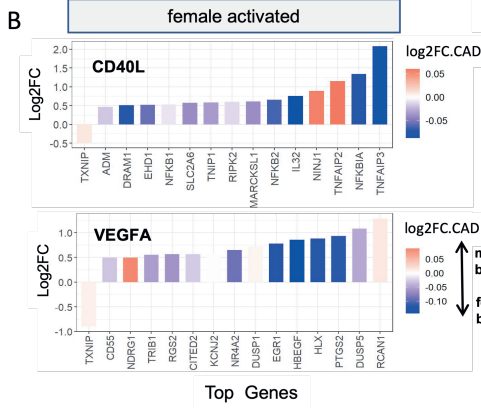
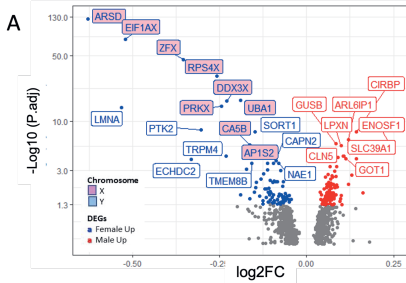
Table S1: Ages of healthy female and males in the monocyte cohort (GSE56034) expressed as mean±SD.

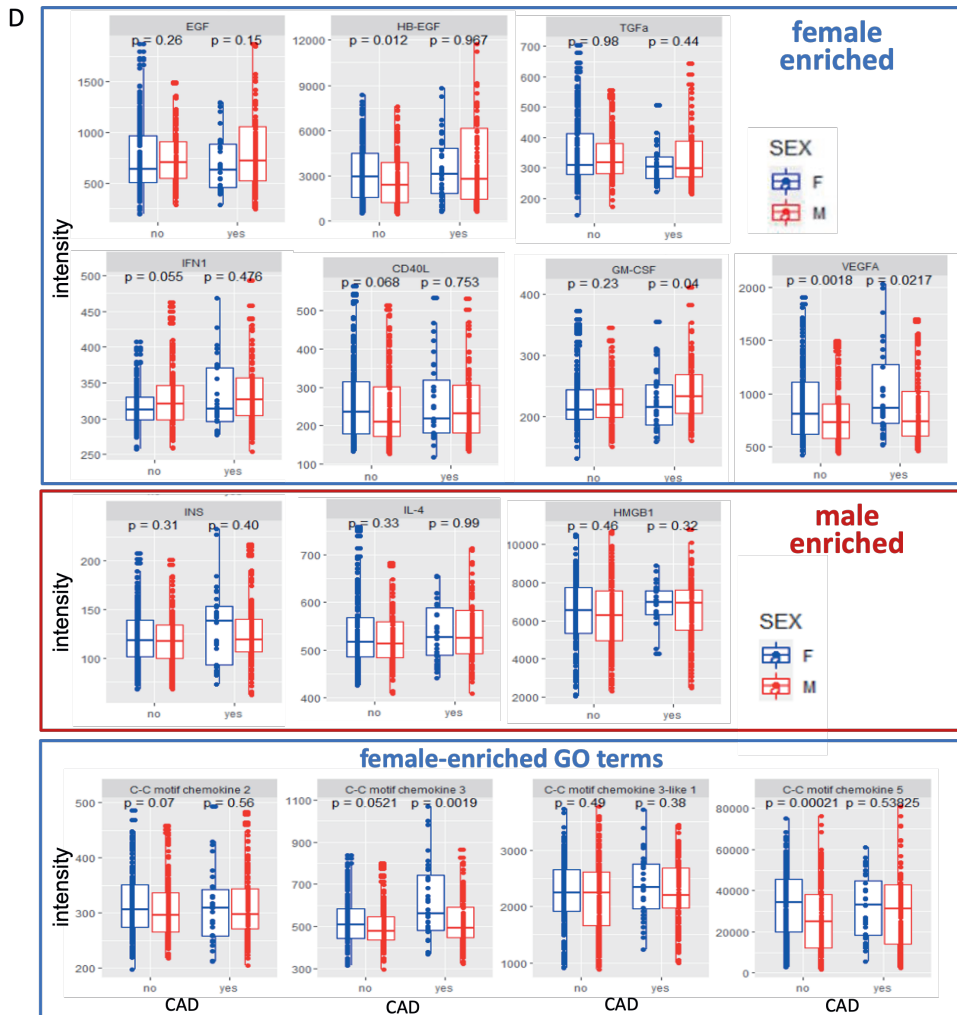
	Gender	Age (Mean±std)
Whole cohort	Male (n=213)	32.93±10.96
	Female (n=272)	28.49±9.11
Sub-cohort (Age > 45)	Male (n=43)	48.81±2.10
	female(n=29)	48.93±2.49

4

Table S 2: Demographics of females and males who participated in the cohort of proteomics study expressed as mean ± SD or proportion. P-values were calculated by Student's t-test

	CAD			Healthy		
	Female (n=32)	Male (n=155)	P-value	Female (n=182)	Male (n=159)	P-value
Diabetes (proportion)	0.16	0.19	0.62	0.09	0.08	0.68
Smoking (proportion)	0.56	0.26	<b>9.59E-04</b>	0.21	0.19	0.75
Age (years)	65.38± 6.36	62.22±7.76	<b>0.017</b>	59.86±8.19	56.37±8.16	<b>1.03E-04</b>
BMI (kg/m <sup>2</sup> )	28.01±5.79	27.53±3.94	0.65	25.75±4.47	26.31±3.20	0.18
SBP (mmHg)	132.03±16.37	133.87±15.97	0.56	125.66±14.29	127.25±14.01	0.30
DBP (mmHg)	80.47±9.32	81.59±8.41	0.53	77.46±7.72	78.94±8.09	0.09
HDL-C (mg/dL)	49.75±11.80	46.15±11.99	0.12	58.97±16.78	48.45±12.19	<b>1.03E-10</b>
LDL-C (mg/dL)	126.53±39.64	118.37±32.03	0.28	126.32±34.91	120.29±32.63	0.10
TG (mg/dL)	130.69±52.56	124.87±69.08	0.59	108.47±76.90	114.65±91.69	0.50
Cigarettes/day	8.469±11.16	4.78±10.55	0.092	2.75±6.17	3.13±7.54	0.62
Creatinine (mg/dL)	0.73±0.17	0.90±0.18	<b>6.74E-06</b>	0.69±0.12	0.93±0.133	<b>3.04E-47</b>

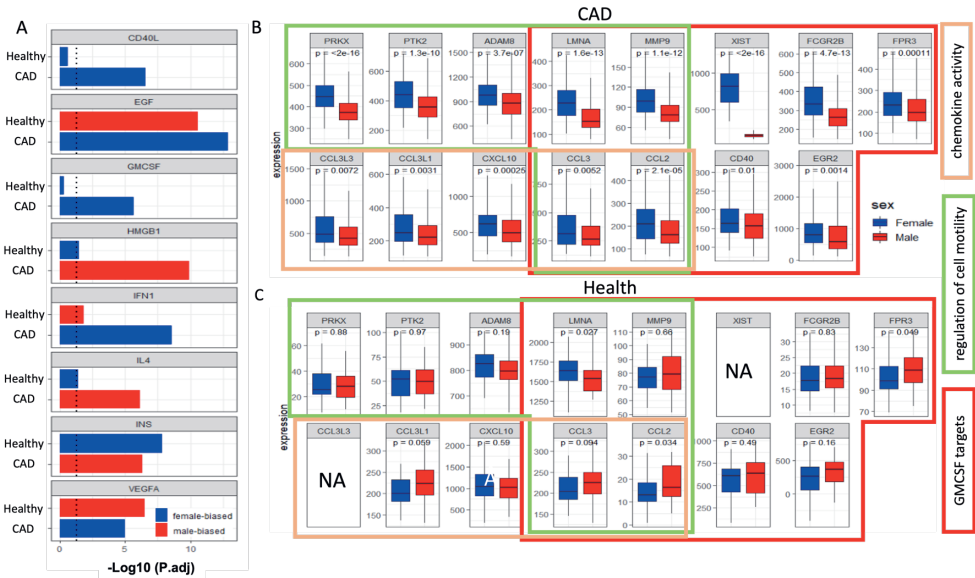




**Fig S1: Inference of cytokine activities based on sex-biased GES using Cytosig.** (A) Volcano plot showing sex-biased genes based on the overlapped genes of CTMM cohort and Cytosig database. (B) & (C) Bar plots showing the top 15 genes (ranked by absolute  $\log_2FC$ ) which are modulated by 2 female-activated cytokines (i.e. GMCSF, and VEGFA, shown in (B)); and male-activated cytokines (INS, HMGB1, EGF, shown in (C)) Bar colors indicate the  $\log_2$  fold change of these genes in CTMM cohort, comparing male vs female patients. (D) Intensities of 14 ligands from a plasma cohort of individuals with (yes) and without (no) CAD. These ligands show enhanced signaling activities calculated based on the gene profile of CTMM cohort.



**Fig S2:** Scatters showing the regulations of genes driven by a female-activated TF JUNB (A), CEBPB (B) and FOXL2 (C) in CAD and female-activated cytokines (i.e. VEGFA, CO40L, GMCSF and EGF). The X-axis indicates the log 2-fold changes of gene expressions comparing female and male patients in the CTMM cohort, and Y-axis is the log 2-fold changes of cytokine-induced gene expressions in the Cytosig database. The names of genes which absolute log2FC > 0.1 or pvalue < 0.05 in CTMM cohort (males verse females) were visualized.



**Fig S3: Female-/male-biased genes and cytokines inferred from CTMM CAD cohort were validated on a sub-healthy cohort (GSE56034, subjects above 45 years old). (A)** Activities of 8 CAD-related sex-biased cytokines on CAD and healthy subcohort of all subjects above 45 years old ( $n=43$  males, age=  $48.81 \pm 2.10$ ;  $n=29$  females; age=  $48.93 \pm 2.49$ ). Dashed lines indicate the significant threshold 0.05. **(B)** Expression levels in the CTMM cohort for 15 differentially expressed genes (male vs. female) that are involved in chemokine activity (yellow box), regulation of cell motility (green box) and GMCSF-driven target genes (red box). **(C)** Expression levels of the same 15 sex-biased genes as in panel B in a subcohort of all subjects over 45 years of age (43 males, age=  $48.81 \pm 2.10$ ; 29 females, age=  $48.93 \pm 2.49$ )



## References

1. Man JJ, Beckman JA, Jaffe IZ. Sex as a Biological Variable in Atherosclerosis. *Circ Res.* 2020;126:1297–319.
2. Park SM, Merz CN. Women and Ischemic Heart Disease: Recognition, Diagnosis and Management. *Korean Circ J.* 2016;46:433–42.
3. Vakhtangadze T, Singh Tak R, Singh U, Baig MS, Bezsonov E. Gender Differences in Atherosclerotic Vascular Disease: From Lipids to Clinical Outcomes. *Front Cardiovasc Med.* 2021;8.
4. Andreotti F, Marchese N. Women and coronary disease. *Heart.* 2008;94:108–16.
5. Anand SS, Islam S, Rosengren A, Franzosi MG, Steyn K, Yusufali AH, et al. Risk factors for myocardial infarction in women and men: insights from the INTERHEART study. *Eur Heart J.* 2008;29:932–40.
6. Libby P. Inflammation in atherosclerosis. *Arter Thromb Vasc Biol.* 2012;32:2045–51.
7. Lofblad L, Hov GG, Asberg A, Videm V. Inflammatory markers and risk of cardiovascular mortality in relation to diabetes status in the HUNT study. *Sci Rep.* 2021;11:15644.
8. Ridker PM, Rifai N, Clearfield M, Downs JR, Weis SE, Miles JS, et al. Measurement of C-reactive protein for the targeting of statin therapy in the primary prevention of acute coronary events. *N Engl J Med.* 2001;344:1959–65.
9. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med.* 2017;377:1119–31.
10. Barrett TJ. Macrophages in Atherosclerosis Regression. *Arterioscler Thromb Vasc Biol.* 2020;40:20–33.
11. Gupta RM, Lee-Kim VS, Libby P. The March of Monocytes in Atherosclerosis: One Cell at a Time. *Circ Res.* 2020;126:1324–6.
12. Stout RD, Suttles J. Functional plasticity of macrophages: reversible adaptation to changing microenvironments. *J Leukoc Biol.* 2004;76:509–13.
13. Beenakker KGM, Westendorp RGJ, de Craen AJM, Chen S, Raz Y, Ballieux B, et al. Men Have a Stronger Monocyte-Derived Cytokine Production Response upon Stimulation with the Gram-Negative Stimulus Lipopolysaccharide than Women: A Pooled Analysis Including 15 Study Populations. *J Innate Immun.* 2020;12:142–53.

14. Chen KE, Lainez NM, Coss D. Sex Differences in Macrophage Responses to Obesity-Mediated Changes Determine Migratory and Inflammatory Traits. *J Immunol.* 2021;206:141–53.
15. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol.* 2016;16:626–38.
16. Folsom AR, Aleksic N, Sanhueza A, Boerwinkle E. Risk factor correlates of platelet and leukocyte markers assessed by flow cytometry in a population-based sample. *Atherosclerosis.* 2009;205:272–8.
17. Heimbeck I, Hofer TP, Eder C, Wright AK, Frankenberger M, Marei A, et al. Standardized single-platform assay for human monocyte subpopulations: Lower CD14+CD16++ monocytes in females. *Cytom A.* 2010;77:823–30.
18. Chadwick CC, Chippari S, Matelan E, Borges-Marcucci L, Eckert AM, Keith JC Jr, et al. Identification of pathway-selective estrogen receptor ligands that inhibit NF-kappaB transcriptional activity. *Proc Natl Acad Sci U A.* 2005;102:2543–8.
19. Demyanets S, Pfaffenberger S, Kaun C, Rega G, Speidl WS, Kastl SP, et al. The estrogen metabolite 17beta-dihydroequilenin counteracts interleukin-1alpha induced expression of inflammatory mediators in human endothelial cells in vitro via NF-kappaB pathway. *Thromb Haemost.* 2006;95:107–16.
20. Jiang P, Zhang Y, Ru B, Yang Y, Vu T, Paul R, et al. Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat Methods.* 2021;18:1181–91.
21. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* 2019;29:1363–75.
22. Hoefler IE, Sels JW, Jukema JW, Bergheanu S, Biessen E, McClellan E, et al. Circulating cells as predictors of secondary manifestations of cardiovascular disease: design of the CIRCULATING CELLS study. *Clin Res Cardiol.* 2013;102:847–56.
23. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinforma Oxf Engl.* 2008;24:1547–8.
24. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinforma Oxf Engl.* 2012;28:882–3.

25. Ferrannini E, Manca ML, Ferrannini G, Andreotti F, Andreini D, Latini R, et al. Differential Proteomics of Cardiovascular Risk and Coronary Artery Disease in Humans. *Front Cardiovasc Med.* 2022;8.
26. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
27. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284–7.
28. Cule E, De Iorio M. Ridge regression in prediction problems: automatic choice of the ridge parameter. *Genet Epidemiol.* 2013;37:704–14.
29. Cule E, Vineis P, De Iorio M. Significance testing in ridge regression for genetic data. *BMC Bioinformatics.* 2011;12:372.
30. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell.* 2017;171:1437-1452.e17.
31. Lau ES, Paniagua SM, Guseh JS, Bhambhani V, Zanni MV, Courchesne P, et al. Sex Differences in Circulating Biomarkers of Cardiovascular Disease. *J Am Coll Cardiol.* 2019;74:1543–53.
32. Varo N, Vicent D, Libby P, Nuzzo R, Calle-Pascual AL, Bernal MR, et al. Elevated plasma levels of the atherogenic mediator soluble CD40 ligand in diabetic patients: a novel target of thiazolidinediones. *Circulation.* 2003;107:2664–9.
33. Webb K, Peckham H, Radziszewska A, Menon M, Oliveri P, Simpson F, et al. Sex and Pubertal Differences in the Type 1 Interferon Pathway Associate With Both X Chromosome Number and Serum Sex Hormone Concentration. *Front Immunol.* 2018;9:3167.
34. Chen L, Zhu H, Su S, Harshfield G, Sullivan J, Webb C, et al. High-Mobility Group Box-1 Is Associated With Obesity, Inflammation, and Subclinical Cardiovascular Risk Among Young Adults: A Longitudinal Cohort Study. *Arterioscler Thromb Vasc Biol.* 2020;40:2776–84.
35. Ushach I, Zlotnik A. Biological role of granulocyte macrophage colony-stimulating factor (GM-CSF) and macrophage colony-stimulating factor (M-CSF) on cells of the myeloid lineage. *J Leukoc Biol.* 2016;100:481–9.

36. Kim Y, Bayona PW, Kim M, Chang J, Hong S, Park Y, et al. Macrophage Lamin A/C Regulates Inflammation and the Development of Obesity-Induced Insulin Resistance. *Front Immunol.* 2018;9:696.
37. Marelli-Berg FM, Nadkarni S. Displacing, squeezing, and ramming: The role of nuclear lamins in leukocyte migration. *J Leukoc Biol.* 2018;104:235–6.
38. Saez A, Herrero-Fernandez B, Gomez-Bris R, Somovilla-Crespo B, Rius C, Gonzalez-Granado JM. Lamin A/C and the Immune System: One Intermediate Filament, Many Faces. *Int J Mol Sci.* 2020;21:E6109.
39. Lutgens E, Lievens D, Beckers L, Donners M, Daemen M. CD40 and its ligand in atherosclerosis. *Trends Cardiovasc Med.* 2007;17:118–23.
40. Shami A, Edsfeldt A, Bengtsson E, Nilsson J, Shore AC, Natali A, et al. Soluble CD40 Levels in Plasma Are Associated with Cardiovascular Disease and in Carotid Plaques with a Vulnerable Phenotype. *J Stroke.* 2021;23:367–76.
41. Schönbeck U, Varo N, Libby P, Buring J, Ridker PM. Soluble CD40L and cardiovascular risk in women. *Circulation.* 2001;104:2266–8.
42. Pi L, Fang B, Meng X, Qian L. LncRNA XIST accelerates burn wound healing by promoting M2 macrophage polarization through targeting IL-33 via miR-19b. *Cell Death Discov.* 2022;8:1–10.
43. Veremeyko T, Yung AWY, Anthony DC, Strelakova T, Ponomarev ED. Early Growth Response Gene-2 Is Essential for M1 and M2 Macrophage Activation and Plasticity by Modulation of the Transcription Factor CEBP $\beta$ . *Front Immunol.* 2018;9.
44. Santegoets KCM, Wenink MH, Berg WB van den, Radstake TRDJ. Fc Gamma Receptor IIb on GM-CSF Macrophages Controls Immune Complex Mediated Inhibition of Inflammatory Signals. *PLOS ONE.* 2014;9:e110966.
45. Mathur P, Ostadal B, Romeo F, Mehta JL. Gender-Related Differences in Atherosclerosis. *Cardiovasc Drugs Ther.* 2015;29:319–27.
46. Deshpande R, Khalili H, Pergolizzi RG, Michael SD, Chang MD. Estradiol down-regulates LPS-induced cytokine production and NF $\kappa$ B activation in murine macrophages. *Am J Reprod Immunol N Y N* 1989. 1997;38:46–54.
47. Huang H, He J, Yuan Y, Aoyagi E, Takenaka H, Itagaki T, et al. Opposing effects of estradiol and progesterone on the oxidative stress-induced production of chemokine and proinflammatory cytokines in murine peritoneal macrophages. *J Med Investig JMI.* 2008;55:133–41.

48. Villa A, Rizzi N, Vegeto E, Ciana P, Maggi A. Estrogen accelerates the resolution of inflammation in macrophagic cells. *Sci Rep.* 2015;5:15224.
49. Man JJ, Beckman JA, Jaffe IZ. Sex as a Biological Variable in Atherosclerosis. *Circ Res.* 2020;126:1297–319.



## Chapter 5

Identification of a gene network driving the hypertension associated dampened monocyte LPS response in coronary artery patients

EMBARGOED

**Chang Lu**<sup>†</sup>, Marjo M.P.C. Donnerst<sup>†</sup>, Julius B. J. de Baaij, Han Jin, Jeroen Otten, Marco Manca, Barend Mees, Mat J.P. Daemen, Judith C. Sluimer, Rachel Cavill, Joël M.H. Karel, Pieter Goossens, Erik A.L. Biessen

*In preparation*

<sup>†</sup> Authors contributed equally

# **Chapter 6**

General discussion





Cardiometabolic diseases (CMDs), including cardiovascular diseases (CVDs) and non-alcoholic fatty liver disease (NAFLD), are characterized by chronic inflammation with a key role for monocytes and macrophages [1–4]. In particular, macrophages are highly plastic cells, which adjust to their environment to display a high level of heterogeneity [4, 5]. Recent advances in omics technology have allowed us to characterize genetic, protein, and metabolic makeup of not only bulk tissue, but even of individual cells in complex tissue. Rapid technological developments in imaging enable us to capture fluorescence signals at high resolution while preserving morpho-spatial features [6, 7]. This thesis deployed a series of computational strategies based on statistics, machine learning, and image processing to analyse high-throughput or imaging data, aiming (1) to dissect the heterogeneity of the macrophage population and their molecular environment in CVD-affected arteries (more specifically, in atherosclerotic plaque); (2) to analyse the heterogeneity of liver compartments and metabolic changes during the development of NAFLD; (3) to infer the transcriptional and functional profiles of monocytes in the progression of CVD.

The main results of this thesis are as follows:

1. We developed a novel approach to distinguish cell types and characterize their microenvironments by integrating multiple imaging data layers from different platforms (i.e., multispectral, H&E-stained and MALDI-mass spectroscopy imaging (MSI)), which has been applied to dissect macrophage heterogeneity, as well as their local niche in atherosclerotic plaque (**chapter 2**).
2. We established a computational pipeline to analyse the spatiotemporal heterogeneity of liver tissues from mice with NAFLD based on metabolite levels from MALDI-MSI data. This study not only mapped the different compartments of liver tissue solely based on their metabolic profile, but also enabled to determine the dynamics of metabolite profiles in each compartment during disease progression (**chapter 3**).
3. We identified differential gene expression in monocytes from male and female CAD patients, inferred the transcription factors (TFs) driving this difference between sexes and the cytokines associated with it. The findings were validated on independent cohorts (**chapter 4**).
4. We constructed a gene co-regulation and corresponding regulatory network by comparing the gene expression profiles of monocytes from a cohort of CAD patients before and after LPS stimulation and identified a subnetwork with strong correlation with blood pressure levels, that may link inflammation to hypertension (**chapter 5**).

The results of my studies not only demonstrate the power of advanced machine

learning and statistical models in analysing high-throughput multi-omics data to unravel disease mechanisms, but also exemplify the effectiveness and broad applicability of the pipelines we developed for cell/tissue heterogeneity analysis. This chapter will evaluate the computational methods employed in the thesis, describing potential application scenarios in the analysis of cardiometabolic diseases and advantages and limitations compared to reported approaches, and provide directions for future research.

## Machine learning versus statistical models in cardiometabolic diseases

Although statistical models have been and continue to be widely used for biomedical data analysis, machine learning has gained momentum in the past few years. It has achieved increasingly impressive performances for analysis of high-dimensional datasets, which are ever expanding in size and number [8]. In this thesis, I applied different computational strategies in various contexts to explore the mechanisms of CMDs.

Although statistical models and machine learning may employ the same approaches, such as linear regression, they serve diverse purposes. (Supervised) machine learning is mainly focussed on the model's prediction performance; therefore, it requires a test set to evaluate the performance of a trained model. A statistical model does not require such a test dataset, but analyses the model for example in terms of confidence intervals. The main aim of the statistical model is not the performance on future data, but to infer relationship between variables [9, 10]. This distinction is significant and has major implications for their application on biomedical data.

In the biomedical field, the number of samples available in an experiment (e.g., patients, mice) is often limited, and statistical methods can estimate information about the populations based on these limited sample sizes. For example, biologists are generally interested in which genes, proteins, or metabolites are significantly different between separate patient groups, tissues or cell populations, or differ with treatment. Statistical methods such as ANOVA and linear-model based moderated t-statistics [11] are commonly used approaches to assess differential expression in the context of multifactor-designed experiments. In this thesis, I applied limma [11] on the CTMM [12] cohort, to compare the gene expression differences between male and female CVD patients (**chapter 4**), revealing female-biased genes (e.g., LMNA) associated with cell migration. The selected differentially expressed genes could be used for predictions, but the performance is often barely satisfactory.

As the size and dimensions of the biomedical data has increased dramatically in recent years, a number of problems regarding their analysis have arisen. For example, as long as the data size is large enough, the results of hypothesis testing will always be significant [13]. Machine learning, which can handle large-scale and high-dimensional datasets, has been widely used for CMD prediction, where they have shown extraordinary predictive performances [14, 15]. In **chapter 3**, we have studied metabolic changes over time during the progression of NAFLD and in different liver compartments; hereto I trained an spatial shrunken centroids (SSC) [16] classification model based on pixels on each MALDI-MSI image (with more than

10,000 pixel points (samples) on each image, and 504 features (i.e. m/z peaks) per pixel point) from healthy mice and then predicted the pixels of the liver tissues from mice that were fed different regimens of a high-fat diet, ranging from 16 hours to 3 weeks. The validity of the compartment classification (i.e., parenchyma, sinusoid, and vessels) was confirmed at histological level. On the other hand, a small sample size, especially when the feature number by far exceeds sample size, will result in the notorious overfitting issue, where the model is only shaped to the training set only and performs much worse on the test data set than it did on the training data set. In addition, high prediction performance of machine learning unfortunately often is at the expense of a lower model interpretability, which leads to its still limited application in the study of CMD pathology.

Although feature selection in machine learning can also be used for identifying gene sets that are characteristic of a tissue or condition, its primary goal is to improve the model performance by eliminating redundant and irrelevant features [17]. Thus, co-linear characteristics will often be deleted since they do not contribute much to model improvement. However, it is well known that genes (but also proteins and metabolites) do not function independently, but rather in collaboration with others. Therefore, analysis of groups of genes with comparable expression patterns is more likely to reveal the included biological significance, than selecting a set of genes that will give the model the highest prediction accuracy. This is why gene clustering strategies are so widely applied in biomedical data analysis. An example can be found in **chapter 5**, where we employed Weighted Gene Co-expression Network Analysis (WGCNA) [18] to group genes that are co-regulated in response to a lipopolysaccharide (LPS) challenge into functional gene modules. This strategy led to the identification of a gene module with a strong correlation to the diastolic blood pressure. In **chapter 3**, a temporal clustering method GPDP [19] was applied on the MALDI-MSI data of livers from mice with various levels of NAFLD, for pinpointing stage-dependent changes in metabolic pathways (as represented by the clusters). Such identification would not have been possible without prior grouping of co-expressed metabolites (for instance by regular univariate analysis).

The recent development of single-cell technology has dramatically increased the data size, but fortunately, this was accompanied by a huge increase in the number of units (“samples”) per analysis. Cell type identification based on such data requires clustering as well, which can either be done guided by signature gene sets for single cell RNA-Seq (scRNA-Seq) analysis, or based on marker expression for multiplex, multispectral or mass spectrometry imaging analysis. A general pipeline for cell type identification first clusters individual cells according to their gene or marker expression in an unbiased manner, and then assigns clusters to a particular cell type based on a known biomarker signature. In **chapter 2**, we discovered multiple macrophage phenotypes in murine atherosclerosis plaques by grouping the cell segments based on the intensity of 12 phenotypic markers (detected by

multiplex multispectral immunofluorescent imaging) using cosine-based k-medoid clustering. The resulting clusters were then assigned to previously identified cell types and subsets, based on similarities in marker profile, albeit with the limitation that the cell signature and image markers involved different feature domains (genes versus proteins).

Compared with the above approach, cell clustering based on scRNA-Seq data will be more complicated, because of the huge feature (gene) and sample size. As a result, one usually needs to perform a gene filtering or dimensionality reduction step prior to clustering. Since the multispectral imaging data in **chapter 2** only contains 12 biomarkers (features), dimensionality reduction is not obligatory and may result in loss of information. On the other hand, spillover compensation is recommended before clustering these imaging data, because signals from adjacent cells could contaminate each other. Therefore, in **chapter 2**, we have adapted the spillover compensation method from Goltsev et al.[20] to reduce the risk of overcompensation and applied it before cell clustering.

How to determine the optimal number of cell clusters or how to evaluate the best clustering solution has been an open question. Some groups propagate the use of inter-class or intra-class distances (or similarities) (e.g., Silhouette Coefficient and Davies-Bouldin Index) as an effective measure to select the most appropriate clustering solution[21]. Alternatively, other methods have been described, such as Dbscan [22], that circumvent specifying the number of clusters by designing and adjusting the parameters 'minPts' (the minimum number of points clustered together for a region to be considered dense) and 'eps' (a distance measure that will be used to locate the points in the neighbourhood of any point). In the **chapter 2** of this thesis, I have compared the Silhouette coefficients of five clustering methods based on different distance measures for the same number of clusters ( $k=70$ ), and the best results were obtained with k-medoid clustering based on cosine distance. Additionally, to bypass the cluster number uncertainty, a Dirichlet process Gaussian process mixture model (DPGP)[19] has been utilized to identify metabolite groups with consistent trajectories over time without pre-specified cluster counts.

It should be noted however that without guidance from expert knowledge, gene or cell clusters delineated by unsupervised clustering methods sometimes do not reflect biologically plausible gene modules or cell phenotypes. To address this issue and the fact that clustering often strives for more or less balanced cluster sizes (ignoring the presence of rare but true populations), in **chapter 2** and **3**, I adopted an interactive approach to cluster cells or genes, involving a deliberate overclustering step after which clusters were merged combining information on location and morphology (inspection by a biologist) with simple hierarchical clustering to obtain biologically more plausible outcome.

In conclusion, statistical models, or machine learning both have their merits, and should be applied with care, depending on the research purpose and the data type. In general, one could say that, when the sample size is limited and most interest lies in the association between variables (e.g., disease risk factors), statistical models are typically more appropriate, whereas supervised machine learning is regularly a better option when the goal is to predict the patient's risk of a disease and there is a sufficient number of training samples to build a reliable predictive model. Unsupervised clustering is able to systematically analyse co-expressed genes or quickly group cell types without the knowledge of the ground truth cell type. Therefore, statistical methods are essential for assessing CVD risk and developing prevention strategies, while machine learning based predictive models for patient's medical data are useful to guide CVD prevention and support the clinical decision-making process [23, 24].

## Enrichment analysis: a knowledge-based approach for interpreting omics profiles

With the exponential growth of available DNA, RNA, protein, and metabolite datasets, we are no longer limited to solely relying on manual searches from the literature to evaluate and interpret the results generated by omics analysis. Pathway enrichment analysis has become a common approach for solving this problem, allowing researchers to obtain mechanistic insight into their gene lists [25]. The main procedures for pathway enrichment analysis are similar, whether for genes, proteins, or metabolites. In brief, the omics data collected during the experiment is utilized to establish the feature list of interest, and statistical methods are then employed to determine whether the gene list is enriched in members of the pathways from a reference database (e.g. KEGG [26] and wikipathway [27]). The results of enrichment analysis are then visualized for further in-depth inspection, evaluation, and experimental validation. In this thesis, I have employed this process several times to infer the pathways involved in sex-biased gene expression signature (GES) from CVD patients (**chapter 4**), a gene module of monocytes that was suppressed by LPS and positive correlated with blood pressure levels (**chapter 5**), as well as the NAFLD-stage-related metabolite clusters (**chapter 3**). Furthermore, this procedure has also been utilized for additional tasks, such as TF activity inference [28] and drug repurposing [29], as long as we have the 3 elements of inference: a list of genes of interest, a (public online) database to serve as reference, and an appropriate inference method. The performance of the analysis results depends on the reliability of the reference database and the enrichment method we used.

Since enrichment analysis aims to quantify the similarity (or overlap) between the list of interest and the list in the reference database, the choice and design of the enrichment method depends on the data types of the two lists. For over-representation analysis (ORA), both the gene list in the public database and the gene list of interest (generally a significantly up- or down-regulated gene set) are unordered gene sets involved in a pathway, so Fisher's exact test is usually used to calculate the enrichment of unordered genes of interest in a certain pathway. If the generated dataset includes information on differential gene expression (such as t-values or log<sub>2</sub> fold changes of differential expression), ranking-based tests (such as the weighted Kolmogorov-Smirnov-like test implemented in GSEA [30]) will perform better because they allow detection of small but consistent changes in gene expression. However, for some specific tasks (e.g., TF activity inference and drug repurposing), the gene list in public databases are in order or/and directed, so the weight and direction of the two lists should be taken into account during the statistical test. In **chapter 4**, we inferred TF activity based on the gender-biased GES in monocytes from CVD patients and the TF-target database from Dorothea [28],



but now utilizing the Weighted Connectivity Score (WTCS) [29], an inference method designed for drug repurposing, instead of the default inference method (VIPER (Virtual Inference of Protein-activity by Enriched Regulon analysis) [31]). The latter approach, originally designed for directed continuous vectors, is less suitable for reference gene lists containing only the directions but no weights (such as Dorothea) and for an unbalanced gene signature as was the case for the CTMM study, where we had more female-biased than male-biased genes.

VIPER virtually infers protein activity by calculating the enrichment of each regulon of the gene regulatory network generated by ARACNe ([32]) in the GES list using an analytic rank-based enrichment analysis (aREA). To be specific, for a GES of length  $n$ , the enrichment score of a TF  $j$  with  $k$  target genes is defined as follows:

$$ES_j = \begin{cases} ES_{2j} + aES_{1j} & \text{if } ES_{1j} > 0 \\ ES_{2j} & \text{otherwise} \end{cases}$$

$$ES_{2j} = \sum_{i=1}^n m_{ij} w_{ij} t_i \quad \text{and} \quad ES_{1j} = \sum_{i=1}^n (1 - |m_{ij}|) w_{ij} q_i$$

where  $m_{ij}$  and  $w_{ij}$  describe the correlation and mutual information (MI) between TF  $j$  and its target gene  $i$  according to the ARACNe network [32], respectively, if  $i$  is in the set of target genes of  $j$ , or equals zero if  $i$  is not a target gene. As such, the vectors  $m_{.j}$  and  $w_{.j}$  are extended from size  $k \times 1$  to  $n \times 1$ .  $t$  is a vector of length  $n$ , indicating the quantile-transformed rank of the GES (in ascending order), and  $q$  is a vector of length  $n$ , meaning the quantile-transformed symmetrized rank (middle rank becoming the lowest rank) of the GES. In brief, VIPER calculates the extent to which the orientation and intensity of TF-regulated genes are consistent with their orientation and intensity in the GES list (for a detailed description of the aREA and VIPER algorithm see [31]).

However, for Dorothea, the relationship between TF and its target genes only contains the directions without weights (i.e., the items in  $m_{ij} = -1$  or  $1$ , where  $1$  represents forward regulation and  $-1$  represents reverse regulation), and  $w_{ij}=1$  because there is no information on MI in Dorothea database.  $ES_{1j}$  of the above equation is therefore equal to 0. Thus, the  $ES_j = \sum_{i=1}^n m_{ij} t_i$ , which is the dot product of the regulatory directions of the  $j$ 's target genes and their quantile-transformed rank positions within the GES.

In **chapter 4** we describe a sex-biased signature of genes with significantly different expression (log 2 fold change) in males versus female CVD patients, and vice versa. This signature is biased as we found more female- than male-biased genes. Hence, the middle rank of GES is not 0. Imagine a TF with all female-biased target genes but only minor differences in gene expression between males and females, then

this TF will be inferred as a male-activated TF if we would use rank-based aREA. In contrast, connectivity scores are based on a weighted Kolmogorov-Smirnov enrichment statistic [30], which directly weighs the ranked genes using GES and can therefore also handle unbalanced numbers of up- and down-regulated genes. Furthermore, WTCS is able to determine that a TF should not be a regulator of a GES if its positively and negatively regulated target genes are enriched on the same side of the GES. Moreover, in **chapter 4**, WTCS was used for generating a TF-cytokine network by assessing TF activity from 51 cytokine-induced GESs of the CytoSig database of cytokine signalling pathways, to probe the association of TFs and cytokines.

In addition to the statistical inference method, also the quality of the public databases has a significant impact on the inference results. For instance, a limitation in the analysis of sex-specific pathways in monocytes from CVD (**chapter 4**) is the incomplete coverage of genes used for cytokine and TF activity inference. The CytoSig database [33] is still in the stage of development and refinement; therefore the overlap between genes in CytoSig and sex-biased GES from the CTMM cohort was limited. Likewise, although Dorothea integrated human TF-target interactions from a variety of sources, including literature-selected sources, a majority of these interactions were inferred by reverse engineering, and lacks experimental or reported evidence [28]. Exclusion of these non-validated interactions will obviously compromise the number of genes available for inference and such sparse or insufficiently large data are more likely to lead to biased inference results. In the future, we expect a steady stream of new datasets to be integrated into such reference databases, through the joint efforts of biologists and data scientists.

## Spatial multi-omics integration across platforms

Atherosclerotic tissue forms a complex microenvironment with a large but heterogeneous population of immune cells, amongst others [34, 35]. It is important to comprehend the differences in these cells within the diseased tissue and how they affect plaque initiation, growth and stability for unravelling and understanding the complicated immunological environment and designing targeted therapeutics. The advent of single-cell omics has allowed the dissection of cell types and phenotypes in tissue at much higher precision than was possible by deconvolution of bulk transcriptomics. However, scRNA-Seq [36] but also cytometry, be it fluorescence-based or by time of flight [37], require prior isolation of single cells from the tissue sample, resulting in loss of information on the cells' location. Some imaging techniques such as immunohistochemistry (IHC) and immunofluorescence microscopy (IF), while capable of measuring the intensity of cell markers in the tissue and at high resolution, suffer from allowing the combination of just a limited number of markers and thus offer a limited phenotypic resolution. Only a few spatial omics techniques can measure and analyse multi-omics profiles of single cells in tissue. To address these problems, in **chapter 2**, we presented a flexible and scalable computational pipeline capable of integrating multispectral IF imaging with mass spectrometry imaging information while preserving the cellular spatial context, combining the strengths of both approaches. With this pipeline we identified a high number of macrophage phenotypes and at the same time dissected the specific cellular and molecular microenvironment of these subsets.

Compared to the conventional spatial or single-cell omics techniques mentioned above, this pipeline offers several advantages: First of all, in addition to the ability to interactively define cell segments and phenotypes, our pipeline is able to project the locations of each phenotype on histological images by automatically aligning multispectral images with their corresponding H&E-stained images. Secondly, we developed a method to accurately quantify and compare the activity of disease-relevant cell processes per phenotype, by including a characteristic marker of that process in the antibody panel (such as KI67, a proliferation marker). Thirdly, we improved the recently reported compensation method (i-NICHE) for cross-cellular spillover of marker signals between neighbouring cells, which considerably reduces the risk of overcompensation. Fourthly, cross-platform spatial omics data integration was achieved by aligning 3 types of imaging data (i.e., multispectral imaging, H&E-stained imaging and MALDI-MSI), despite all having different spatial resolution. Our approach captures fluorescence and mass spectrometry signals from the same region of the tissue (assessed in adjacent sections) without the need for expensive and laborious new technology, paving the way for *in-situ* multi-omics analysis and integration. Fifthly, the pipeline is flexible and can, with minor adaptation, also be used for other (omics) imaging platforms, such as spatial transcriptomics (e.g. seqFISH+[38]) and imaging mass cytometry (IMC). It even

allows linking more layers of omics data, to build an integrated multi-omics model of the tissue's cellular phenotypes.

In the past two years, several approaches have also been proposed to analyse single cell heterogeneity at limited spatial resolution. For example, to compensate for the lack of spatial transcriptomics resolution, STRIDE deconvolutes spatial transcriptomics spots based on the topic profiles trained from a reference scRNA-Seq data, to estimate the proportion of cell types [39]. To improve the resolution of MALDI-MSI, Smets et al. fused reduced-dimensional MSI datasets (based on UMAP) and H&E images using Correspondence-Aware Manifold Learning (CAML) [40]. While an interesting approach, the nonlinear dimensionality reduction failed to extract the ion characteristics from the fused images. Some algorithms focus on integrated analysis of multi-omics data from the same cell (e.g., MUSE [41]). Nevertheless, when the input image resolution is insufficient (i.e. larger than a cell), MUSE can only identify the heterogeneity of tissue regions based on annotated information. In contrast to these methods, we were the first to integrate multispectral, histological, and MALDI-MSI data to identify individual cell phenotypes and analyse cellular communities and lipid profiles.

Despite the innovations mentioned above, a potential limitation of our approach remains the limited spatial resolution of MALDI-MSI. As the MSI pixel size of atherosclerotic lesions collected from mice aorta was approximately equal to the size of a cell segment from the multispectral imaging layer, the phenotype to which an MSI pixel corresponded was assigned according to the coordinates of the centroid of cell segments. Thus, the signal of an MSI pixel may be contaminated by that from adjacent cell segments. To address this problem, in **chapter 2**, we examined the cell-pixel correspondence and removed a few (21 of 12,016) pixels that covered the centroids of two cell segments, to reduce the risk of cross-cell contamination. To verify the effectiveness of this strategy, we weighed the mass spectra of each cell by the proportion of a cell's area occupied by a pixel and obtained essentially similar results. However, recent technological improvements have tackled this bottleneck of insufficient spatial resolution. For example, recently a High-plex Multiomic Analysis platform, Nanostring CosMX [42], was introduced; it offers increased spatial resolution to the subcellular level (< 50 nm) for protein and transcriptomics measurements. In addition, t-MALDI-2 claims to achieve a spatial resolution of 600 nm [43]. By integrating the imaging data generated by these novel technologies, we anticipate that our approach will enable real spatial single-cell multi-omics integration.

Furthermore, the molecular resolution of MALD-MSI is also an issue, as it limits the unequivocal assignment of peaks to chemical entities and thus the biological interpretation of findings. Even though Tandem Mass Spectrometry (MS/MS)[44] was applied in **chapter 2** for validating a selection of lipid identities, the limited

number of ions that can be analysed simultaneously by MS/MS makes it difficult to apply pathway enrichment analyses to infer the subset-specific lipidomic pathways.

In conclusion, we have pioneered an *in-situ* cross-platform multi-omics computational analysis approach, which opens a new avenue for integrating multi-omics imaging data with non-equal resolutions. By applying our pipeline on aortic tissue from mice with atherosclerosis, we observed that the cell phenotypes defined by multispectral imaging intensities differed in mass spectra, demonstrating the validity of our approach. We are currently designing a user-friendly interactive interface to facilitate the analysis and integration of multispectral and MSI findings for biologists; such tool will benefit researchers in their exploration of cell composition, function, and context *in vitro* and in healthy and diseased tissue.

## Space- and time-resolved omics analysis: a new direction for future study of cell heterogeneity

MALDI-MSI is an advanced imaging technique for visualizing information on the spatial distribution of various molecular types (e.g., metabolites, lipids, peptides) and has been used to study the spatial heterogeneity of various tissues [45–47]. Time course studies with MALDI-MSI allowed us to dissect the spatial dynamics of metabolism in response to a disease trigger. In **chapter 3**, we analysed metabolomic MALDI-MSI data from the livers of *Lldr<sup>-/-</sup>* mice during the early stages of NAFLD, to understand the metabolic changes associated with progression of NAFLD in different liver compartments.

A few studies have already applied MALDI-MSI to explore the heterogeneity of liver tissue [46, 48]. However, the pursued strategy required manual selection of “region of interests” in immunohistochemically stained images flanking the MSI-analysed section, subsequent alignment thereof with the MSI images and finally identification of metabolic features in these regions using supervised or unsupervised methods. Our approach does not require H&E-stained images, but directly trains the SSC classification model based on MALDI-MSI data from healthy mouse livers, to identify three compartments in the liver tissues of the disease stages: parenchyma, sinusoids, and vessels. Our study not only demonstrates the presence of compartment-specific metabolites, but also avoids errors arising from manual co-registration of images and time-consuming manual annotation. However, it is less accurate, as it relies on metabolic differences between compartments and does not take the heterogeneity of the compartments itself into account. For instance, sinusoids are known to harbor both endothelial cells, Kupffer cells, perivascular macrophages and stellate cells, each likely to display their own metabolomic identity. Furthermore, some MALDI-MSI-based methods for liver tissue analysis focus on static comparison of tissues in disease and control states for characterization and subsequent visualization of the corresponding ion images, and therefore fail to capture the metabolic dynamics involved in the progression of NAFLD [49–51]. In **chapter 3**, we adopted a Dirichlet process Gaussian process mixture model (DPGP), originally designed for time series of gene expression profiles, to metabolomics data and applied it on the compiled metabolite datasets from two identified liver compartments: parenchyma and sinusoids. For each compartment, metabolites were clustered according to their trajectories over time, thus providing a comprehensive profile of the metabolic changes of that particular liver compartment during the early stages of NAFLD.

In addition to the limitations discussed in **chapter 3**, the metabolite clusters were limited by the clustering method, since one of the assumptions of GPDs is that each  $m/z$  peak belongs to only one class. Given the complex chemical reactions and interactions between metabolites, future research may involve soft clustering or

graph-based clustering approaches to further improve the consistency of clustering results with the actual biological systems.

Overall, the analysis pipeline we proposed is able to not only distinguish liver compartments based on MALDI-MSI data, but also identify metabolic changes associated with the initiation and progression of NAFLD, thus providing new leads for the discovery of biomarkers for liver compartments and early stage of NAFLD. Although spatio-temporal omics analysis is still in its infancy, we expect that as the resolution (both in mass and space) of spatial omics improves and the costs of detection decrease, more and more computational strategies will be developed to herald a new era of spatio-temporal histological analysis and really understand the dynamics of disease at molecular and cellular level.

## Concluding remarks and future perspective

In this chapter, we have compared the potential of statistical and machine learning-based computational approaches for analysis of different biomedical data types and research purposes. We highlighted the application scenarios of these approaches in this thesis (**chapter 2 to 5**) and showed their effectiveness in analysing omics data to dissect molecular mechanisms in disease processes. Moreover, we described the basic concepts behind and procedure of enrichment analysis methods, discussed the applicability of various types of approaches with examples from this thesis (**chapter 3, 4 and 5**), and emphasized the important impact of public resources and statistical methods on inferred results. We demonstrated the superiority of a cross-platform spatial multi-omics integration pipeline (**chapter 2**) as compared to conventional and recent spatial omics analysis methods in dissecting cell heterogeneity as well as the cells' microenvironment, and explored the flexibility and broader applicability of the approach. Finally, we discussed advances in MALDI-MSI-based spatio-temporal metabolomic analysis for liver tissue heterogeneity and pointed out the advantages and limitations of the approach presented in **chapter 3** for identifying liver compartments and capturing metabolite dynamics associated with disease processes.

Omics technology is rapidly developing at several levels: from bulk tissue analysis to single-cell analysis, from individual images to image cubes, from single timepoint to time series studies, and from single-omics to multi-omics. Updates to these techniques are accompanied by increasingly complex data, attracting data scientists and bioinformaticians to design effective tools to assist biologists in solving specific challenges. Examples include single-cell analyses to map cell heterogeneity in tissue; imaging analyses that probe the association between cells and their microenvironment, and time-series analyses that track changes in genes, proteins, or metabolites upon a disease or stimulus challenge. However, current computational methods are still constrained by the spatial resolution and the number of features that can be analysed simultaneously. As the spatial resolution of imaging technologies increases, I anticipate that, eventually, approaches that can integrate multi-omics information and simultaneously obtain spatial distribution and temporal dynamics at single-cell resolution will become a new hotspot. Undoubtedly, this will allow us to gain a more comprehensive insight into the complexity and diversity of cells in diseased tissues, contributing significantly to deciphering the mechanisms of cardiometabolic and other diseases and thereby reducing morbidity and mortality.



## References

1. Aksentijevich M, Lateef SS, Anzenberg P, Dey AK, Mehta NN. Chronic Inflammation, Cardiometabolic Diseases and Effects of Treatment: Psoriasis as a Human Model. *Trends Cardiovasc Med.* 2020;30:472–8.
2. Schüler R, Brand A, Klebow S, Wild J, Veras FP, Ullmann E, et al. Antagonization of IL-17A Attenuates Skin Inflammation and Vascular Dysfunction in Mouse Models of Psoriasis. *Journal of Investigative Dermatology.* 2019;139:638–47.
3. Yang J, Zhang L, Yu C, Yang X-F, Wang H. Monocyte and macrophage differentiation: circulation inflammatory monocyte as biomarker for inflammatory diseases. *Biomarker Research.* 2014;2:1.
4. Ley K, Miller YI, Hedrick CC. Monocyte and Macrophage Dynamics during Atherogenesis. *Arterioscler Thromb Vasc Biol.* 2011;31:1506–16.
5. Orihuela R, McPherson CA, Harry GJ. Microglial M1/M2 polarization and metabolic states. *Br J Pharmacol.* 2016;173:649–65.
6. Chang Q, Ornatsky OI, Siddiqui I, Loboda A, Baranov VI, Hedley DW. Imaging Mass Cytometry. *Cytometry Part A.* 2017;91:160–9.
7. Buchberger AR, DeLaney K, Johnson J, Li L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal Chem.* 2018;90:240–65.
8. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnology Advances.* 2021;49:107739.
9. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nature Methods.* 2018;15:233–4.
10. Stewart M. The Actual Difference Between Statistics and Machine Learning. Medium. 2020. <https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3>. Accessed 10 Aug 2022.
11. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research.* 2015;43:e47.
12. Hoefler IE, Sels JW, Jukema JW, Bergheanu S, Biessen E, McClellan E, et al. Circulating cells as predictors of secondary manifestations of cardiovascular disease: design of the CIRCULATING CELLS study. *Clinical Research in Cardiology.* 2013;102:847–56.
13. Statistics Ready for a Revolution | Amstat News. 2010. <https://magazine.amstat.org/blog/2010/09/01/statrevolution/>. Accessed 24 Aug 2022.

14. Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, et al. Study of cardiovascular disease prediction model based on random forest in eastern China. *Sci Rep.* 2020;10:5245.
15. Jiang Y, Zhang X, Ma R, Wang X, Liu J, Keerman M, et al. Cardiovascular Disease Prediction by Machine Learning Algorithms Based on Cytokines in Kazakhs of China. *Clin Epidemiol.* 2021;13:417–28.
16. Bemis KD, Harry A, Eberlin LS, Ferreira CR, Ven SM van de, Mallick P, et al. Probabilistic Segmentation of Mass Spectrometry (MS) Images Helps Select Important Ions and Characterize Confidence in the Resulting Segments \*. *Molecular & Cellular Proteomics.* 2016;15:1761–72.
17. Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics.* 2006;7:3.
18. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
19. McDowell IC, Manandhar D, Vockley CM, Schmid AK, Reddy TE, Engelhardt BE. Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLOS Computational Biology.* 2018;14:e1005896.
20. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, et al. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell.* 2018;174:968-981.e15.
21. Pollard KS, Laan MJVD. A method to identify significant clusters in gene expression data. In: *Sixth World Multiconference on Systemics, Cybernetics, and Informatics.* 2002. p. 318–25.
22. Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining.* Portland, Oregon: AAAI Press; 1996. p. 226–31.
23. Jiang Y, Zhang X, Ma R, Wang X, Liu J, Keerman M, et al. Cardiovascular Disease Prediction by Machine Learning Algorithms Based on Cytokines in Kazakhs of China. *CLEP.* 2021;13:417–28.
24. Pal M, Parija S, Panda G, Dhama K, Mohapatra RK. Risk prediction of cardiovascular disease using machine learning classifiers. *Open Medicine.* 2022;17:1100–13.
25. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc.* 2019;14:482–517.

26. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*. 2016;44:D457–62.
27. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al. WikiPathways: connecting communities. *Nucleic Acids Research*. 2021;49:D613–21.
28. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res*. 2019;29:1363–75.
29. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171:1437–1452.e17.
30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005;102:15545–50.
31. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet*. 2016;48:838–47.
32. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, et al. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*. 2006;7:S7.
33. Jiang P, Zhang Y, Ru B, Yang Y, Vu T, Paul R, et al. Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat Methods*. 2021;18:1181–91.
34. Yurdagul A Jr, Finney AC, Woolard MD, Orr AW. The arterial microenvironment: the where and why of atherosclerosis. *Biochemical Journal*. 2016;473:1281–95.
35. Hansson GK. Immune and inflammatory mechanisms in the development of atherosclerosis. *Br Heart J*. 1993;69 1 Suppl:S38–41.
36. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med*. 2018;50:1–14.
37. Spitzer MH, Nolan GP. Mass Cytometry: Single Cells, Many Features. *Cell*. 2016;165:780–91.
38. Eng C-HL, Lawson M, Zhu Q, Dries R, Kouloua N, Takei Y, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*. 2019;568:235–9.

39. Sun D, Liu Z, Li T, Wu Q, Wang C. STRIDE: accurately decomposing and integrating spatial transcriptomics using single-cell RNA sequencing. *Nucleic Acids Research*. 2022;50:e42.
40. Smets T, De Keyser T, Tousseyn T, Waelkens E, De Moor B. Correspondence-Aware Manifold Learning for Microscopic and Spatial Omics Imaging: A Novel Data Fusion Method Bringing Mass Spectrometry Imaging to a Cellular Resolution. *Anal Chem*. 2021;93:3452–60.
41. Bao F, Deng Y, Wan S, Shen SQ, Wang B, Dai Q, et al. Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. *Nat Biotechnol*. 2022;40:1200–9.
42. He S, Bhatt R, Brown C, Brown EA, Buhr DL, Chantranuvatana K, et al. High-plex Multiomic Analysis in FFPE at Subcellular Level by Spatial Molecular Imaging. 2022;:2021.11.03.467020.
43. Niehaus M, Soltwisch J, Belov ME, Dreisewerd K. Transmission-mode MALDI-2 mass spectrometry imaging of cells and tissues at subcellular resolution. *Nat Methods*. 2019;16:925–31.
44. Bythell BJ, Hendrickson CL, Marshall AG. Relative Stability of Peptide Sequence Ions Generated by Tandem Mass Spectrometry. *J Am Soc Mass Spectrom*. 2012;23:644–54.
45. Shi Y, Johnson J, Wang B, Chen B, Fisher GL, Urabe G, et al. Mass Spectrometric Imaging Reveals Temporal and Spatial Dynamics of Bioactive Lipids in Arteries Undergoing Restenosis. *J Proteome Res*. 2019;18:1669–78.
46. Saito T, Watanabe A, Nakano M, Matsuo K. MALDI-TOF mass spectrometry imaging for N-glycans on FFPE tissue sections of mouse NASH liver through Sialic acid Benzylamidation. *Glycoconj J*. 2021;38:167–75.
47. Grove KJ, Hoque S, Rudewicz PJ. Investigation of amodiaquine localization in liver lobules using matrix-assisted laser desorption/ionization imaging mass spectrometry. *Rapid Communications in Mass Spectrometry*. 2019;33:252–8.
48. Zhang Q, Wu Z-H, Zhao S-S, Yang J, Chen L, Wang X-Y, et al. Identification and Spatial Visualization of Dysregulated Bile Acid Metabolism in High-Fat Diet-Fed Mice by Mass Spectral Imaging. *Frontiers in Nutrition*. 2022;9.
49. Ščupáková K, Soons Z, Ertaylan G, Pierzchalski KA, Eijkel GB, Ellis SR, et al. Spatial Systems Lipidomics Reveals Nonalcoholic Fatty Liver Disease Heterogeneity. *Anal Chem*. 2018;90:5130–8.
50. Rodríguez-Calvo R, Samino S, Girona J, Martínez-Micaelo N, Ràfols P, García-Altres M, et al. Hepatic Lipidomics and Molecular Imaging in a Murine Non-

Alcoholic Fatty Liver Disease Model: Insights into Molecular Mechanisms. *Biomolecules*. 2020;10:1275.

51. Alamri H, Patterson NH, Yang E, Zoroquiain P, Lazaris A, Chaurand P, et al. Mapping the triglyceride distribution in NAFLD human liver by MALDI imaging mass spectrometry reveals molecular differences in micro and macro steatosis. *Anal Bioanal Chem*. 2019;411:885–94.

# **Chapter 7**

Summary

Samenvatting

总结



Worldwide, cardiometabolic diseases (e.g., cardiovascular disease (CVD), diabetes, and non-alcoholic fatty liver disease (NAFLD)) have acquired almost epidemic proportions in the past few decades due to the widespread adoption of a western lifestyle. They compromise heart and liver functions and underlie the two main causes of death worldwide, ischemic heart disease and stroke. The development of these diseases is characterized by lipid accumulation, inflammatory responses, and metabolic dysfunction in the arterial wall (a process referred to as atherosclerosis) and liver (“fatty liver disease” or NAFLD). For both, macrophages and their precursors, monocytes, are important contributors.

The rapid development in high-throughput and imaging technologies have enabled researchers to detect profiles of genes, proteins, and metabolites within individuals and cells. These techniques are increasingly applied to study cardiometabolic diseases. However, their potential to explore the pathogenesis has not been fully exploited. In this thesis, we aim to analyse high-dimensional omics and imaging data of plaque and liver through a combination of computational strategies including statistical inference, machine learning and image processing, to dissect the pathogenesis of atherosclerosis and NAFLD at the molecular and spatial level.

Macrophages exhibit extreme plasticity, adopting a broad range of micro-environment-driven phenotypes. Exploring their heterogeneity throughout the course of atherosclerosis will be crucial to the discovery of new diagnostic and therapeutic measures. However, most of the commonly used techniques for quantifying cell markers cannot simultaneously measure sufficient features to properly describe the heterogeneity or do not provide spatial information at single-cell resolution. In **chapter 2**, we proposed a computational pipeline to integrate multiple imaging data layers from the same sample but derived from different platforms (i.e., multispectral imaging, MALDI mass spectrometry imaging (MALDI-MSI), and histologic imaging). Our approach enabled to identify cell phenotypes, pinpoint their locations in the tissue, reveal the cell community landscape, and characterize the metabolic environment of identified cell subsets. Our method demonstrated its effectiveness in analysing macrophage heterogeneity in atherosclerotic plaques but will also be applicable to other types of tissues, diseases and cells.

The metabolism of hepatocytes has been shown to be severely disrupted with the development of NAFLD, as metabolomic studies have shown. However, most metabolomics efforts that mapped the dynamics of these changes lacked spatial information. In **chapter 3**, we fused both supervised and unsupervised MSI image segmentation algorithms to deploy a liver compartment recognition model, resulting in the identification of three distinct liver compartments in the early stage of murine NAFLD, which were assigned “parenchyma”, “sinusoids” and “vessels”, based on their localisation and morphology. Metabolites were then clustered



according to their trajectories over the disease progression for the individual compartments. Our approach identified liver compartments solely based on the metabolite features from MALDI-MSI, indicating the presence of compartment-specific metabolic pathways, while avoiding the potential errors in aligning histological images with MALDI-MSI images, as well as extensive manual annotation on histological images. In addition, spatio-temporal analysis of mouse liver helps to monitor the metabolite dynamics during the disease progression within the separate compartments, leading to a deeper understanding of the processes that occur in the early stage of NAFLD.

Several risk factors have been identified, that increase the chance of developing cardiometabolic diseases, of which some are linked to an unhealthy lifestyle. However, it is unclear if these risk factors impact circulating monocytes, precursors of macrophages and thereby important players in many diseases. Therefore, in **chapters 4 and 5**, we aimed to identify monocyte key gene programs and pathways associated with the risk factors of CVD (which is mainly caused by atherosclerosis) by analysing the monocyte expression profiles of a CVD cohort collected by the Center for Translational Molecular Medicine (CTMM). We first studied disease-relevant sex differences in the transcriptional makeup of monocytic cells (**chapter 4**). For this purpose, we compared monocyte expression profiles of male and female CVD patient in the CTMM cohort, and generated sex-biased gene signatures. Then we adopted multiple enrichment analysis approaches based on several publicly available pathway, transcription factor (TF)-target interaction, and cytokine data resources, to infer pathway, TF, and cytokine signalling activities of male vs female monocytes. This led us to pinpoint female- and male-skewed cytokine signalling cascades in CVD. Validation of our findings in an independent healthy subject cohort showed that the observed sex differences were CVD-specific. In addition, we constructed a cytokine-TF network by inferring TF activities for the expression profiles of the identified sex-specific cytokine signalling pathways, which revealed Jun-B as central activator of most female-specific pathways. In **chapter 5**, we investigated the correlation between other CVD risk factors and the response capacity of CVD patient-derived monocytes to an inflammatory stimulus (lipopolysaccharide). As we show, CVD patients with high (diastolic) blood pressure tended to have a weaker monocyte inflammatory response. We then constructed a gene co-regulation and corresponding regulatory network and identified a sub-network with strong correlation with blood pressure. Moreover, this network pointed to a dysfunction in the cells' energy production (mitochondrial respiration). Finally, we pursued to a network-guided drug repurposing approach to identify iloprost as potential candidate to target this network. This drug, which is currently used to treat pulmonary hypertension, could potentially target the network that connects a weakened inflammatory response in monocytes (and potentially other cells) to diastolic hypertension.

In **chapter 6**, I summarised the main findings of my thesis, further explained the computational approaches used from **chapters 2 to 5** and explored potential application scenarios thereof in the study of cardiometabolic diseases, including their advantages and remaining limitations, and provided an outlook on future related areas.

Wereldwijd hebben cardiometabole aandoeningen, zoals aderverkalking, diabetes, suikerziekte en leververvetting en -ontsteking, de laatste decennia bijna epidemische proporties aangenomen, deels als gevolg van een westerse levensstijl. De verminderde hart- en leverfunctie waarmee zij gepaard gaan liggen ten grondslag aan de twee belangrijkste doodsoorzaken wereldwijd, ischemische hartziekten en beroerten. De ontwikkeling van deze ziekten wordt gekenmerkt door vetophoping, ontstekingsreacties en metabolische disfunctie in de slagaderwand (een proces dat atherosclerose wordt genoemd) en de lever (leververvetting of NAFLD). Macrofagen en hun voorlopercel, de monocyt, dragen voor beide aandoeningen belangrijk bij aan het ziekteproces .

Snelle ontwikkelingen op het gebied van high-throughput en beeldvormingstechnologieën hebben onderzoekers in staat gesteld om gen-, eiwit- en metabolietprofielen van weefsel en cellen snel en volledig in kaart te brengen. Deze technieken worden in toenemende mate toegepast om cardiometabole ziekten te bestuderen. Hun onderzoekspotentieel voor deze ziekten is echter nog niet volledig benut. In dit proefschrift willen we hoog-dimensionale omics en beeldvormingsdata van plaque en lever analyseren door gebruik te maken van een combinatie van computationele strategieën, waaronder statistische inferentie, machinaal leren en beeldverwerking, met als uiteindelijke doel de ontwikkeling van atherosclerose en NAFLD op moleculair en ruimtelijk niveau te ontrafelen.

Macrofagen vertonen een extreme plasticiteit en kunnen een breed scala aan functies aannemen afhankelijk van de locatie en de fase van het ziekteproces. Het onderzoeken van hun heterogeniteit zal van cruciaal belang zijn voor de ontdekking van nieuwe diagnostische en therapeutische mogelijkheden. Echter, de meest gebruikte technieken voor het kwantificeren van cel-markers kunnen niet voldoende kenmerken tegelijk meten om de cellulaire heterogeniteit voldoende in kaart te brengen, of geven geen ruimtelijke informatie over de cellen. In **hoofdstuk 2**, stelden we een integratie-pijplijn voor om meerdere lagen van beeldvorming, uitgevoerd op eenzelfde monster via verschillende microscopische en chemische analysetechnieken (multispectrale beeldvorming, MALDI massaspectrometrie beeldvorming (MALDI-MSI) en histologische beeldvorming), te integreren tot een ruimtelijk model van het zieke weefsel. Onze aanpak maakte het mogelijk om de verschijningsvorm, locatie en directe moleculaire (en cellulaire) omgeving van elke cel in het zieke weefsel in kaart te brengen. Onze methode is al succesvol toegepast bij het analyseren van macrofaag heterogeniteit in aderverkalking maar is zeker ook toepasbaar op andere types van weefsels, ziekten en cellen.

Het metabolisme van levercellen blijkt ernstig verstoord te zijn tijdens de ontwikkeling van NAFLD, zoals metabolomics studies al eerder hebben aangetoond. Echter, de meeste metabolomics studies waarin de dynamiek van deze veranderingen geanalyseerd is misten ruimtelijke informatie, terwijl MALDI-MSI

analyses van leverweefsel geen informatie verstrekten over de veranderingen doorheen het ziekteverloop. In **hoofdstuk 3** hebben we zowel gestuurde als ongestuurde MSI beeldbewerkingsalgoritmen toegepast om de diverse levercompartimenten te kunnen onderscheiden; dit resulteerde in de identificatie van drie verschillende compartimenten in muizenlever: "parenchym", "sinusoïde" en "bloedvat". Voor elk individueel compartiment werden vervolgens de metabolieten geclusterd op basis van hun expressieprofiel in de loop van de ziekteprogressie. Onze aanpak bleek in staat de belangrijke structurele compartimenten in muizenlever te identificeren, enkel op basis van hun metabolietprofiel en ongeacht het ziektestadium. Dit wijst niet alleen op de aanwezigheid van compartiment-specifieke metabolieten, maar maakt het ook mogelijk potentiële fouten in het uitlijnen van histologische beelden bij het integreren van verschillende datalagen te omzeilen. Bovendien stelde dit ons in staat om het ziekteproces voor elk compartiment afzonderlijk te volgen in de tijd. Dit laatste leidde tot een dieper inzicht in de metabole processen in het beginstadium van NAFLD.

De risicofactoren die bijdragen aan de ontwikkeling van cardiometabole ziekten zijn inmiddels grotendeels bekend en diverse factoren blijken daarbij verband te houden met een ongezonde levensstijl. Het is echter onduidelijk of deze risicofactoren ook invloed hebben op een ontstekingsceltype in het bloed, de monocyt, die betrokken is bij de ontwikkeling van de bovengenoemde cardiometabole aandoeningen. Daarom hebben wij in hoofdstuk 4 en 5 getracht de belangrijkste genprogramma's en signaaltransductieroutes te identificeren die in dit celtype correleren met risicofactoren voor hart- en vaatziekten. Hiervoor hebben we de beschikbare expressieprofielen geanalyseerd van monocytten uit een CVD-cohort dat verzameld is in het kader van het Center for Translational Molecular Medicine programma (CTMM). We bestudeerden daarbij voor het ziekteproces relevante man-vrouw verschillen in het expressieprofiel van monocytten (**hoofdstuk 4**) en waren in staat geslachts-specifieke genetische signaturen voor hart- en vaatziekten aan te tonen. Vervolgens hebben we meerdere analyses uitgevoerd op basis van openbaar beschikbare gegevensbronnen om de belangrijkste signaaltransductieroutes in dit celtype in kaart te brengen die kenmerkend waren voor mannelijke en vrouwelijke patiënten met hart- en vaatziekten. Validatie van onze bevindingen in een onafhankelijk cohort van gezonde proefpersonen toonde aan dat de waargenomen sekseverschillen inderdaad enkel golden voor hartpatiënten. Daarnaast hebben we een genregulatiernetwerk gebouwd door transcriptiefactor-activiteiten af te leiden uit de expressieprofielen van de geïdentificeerde sekse-specifieke cytokine-signaleringspaden. Jun-B kwam daarbij naar voren als centrale activator van de het gros van de vrouw-specifieke paden.

In **hoofdstuk 5** hebben we de correlatie onderzocht tussen andere risicofactoren en de reactie van monocytten in patiënten met hart- en vaatziekten op een

ontstekingsprikkel (lipopolysaccharide). We hebben aangetoond dat CVD-patiënten met een hoge (diastolische) bloeddruk een zwakkere ontstekingsreactie vertoonden. Vervolgens hebben we een gen (co)regulatiernetwerk geconstrueerd en daarbij een regulatiernetwerk geïdentificeerd met een sterk correlatie met hoge bloeddruk. Dit netwerk wees op een verminderd functioneren van de cellulaire energieproductie (mitochondriale ademhaling). Tenslotte hebben we een aantal geneesmiddelen kunnen selecteren, via een computeralgoritme, waarvan het aannemelijk is dat ze dit netwerk (en daarmee de bloeddruk) kunnen corrigeren. Een van deze geneesmiddelen, iloprost, wordt momenteel al gebruikt om pulmonale hypertensie te behandelen, maar zou dus deels via correctie van de verzwakte afweerreactie van monocytten kunnen werken.

In **hoofdstuk 6** heb ik de belangrijkste bevindingen van mijn proefschrift samengevat, de in de hoofdstukken 2 tot en met 5 gebruikte rekenkundige technieken en mogelijke toepassingsscenario's daarvan in het onderzoek naar cardiometabole ziekten nader toegelicht, de voordelen en tekortkomingen ervan besproken en ideeën voor verdere verfijning en verbetering van deze technologieën gepresenteerd.

由于不健康的生活方式的广泛流行，在过去的几十年里代谢性心血管病 (cardiometabolic diseases)，包括心血管疾病 (CVD)、糖尿病和非酒精性脂肪肝 (NAFLD)，已成为影响全球经济社会发展的重大公共卫生问题。这些疾病严重损害心脏和肝脏的功能，已成为全世界两个主要的死亡原因--缺血性心脏病和中风的首要致病因素。这些疾病的发展特点是脂质积累、炎症反应和动脉壁（被称为动脉粥样硬化的过程）和肝脏 (NAFLD) 的代谢功能紊乱。对于这两种情况，巨噬细胞和它们的前体单核细胞，都是重要的贡献者。

高通量和成像组学技术的快速发展使研究人员能够检测个人和细胞内的基因、蛋白质和代谢物的概况。然而，目前的技术在探索心脏代谢性疾病的发病机制方面的潜力还没有得到充分的开发。在这篇论文中，我们旨在通过统计推理、机器学习和图像处理等计算策略的组合来分析高维的全息影像数据，在分子水平上研究两种最常见的代谢性心血管病--心血管疾病(CVD)和非酒精性脂肪肝(NAFLD)的发病机理。

巨噬细胞表现出极强的可塑性，在动脉粥样硬化的过程中根据环境的不同而演化出一系列表型。探索它们的异质性对于发现新的诊断和治疗措施至关重要。然而，大多数量化细胞标志物的技术不能测量足够的特征来正确描述异质性，或者不能提供单细胞分辨率下细胞的位置信息。在第二章中，我们提出了一个综合计算管道，以整合来自同一样品但不同平台的多种成像数据（即多光谱成像、MALDI 质谱成像 (MALDI-MSI) 和组织学成像）。我们的方法能够识别细胞表型，准确定位它们在组织上的位置，提供一个细胞类型相互作用的视图，并描述它们的代谢环境。我们的方法不仅证明了其在分析动脉粥样硬化斑块中巨噬细胞异质性的有效性，而且也适用于其他类型的组织、疾病和细胞。

已有研究表明，随着非酒精性脂肪肝的发展，肝细胞的代谢会受到严重破坏。然而，许多关于肝脏组织 MALDI-MSI 的研究仍然缺乏对代谢物随疾病进展的动态变化的追踪以及对空间异质性的剖析。在第三章中，我融合了有监督和无监督的 MSI 图像分割算法，部署了一个肝脏分区识别模型，从而识别了非酒精性脂肪肝早期阶段的小鼠的三个不同的肝脏分区：实质、肝窦和血管。然后，根据各个区间的疾病进展轨迹对代谢物进行聚类。该方法仅根据 MALDI-MSI 的代谢物特征来确定肝脏分区，表明分区特定代谢物的存在，从而避免了将组织学图像与 MALDI-MSI 图像对齐的潜在错误，以及对组织学图像的大量人工注释。此外，小鼠肝脏的时空分析有助于识别疾病发展过程中不同区间内代谢物的动态变化，从而更深入地了解非酒精性脂肪肝的代谢紊乱。

7

尽管一些生物学特征和不健康的生活方式与心血管疾病风险的增加有关,但目前还不清楚这些因素如何影响心血管疾病中最重要的细胞类型之一:循环单核细胞的活动。因此,在第四章和第五章中,我们旨在通过分析转化分子医学中心(CTMM)收集的 CVD 队列的单核细胞表达谱,确定与 CVD 风险因素相关的单核细胞关键基因程序和途径。我们首先在第四章中研究了明显的性别差异在多大程度上反映了单核细胞中转录构成的差异。我们通过比较 CTMM 队列中的男性和女性 CVD 患者,产生了性别偏见的基因特征,然后根据一些公开的路径、转录因子-靶基因相互作用和细胞因子资源,采用多种富集分析方法推断路径、转录因子和细胞因子信号活动。这使我们确定了心血管疾病中女性和男性激活的途径和细胞因子。在一个独立的健康队列中的验证结果也显示,这些基于单核细胞转录谱推断的性别差异是心血管疾病特有的。此外,我们通过推断细胞因子诱导的转录谱的转录因子活性,构建了一个细胞因子-转录因子网络,该网络显示 *jun-B* 是大多数女性特定途径的激活剂。在第五章中,我们研究了心血管疾病风险因素与心血管疾病受试者单核细胞先天免疫反应能力之间的相关性,这使我们观察到高血压的 CVD 患者的单核细胞对 LPS 的反应较弱。然后,我们根据 CVD 患者的 LPS 反应谱构建了一个基因共调和相应的调控网络,发现与血压水平呈现最强相关性的子网络与氧化磷酸化途径和呼吸电子链运输有关。药物再利用锁定了伊洛前列腺素,一种用于治疗肺动脉高压的药物,能够增强 LPS 反应,这在一定程度上证实了我们对高血压和 LPS 反应之间关联的推断。

第六章总结了本论文的主要发现,进一步解释了第二章至第五章所使用的计算方法及其在心脏代谢性疾病研究中的应用场景,探讨了它们的优势和局限性,并对相关领域的未来进行了展望。

# Chapter 8

Impact





## Scientific Impact

In the past decade, high-throughput technologies are booming and have propelled biomedical research to a new level. The development of single-cell omics (e.g., scRNA-Seq and CyTOF) has enabled researchers to uncover rare cell types and detect cellular heterogeneity. This has made it a valuable tool not only for studying cardiometabolic diseases, but also in other disease domains such as cancer, immunology, microbiology, and neurology [1, 2]. The pathologies of these diseases are typically complex and closely related to the microenvironment of the cells in the diseased tissue. However, as all information of the cell's location is lost due to tissue dissociation steps before analysis, these single-cell technologies are unable to analyse the microenvironment of cells. Some spatial omics techniques, such as 10x Visium[3] (transcriptomics), are unable to identify the phenotypes of individual cells due to their low spatial resolution. In addition, some imaging techniques either do not capture enough features (e.g., immunofluorescent microscopy (IF)) or are expensive (e.g., CODEX [4]). Therefore, biologists urgently require affordable and effective ways to identify cell types while visualizing their spatial distributions and dissecting the cellular and molecular microenvironments. In response to this pressing requirement, I proposed a novel computational pipeline in **chapter 2**. The pipeline implements several functions. First, it allows biologists to distinguish cell phenotypes in an interactive manner based on the relative intensities of multiplex signals in multispectral imaging. Secondly, it visualizes individual cells and cell communities on histological images, helping pathologists to quickly verify their identities. Thirdly, this approach links multiple omics imaging data (i.e., multispectral imaging and MSI) with the corresponding histological images in tissue, allowing to dissect the molecular environment of cells from multiviews. As this approach is affordable and applicable to a wide range of cell types and tissues, it may represent a breakthrough in linking molecular context to cellular phenotype and function in healthy and diseased tissues. In the end, this not only provides new insights into pathogenesis of disease and leads for prevention and/or treatment of cardiovascular disease, cancer, neurological disorders, and a range of other diseases, the recently developed user-friendly interface also brings MSI/multispectral image analysis within the reach of biologists and pathologists, who are not or marginally skilled in R or Matlab.

In the field of spatial metabolomics, MALDI-MSI detects metabolite levels in tissue while preserving spatial information [5]. However, to study metabolite differences across different sections of the tissue, researchers always manually designate the regions of interest (ROIs) on the histological image for each section, and then align the histological image with the optical image from MALDI-MSI. This requires a significant amount of time and effort. Moreover, this approach is error-prone especially if tissue sections used for histology and MS imaging differ (for instance

due to the laser-inflicted tissue damage). In **chapter 3**, I presented a new strategy to identify liver compartments using a combination of supervised and unsupervised MSI segmentation algorithms, with circumvents the need of a histological image. Accurate segmentation of parenchyma, sinusoid, and vessel on MALDI-MSI images indicates the presence of compartment-specific metabolites that are not associated with disease progression. Validation for these identified metabolites is expected to completely free researchers from the time-consuming and repetitive annotation work and to also bring breakthroughs in future studies of metabolism in liver.

In **chapter 4** and **5**, I interrogated a monocyte transcriptomics dataset to extract sex-specific cytokine signalling pathways and hypertension-associated gene networks, respectively. Although both findings still need validation in an independent genetic or genomics dataset, they could pave the way for the design of new drugs for tailored intervention in ischemic heart disease or hypertension-associated heart failure. Moreover, as monocyte isolation is only mildly invasive, the leads could represent new genetic diagnostics for early stages of heart failure and microcirculatory dysfunction, or for the effectiveness of sex-specific interventions in CVD development in women.

### Societal Impact

Therapeutic intervention in cardiometabolic diseases is complicated by the profound heterogeneity of disease-driving inflammatory cells. However, this heterogeneity was hitherto only poorly studied. The comprehensive pipeline proposed in **chapter 2** allows to identify an unprecedented number of myeloid phenotypes in murine atherosclerotic tissues and to dissect the cellular and molecular microenvironment associated with these phenotypes. With this approach we are able to demonstrate that plaque myeloid phenotypes often have a unique and characteristic cellular and molecular environment, which offers the possibility of treating atherosclerosis by altering the macrophage microenvironment. Such new insights could eventually benefit many patients as they allow the design of more tailored precision medicines that target the right subset in the right patient.

Non-alcoholic fatty liver disease (NAFLD) is a heterogeneous and complex disease that affects approximately 20% to 25% of Europeans and 30% to 40% of Americans[6, 7]. Metabolic disturbances are one of the main features of NAFLD progression [8]. Therefore, determining the spatial distribution of metabolites on liver tissue and the metabolic changes with disease progression is essential to decipher the heterogeneity of liver tissue and to gain insight into NAFLD. In **chapter 3**, our spatio-temporal analysis of mouse liver tissue and enrichment analysis for metabolic clusters revealed that metabolite clusters within sinusoids and parenchyma during the early stage of NAFLD involve mostly identical and a small

number of differential pathways. While our studies are basic in nature, findings may benefit the design of a new generation of precision medicines or life-style interventions that target the aberrant metabolism in NAFLD. Moreover, the identified disease-associated metabolites could serve as early reporters of the disease.

Substantial evidence has shown that male and female CVD patients differ in terms of underlying causes, presentation, and prognosis of the disease. Meanwhile, most studies on CVD risk factors still mainly focus on male patients [9]. In **chapter 4**, we observed distinct CVD-specific sex differences in monocyte transcriptional profiles and cytokine activities, and female patients are more activated in EGF, IFN1, CD40L, GM-CSF and VEGF signalling pathways than males. More relevant, we identified a common regulator of the disturbed signalling pathways that may serve as target for genetic linkage studies or for interventions specifically in women. Translating these findings into practice will contribute to reducing gender disparities in preventive care and improving clinical CVD treatments for women.

Hypertension is widely acknowledged as major risk factor for cardiovascular disease and blood pressure control has become one of the main means to prevent cardiovascular diseases [10, 11]. Nevertheless, the underlying mechanism behind the link between them is not clear. In **chapter 5**, we uncovered the negative correlation between blood pressure and LPS response of monocytes, indicating the reason why hypertension is a risk factor for CVD may be related to the suppressed immune response. This finding could improve treatment options for CVD disease associated with hypertension in the future. Furthermore, we inferred that iloprost, a drug capable of targeting this disease network, may not only enhance a compromised LPS response in monocytes in CVD-susceptible subjects, but may also represent an effective drug in the treatment of diastolic hypertension-related CVD.

## Conclusion

The computational approaches developed in this thesis greatly facilitate the work of biomedical researchers. Additionally, the findings based on these computational analyses deepen the understanding of cardiometabolic diseases and contribute to the development of more effective therapeutic options for an early conquest of this disease.

## References

1. Li Y, Ma L, Wu D, Chen G. Advances in bulk and single-cell multi-omics approaches for systems biology and precision medicine. *Briefings in Bioinformatics*. 2021;22:bbab024.
2. Tang X, Huang Y, Lei J, Luo H, Zhu X. The single-cell sequencing: new developments and medical applications. *Cell & Bioscience*. 2019;9:53.
3. Asp M, Bergenstråhle J, Lundeberg J. Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration. *BioEssays*. 2020;42:1900221.
4. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, et al. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell*. 2018;174:968–981.e15.
5. Buchberger AR, DeLaney K, Johnson J, Li L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal Chem*. 2018;90:240–65.
6. Nonalcoholic Fatty Liver Disease (NAFLD) & NASH | All Content | NIDDK. National Institute of Diabetes and Digestive and Kidney Diseases. <https://www.niddk.nih.gov/health-information/liver-disease/nafl-d-nash/all-content>. Accessed 21 Aug 2022.
7. EASL–EASD–EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *Journal of Hepatology*. 2016;64:1388–402.
8. Lu Q, Tian X, Wu H, Huang J, Li M, Mei Z, et al. Metabolic Changes of Hepatocytes in NAFLD. *Frontiers in Physiology*. 2021;12.
9. Connelly PJ, Azizi Z, Alipour P, Delles C, Pilote L, Raparelli V. The Importance of Gender to Understand Sex Differences in Cardiovascular Disease. *Canadian Journal of Cardiology*. 2021;37:699–710.
10. He J, Whelton PK. Elevated systolic blood pressure and risk of cardiovascular and renal disease: overview of evidence from observational epidemiologic studies and randomized controlled trials. *Am Heart J*. 1999;138 3 Pt 2:211–9.
11. Stamler J, Stamler R, Neaton JD. Blood pressure, systolic and diastolic, and cardiovascular risks. US population data. *Arch Intern Med*. 1993;153:598–615.

# **Appendices**

List of abbreviations



---

7AAD	7-Aminoactinomycin-D
AdaBoost	Adaptive Boosting
aREA	Analytic Rank-Based Enrichment Analysis
BP	Biological Process
BP	Blood Pressure
CAD	Coronary Artery Disease
CC	Cellular Component
CD40L	CD40 Ligand
CLAHE	Contrast-Limited Adaptive Histogram Equalization
Cmap	Connectivity Map
CMDs	Cardiometabolic Diseases
COX	Cyclooxygenases
COX7C	C Oxidase Subunit 7C
CRP	C-reactive protein
CTMM	Center for Translational Molecular Medicine
CVD	Cardiovascular Disease
DBP	Diastolic Blood Pressure
DDA	Data-Dependent Acquisition
DEG	Differential Expressed Gene
DIABLO	Data Integration Analysis for Biomarker Discovery Using Latent Components
DP	Dirichlet Process
DPGP	Dirichlet Process Gaussian Process Mixture Models
DPI	Data Processing Inequality
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor
ES	Enrichment Statistic
FA	Fatty Acid
FCGR2B	Fc-Gamma Receptor 2B
FDR	False Discovery Rate
FFPE	Formalin-Fixed Paraffin-Embedded
FMO	Fluorescence-Minus-One
FOV	Fluorescent Fields of View
GES	Gene Expression Signatures
GM-CSF	Granulocyte-Macrophage Colony-Stimulating Factor
GO	Gene Ontology
GP	Gaussian Process
GRN	Gene Regulatory Network



GSEA	Gene Set Enrichment Analysis
H&E	Haematoxylin and Eosin
HDL	High-Density Lipoproteins
HDST	High-Definition Spatial Transcriptomics
HFD	High-Fat Diet
HMGB	High Mobility Group Box
HSP	Heat Shock Protein
IF	Immunofluorescence
IFN1	Interferon Type I
IFN- $\gamma$	Interferon- $\gamma$
IHC	Immunohistochemistry
IHD	Ischemic Heart Disease
IL-13	Interleukin 13
IL-4	Interleukin 4
IL-4	Interleukin-4
IMC	Imaging Mass Cytometry
INS	Insulin
ITO	Indium-Tin Oxide
K-S	Kolmogorov-Smirnov
LDL	Low-Density Lipoproteins
LMNA	Lamin A
log <sub>2</sub> FC	log 2 fold change
LPBoost	Linear Programming Boosting
LPS	Lipopolysaccharides
m/z	mass-to-charge ratio
MALDI-MSI	Matrix-Assisted Laser Desorption Ionization-MSI
MF	Molecular Function
MF1	Mean Fluorescence Intensity
MI	Mutual Information
MMI	Mattes Mutual Information
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MSI	Mass Spectrometry Imaging
MST	Minimal Spanning Tree
NAFLD	Non-Alcoholic Fatty Liver Disease
NASH	Non-Alcoholic Steatohepatitis
NES	Normalized Enrichment Score

---

NGS	Next-Generation Sequencing
NSTEMI	Non-ST Elevation Myocardial Infarction
ORA	Overrepresentation Analysis
ORB	Oriented FAST and Rotated BRIEF
oxLDL	Oxidized Low-Density Lipoprotein
oxPHOS	Oxidative Phosphorylation
OxPLs	Oxidized Phospholipids
PCA	Principal Component Analysis
PGE2	Prostaglandin E2
PGH2	Prostaglandin H2
RAAS	Renin-Angiotensin-Aldosterone System
RNA-Seq	RNA Sequencing
ROIs	Regions of Interest
RSN	Robust Spline Normalization
RUSBoost	Random Undersampling Boosting
SASA	Spatially Aware Structurally Adaptive
SBP	Systolic Blood Pressure
scRNA-Seq	Single Cell RNA Sequencing
SMC	Smooth Muscle Cell
SNN	Shared Nearest Neighbour
SOM	Self-organized Map
SSC	Spatial Shrunken Centroids
STEMI	ST-Elevation Myocardial Infarction
TF	Transcription Factor
TG	Triglycerides
TLR4	Toll-Like Receptor 4
TOM	Topological Overlap Measure
tSNE	T-Distributed Stochastic Neighbour Embedding
UCP-2	Uncoupling Protein-2
Umap	Uniform Manifold Approximation and Projection
VEGF	Vascular Endothelial Growth Factor
WGCNA	Weighted Gene Correlation Network Analysis
WHO	World Health Organization
WTCS	Weighted Connectivity Score



# Appendices

Acknowledgements

致谢



Five years ago, when I was still in the distant land of China longing for a wonderful life in the Netherlands, I didn't realize how hard it would be to get a PhD degree. Today, with the enormous stress, frustration and massive gray hair, I am coming to the end of my career of a PhD student. However, to be honest, I have greatly enjoyed doing this PhD (not all the time), mainly due to the countless people around me who support and encourage me.

First of all, I would like to express my sincere gratitude to all the members of the dissertation committee, **Prof. I.C.W. Arts**, **Prof. J. Jankowski**, **Prof. T. Uras** and **Dr. C. Seiler** for your time in evaluating my dissertation and for your valuable comments.

Next, I would like to give my deepest appreciation to my supervision team.

**Prof. dr. Erik A.L. Biessen**, you are the most important person for my PhD study. I have always felt that the main reason why our group more like a family rather than just a research group is because of your attractive personality. Therefore, I am actually your big fan (Probably not obvious, but you know Chinese people are always shy). I made some mistakes during my research. However, you have never blamed me even once, but just help me deal with problems. In particular, in the final year of my PhD study, I can't imagine how I would have been able to keep going to finish my thesis without your support and encouragement. You have no idea how loudly I cried when I read your email saying, "there's light at the end of the tunnel, isn't there?". I was fortunate to come to Maastricht from China 5 years ago, and to be your PhD student. Thank you for always encouraging me to be myself and for always listening patiently to my ideas. A thousand words come together to say that you are the best tutor I have ever met. It is a pity that I don't have time to learn more from you, but I look forward to keeping in touch with you in the future!

**Dr. Joël M.H. Karel**, I can't tell you how lucky I feel to have you as my supervisor. Although we were not in the same office building, you still did your best to give me the support and guidance I needed to find my own niche in a team full of biologists. I still remember you coming to our lab and teaching me about "low-pass and high-pass filtering" and I thought you were amazing at that time, always making complex knowledge easy to understand. In addition, I was very impressed by your rigorous approach to academics. You were the one who kept telling me to dissect the distribution of each data set and the principles of the algorithm in depth, rather than just for application. I will always keep this in mind in my future research path.

**Dr. Pieter Goossens**, as your first PhD student, many thanks for your help and supports in both of my research and life. I am glad that you become my co-supervisor. During my time with you, I not only learned how to work with biologists, but also learned a lot of biological knowledge from you. No matter how simple a question I asked, you always explained to me patiently. In addition, I was also very

touched by the care you took in my life. I still remember you taking the drill to my house to install the chandelier. Besides, I preserved the red envelope you gave me on New Year's Eve. I wish you, a rising academic star in the field of biology, a bright future in the research of macrophages.

I would also like to highlight three truly exceptional people who have been extremely supportive and helpful to me in my PhD projects: **dr. Marjo Donners**, **dr. Evgeni Smirnov**, and **dr. Rachel Cavill**.

**Marjo**, first, thank you for your support and help with my thesis and paper. To be honest, I am pretty much enjoying the time together with you, though we have only been working with you for two years. You are very willing to understand my computational work and I enjoy listening to you talk about interesting biology. We always had a clear division of labor and smooth communication. In addition, I am grateful for your help and interest in my life. I will never forget the rainy morning of my car accident when you picked up helpless me from the police.

**Rachel**, I always thought that if only I had met you earlier, the life of my PhD would be much easier. Under your guidance, I saved much time in delving into articles and books to find solutions when my research was stuck. Especially during the time of lock down, our regular meetings and weekly discussions not only helped me in giving my project a boost, but also greatly eased my anxiety.

**Evgeni**, you are an excellent and reliable data scientist in my opinion. Thank you for your patience in answer my questions at each of our regular meetings. I don't know if you remember that we went to the university canteen for discussing problems of my project. Come to think of it, that was really a fun experience.

Next, I would also like to extend my thanks to the people with whom I collaborated in my PhD. First of all, I am grateful for the help provided by **M4I. Prof. dr. Ron M.A. Heeren**, and **dr. Benjamin Balluff**, thanks for providing me the MSI data and the teaching me the data processing procedures. **Eva** and **Britt**, thanks for your help with the validation of lipids and metabolites. **Jianhua (花花)**, 很幸运不仅能够和你成为好朋友, 并且还是项目上的合作伙伴。不得不说, 以中国人的合作方式, 工作效率确实要高出不少。祝贺你找到了爱情, 也希望我们两个有机会继续一起浪呀! In addition, I would like to thank the people I have known from the Department of Advanced Computing Sciences, Maastricht University. **Briassouli**, thanks for teaching me the image processing course, and I am glad that I passed the final exam with a high score! For the people from S&C meeting, although we didn't spend much time together (I was already in the last year of my Ph.D. when I started to participate in the meetings), I enjoyed the time with you, in the way of both online and onsite. Next, I would like to thank **dr. Javier Perales-Patón** from institute of computational biomedicine, Heidelberg University. Thank you for the endless

supports and kind help of the Gender Study. We become colleagues nowadays, so I look forward to more communications and collaborations in our future!

Additionally, I would like to thank the guys in our lab.

**Adele**, I feel lucky to have met and become friends with you in Maastricht. I will miss the time that we cooked, ran and play squash together, and hopefully will meet again in Heidelberg or black forest! **Valeria**, I especially adore that you are passionate about everything and everyone, and always have the courage to speak your mind. **Baixue** and **Zhenyu**, I am very happy that you two Chinese joined our group in my last year of PhD, to make our 'Pathology China Town' possible. I am also thankful that during this limited time, we went on a trip to Iceland, which was tortuous, but full of memories. All the best for your research and hope Zhenyu will come back next year so that we can start new adventure! **Olivia**, I'm so glad you're still in Maastricht, so we got to see each other a few times before I left. Hope you enjoy your current job and life! **Jan**, **Daniella**, and **Han**, thanks for your help and concern as senior PhD students to me when I just come to Maastricht. **Jan**, hope you are doing well in Germany! **Daniella**, you are my first Dutch friend here so thanks for everything you have done for me. I assume will become a mommy soon? Congratulations in advance on the new little addition to your family! **Han**, 作为实验室里和我有相似背景的师兄, 感谢你博士期间的陪伴。如今你虽然身在瑞典, 但我相信, 四年里的“相依为命”的情谊是真切且无需太多言语的。希望你和**仝乐**师姐在瑞典健康顺利! **Laura**, I am glad to see that you are interested in Chinese food, but it is a pity that we did not find an opportunity to cook together again before I moved. You are very welcome to travel to China, to sample the myriad of Chinese cuisine! **Elias**, you are such a funny and nice guy who can keep our group energized even during the epidemic! Unfortunately, I couldn't join the trip to Berlin with you, but maybe you can consider the destination of your next trip as 'Heidelberg' so that we will see each other again! **Renee** and **Sebastian**, you are always hard working so do not forget to enjoy your life and have a good coffee downstairs. Otherwise, the headache and backache will come to you (don't ask me how I know it). **Judith**, although I didn't work with you much during my PhD, I still can feel your rigorous attitude towards academics. On the other hand, you are a lively and interesting person in life. Thank you for inviting me to your pool party, and I look forward to working with you in the future! **Lieve**, I especially enjoyed talking to you because you always came across as kind and approachable. You were the first postdoc I came in contact with after I arrived in Maastricht and I am glad you were still in our group when I graduated. **Margaux**, **Kim**, **Ruud**, **Marion**, **Anke**, **Jasper**, **Jenny**, **Axel**, **Suzan**, **Gregorio**, **Mat**, and **Dlzar**, thank you all for your help and supports in my PhD. I will miss all the members in this big EVP family and wish you all the best in future! !



轮到感谢一起玩耍的小伙伴们了。博士的最后一年里我真的是抑郁又脆弱，生活感情学业一团糟，活着全凭一口气。然而，全靠家人和小伙伴们的支持，才让你们看到仍然健在的我。

**石帅**，不得不说你是我博士最后一年的光。虽然被你治愈的过程蜜汁诡异。当我将百分百经历投入到毕业和科研上，但却毫无进展时，那种绝望感几乎要将我吞没。但因为你的日常捣乱骚扰，我必须分出一半的精力应付你，科研忽然就变得无关紧要了:)。**刘丹**，我真的很佩服你好像有用不完的精力，再大的困难也能微笑面对，我要向你学习，也希望你和唐哥还有你家宝宝在这里茁壮成长！**Pmax**, I'm so glad that we were able to meet again in Maastricht a year after your graduation, and it seemed that nothing had changed. I really miss our travel days in Portugal and Spain, and the time we played Mahjong together ☺. Hope we still have chance to travel together! **陈琳**，还记得我第一次去你家吃火锅吗，哈哈，好久没联系了，你和你家猫在国内都要开开心心的呀。**丁伶俐**和**张礼川**，真的感觉真的好久不见了，真怀念你们还在这里的日子，我们几个一起出去玩，或者在家里聚餐，真的是毕业前为数不多的开心的回忆了，希望你们和宝宝在国内一切都好。**宗盛华**师兄，虽然没太多机会见面，但是我一直记得我生病的时候你对我的帮助，但愿你家小玉米还记得我 ☺。**汤梦资**，很庆幸和你一起来了欧洲，更庆幸在你回国之前跟你去了克罗地亚，作为相识八年的同学加好友，祝愿你武汉事业爱情双顺利。还有**惠冻潼**，**罗倩**，**茆新伟**，**张曼莉**，**王雅文**，**余艺文**……这些同届的小伙伴们，以及**张硕**，**石镇伟**，**冯正**……这些师兄师姐们，虽然大家已经各奔东西，但马城依然有我们共同的回忆，我们一起跨年，参会，过圣诞，包饺子，狼人杀，甚至还有编程比赛……年纪大了脑子也愈发迟钝，所以，这里还要感谢所有我没提及但熟悉的人，谢谢你们在我博士生涯里留下的印记。

最重要的，也是最要感谢的，是无条件支持我的父母：**路书**先生和**马成龙**女士 (The most importantly, I would like to thank my parents, Mr. Shu Lu and Mrs. Chenglong Ma)。我不止一次自我感慨我要多幸运才能遇到你们这么开明的父母，从我去外地上大学，到有留学梦，到千里迢迢奔赴荷兰，再到如今的在德国继续科研之路，纵使有万般的不舍和思念，你们还是全力支持我朝向更好的未来，支持我做我想做的事情。很多时候你们可能并不懂我在干什么，却还是愿意听我抱怨科研的不顺利，并尽自己所能来帮我解决问题和缓解压力。我向来不愿意麻烦别人，也不愿意展示自己脆弱的一面，所以一个人的毕业季有多难熬只有你们知道，有几次感觉要撑不下去崩溃的时候也只有你们知道。我总是想，我能完成并提交这本论文，一半的功劳都要归功于你们，归功于我坚信，纵使什么也做不好，你们也依然会坚定地在我身后支持我。

博士生涯说长很长，说短似乎一瞬间就过去了。呆得时间久了，竟有种“只闻新人笑，不见旧人哭”的悲戚感。离开马城之际，我一边感慨自己对马城太过熟悉，熟悉到凌晨三四点也可以在学校和家之间横行；另一方面，又忽

然间对这里无比怀念，以至于回家路上的一棵树，一个斜坡都能让我没来由鼻子一酸。但是，生活也许就是这样，无论你在哪里在干什么，准没准备好，它都会推着你走。

所以最最后，感谢无论被生活推到何处都仍然没有放弃希望的我自己。

(At last, thank myself for not giving up no matter where I was pushed by life)

山高路远，后会有期。

(There is still a long way to go; we will see each other in the future)



# **Appendices**

Curriculum Vitae



## About the Author

Lu Chang was born on November 13, 1992, in Pingdingshan, Henan Province, China, an industrial city known for its coal mining industry. She has longed for the outside since she was a child, so after graduating from high school, she went to Jiangxi Normal University, 700 kilometers away from her hometown to study computer science and technology. During four years of undergraduate study, she participated in several programming activities, such as ACM International Collegiate Programming Contest, which developed her a strong interest in programming. The summer after her sophomore year, she went to Beijing University to study biometrics identification, which led to her interest in the application of computational methods to biological data and her decision to pursue a master's degree. In September 2014, she started her life as a master's student in the Machine Learning and Data Analytics Lab at Southwestern University, where she gained knowledge of machine learning and data mining and applied them to the prediction of protein function and noise functional annotations. In October 2017, she crossed Eurasia to Maastricht, the Netherlands, to pursue her Ph.D. with her passion for research, where she focused on applying machine learning and statistic approaches to multi-omics and multimodal image data to analyze the mechanism of cardiometabolic diseases. Starting in October 2022, Chang Lu continued her research path as a postdoctoral fellow at the Institute for Computational Biomedicine (Saez-Rodriguez Group) at the University of Heidelberg, Germany, studying computational integration of cellular circuits and immune cell repertoires in spatial and temporal resolution.

## Personal Information

---

**Name:** Chang Lu  
**Date of birth:** 13th of November, 1992  
**Place of birth:** Pingdingshan, Henan, China  
**Nationality:** Chinese  
**E-mail:** [lu755735129@163.com](mailto:lu755735129@163.com)

## Research Career

---

**10/2022– Present** Postdoctoral researcher, Institute for Computational Biomedicine, Heidelberg University, Germany

**10/2017 – 09/2022** Research assistant, Department of Pathology, Maastricht University Medical Center+ (MUMC+), Maastricht, The Netherlands

**09/2014 – 06/2017** Research assistant, Department of Computer and Information Science, Southwest University, Chongqing, China

## Education

---

**10/2017 – 09/2022** **PhD student**  
Computational Biology in cardiometabolic diseases, Department of Pathology, Maastricht University Medical Center+ (MUMC+), Maastricht, The Netherlands  
Promoters: Prof. dr. Erik A.L. Biessen  
Dr. Joël M.H. Karel  
Co-promoter: dr. Pieter Goossens

**09/2014 – 06/2017** **Master of Engineering**  
Computer Application Technology  
Department of Computer and Information Science  
Southwest University  
Chongqing, China  
Supervisor: Prof. dr. Guoxian Yu

**09/2010 – 06/2014** **Bachelor of Science**  
Computer Science and Technology  
Department of Computer Information Engineering

Jiangxi Normal University  
Nanchang, China

## Publications

### **Integrating multiplex immunofluorescent and mass spectrometry imaging to map tissue myeloid heterogeneity in its metabolic and cellular context**

P Goossen<sup>†</sup>, **C Lu**<sup>†</sup>, J Cao, M Gijbels, J M.H. Karel, E Wijnands, G E. Fazzi, K Wouters, E Smirnov, M J.M. van Zandvoort, B Balluff, M M.P.C. Donners, R M.A. Heeren, E A.L. Biessen

*Cell Metabolism*, 2022 Aug 2;34(8):1214-1225.e6.

### **Identification of CD8+ T cell PRDM1 in high-risk human plaques and its regulatory role in murine lesion development**

H. Jin, S.L. Maas, **C. Lu**, J. Nagenborg, M. Manca, J.M.H. Karel, R. Cavill, O. Waring, C.J.J.M. Sikkink, B.M.E. Mees, M.J.A.P. Daemen, E. Smirnov, J. Sluimer, E. Van Der Vorst, E.A.L. Biessen

*Cardiovascular Research*, 2022, 118, *Supplement\_1*, cvac066.186

### **Identification of a PRDM1-regulated T cell network to regulate T cell driving plaque inflammation in human and mouse atherosclerosis**

S L Maas, H Jin, **C Lu**, J Nagenborg, J M H Karel, R Cavill, C J J M Sikkink, S Nadeau, M Gijbels, B M E Mees, E Smirnov, J C Sluimer, G A Martins, E P C Van Der Vorst, E A L Biessen

*European Heart Journal*, 2022, 43, *Supplement\_2*, ehac544.3049

### **MicroRNA-26b Attenuates Platelet Adhesion and Aggregation in Mice**

Peters, L.J.F.; Baaten, C.C.F.M.J.; Maas, S.L.; **Lu, C.**; Nagy, M.; Jooss, N.J.; Bidzhekov, K.; Santovito, D.; Moreno-Andrés, D.; Jankowski, J.; Biessen, E.A.L.; Döring, Y.; Heemskerk, J.W.M.; Weber, C.; Kuijpers, M.J.E.; van der Vorst, E.P.C.

*Biomedicines* 2022, 10: 983.

### **NoGOA: predicting noisy GO annotations using evidence and sparse representation**

G. Yu, **C. Lu**, J. Wang

*BMC Bioinformatics* 2017, 18:350

### **BRWLDA: bi-random walks for predicting lncRNA-disease associations**

G. Yu., G. Fu., **C. Lu.**, Y. Ren, J. Wang

*Oncotarget*. 2017; 8: 60429-60446.



**HashGO: Hashing Gene Ontology for protein function prediction**

G. Yu, Y. Zhao, C. Lu, J. Wang

*Computational Biology and Chemistry*, 2017, 71: 264-273

**Semi-supervised classification based on clustering adjusted similarity**

Xia Chen, Chang Lu, Qiaoyu Tan, Guoxian Yu

*International Journal of Computers and Applications*, 2017, 39(4): 210--219.

**NoisyGOA: Noisy GO annotations prediction using taxonomic and semantic similarity**

Chang Lu, Jun Wang, Zili Zhang, Pengyi Yang, Guoxian Yu.

*Computational Biology Chemistry*, 2016, 65, 203--211.

**Identifying noisy functional annotations of proteins using sparse semantic similarity**

Chang Lu, Xia Chen, Jun Wang, Guoxian Yu, Zhiwen Yu.

*SCIENTIA SINICA Information*, 2018, 48(2): 1.

A

**Cardiovascular disease but not health is hallmarked by sex-specific cytokine signaling pathways in circulating monocytes**

Chang Lu<sup>†</sup>, Marjo M.P.C. Donners<sup>†</sup>, Joël M.H. Karel , Hetty de Boer, Anton Jan van Zonneveld, Hester den Ruijter, J. Wouter Jukema, Adriaan Kraaijeveld, Johan Kuiper, Gerard Pasterkamp, Rachel Cavill, Javier Perales-Patón, Ele Ferrannini, Pieter Goossens, Erik A.L. Biessen

Submitted to *Clinical and Translational Medicine*

**Identification of a gene network driving the hypertension associated dampened monocyte LPS response in coronary artery patients**

Chang Lu<sup>†</sup>, Marjo M.P.C. Donners<sup>†</sup>, Julius B. J. de Baaij, Han Jin, Jeroen Otten, Marco Manca , Barend Mees, Mat J.P. Daemen, Judith C. Sluimera, Rachel Cavillh, Joël M.H. Karel, Pieter Goossens, Erik A.L. Biessen

Submitted to *Hypertension*

**Spatio-temporal Metabolomics Reveals Metabolic Changes in the Early Stage of Non-alcoholic Fatty Liver Disease in Mice**

Chang Lu, Jianhua Cao, Kristiaan Wouters, Marion Gijbels, Marjo M.P.C. Donners, Benjamin Balluff, Evgueni Smirnov, Rachel Cavill, Joël M.H. Karel, Ron M.A. Heeren, Erik A.L. Biessen<sup>†</sup>, Pieter Goossens<sup>†</sup>

*In preparation*

<sup>†</sup> Authors contributed equally

---

## Conference Presentations and Academic Activities

---

- July 2021**      **The 29<sup>th</sup> conference on Intelligent Systems for Molecular Biology and the 20<sup>th</sup> European Conference on Computational Biology (ISMB/ECCB 2021), Online**  
 Selected Poster Presentation  
*Integrative gene expression analysis reveals distinct sex-specific regulatory pathways in circulating monocytes associated with cardiovascular disease*
- November 2020**      **byteMAL conference 2020, Online**  
 Selected Poster Presentation (Best Poster Award)  
*Integrating Multiplex Immunofluorescent and Mass Spectrometry Imaging to Map Tissue Myeloid Heterogeneity in Its Metabolic and Cellular Context*
- July 2019**      **Deep learning Summer School, Warsaw, Poland**
- May 2019**      **European Atherosclerosis Society (EAS) Meeting 2019, Maastricht, the Netherlands**  
 Selected e-poster presentation  
*Towards The Complete Picture: Computational Strategies to Identify Ms-Imaging Derived Molecular Fingerprints Associated with Inflammatory Cell Phenotypes*
- March 2018**      **Computational Pathology Workshop 2018, Athens, Greece**  
 Oral presentation  
*Towards The Complete Picture: Computational Strategies to Identify Ms-Imaging Derived Molecular Fingerprints Associated with Inflammatory Cell Phenotypes*
- July 2012**      **Computer Science Summer School, Peking University, China**

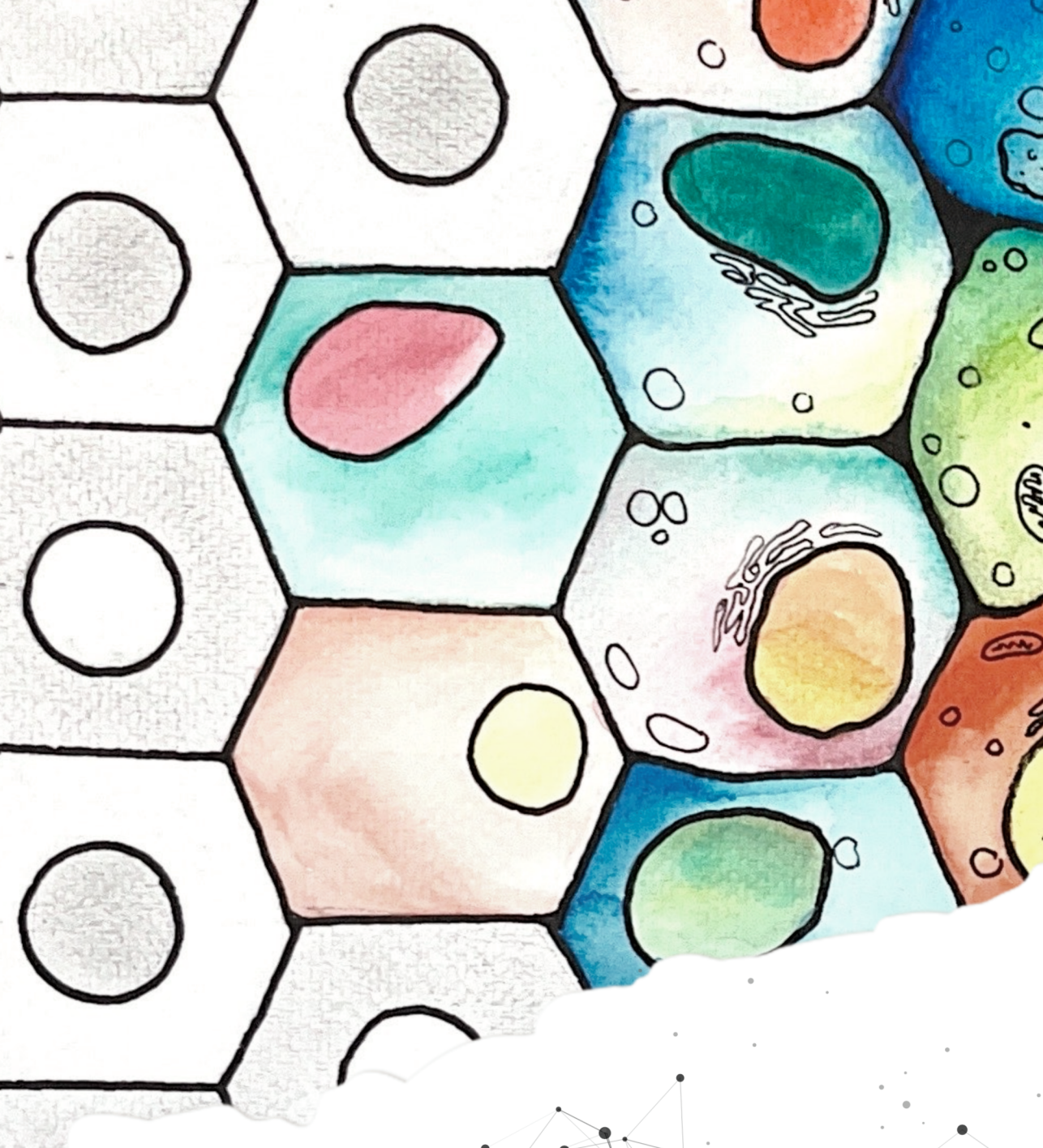
## Key Honours and Awards

---

- February 2022**      NPO Research Fund grant, Netherlands
- January 2018**      Excellent MA theses of Chongqing, China
- July 2017**      Excellent Graduate of Southwest University, China

- May 2017**                      Research fellowship from China Scholarship Council (CSC), China
- October 2016**                National Graduate Fellowship, Ministry of Education and Ministry of Finance of PRC, China
- April 2016**                    Fundamental Research Funds for the Central Universities (XDJK2016E076), China





01010100011010000•1010010111001100100000010101001  
100110010000001110011011011101101101  
001110010011000010110111001100100010  
0000011010001001010101110000111010001000000  
1101101110101010110110  
110001000000110100011001010111001  
000110011001101110111  
1101000111010001101001011011

