NORTHWESTERN UNIVSERSITY

Characterizing Macrophage Heterogeneity in Tissues Through High-Throughput Transcriptomics Technologies and Algorithms

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Driskill Graduate Program in Life Sciences, Biomedical Informatics Track

By Shang-Yang Chen

EVANSTON, ILLINOIS

JUNE 2022

ABSTRACT

Macrophages are innate immune cells that are traditionally thought to be specialists in phagocytosis. More recent evidence suggest that macrophages reside in nearly every organ and readily adapt to local microenvironmental signals, leading to highly plastic phenotypes across and within tissues. Therefore, rather than treating it as a homogenous cell population, new studies should consider the functional heterogeneity that may exist among macrophages. A better understanding of macrophage heterogeneity can inform on potential therapeutic strategies, as macrophages play central pathological roles in numerous diseases. The emergence of high-throughput RNA profiling assays offers great potential in characterizing macrophage heterogeneity by facilitating direct comparison of gene expression profiles between different subpopulations. The introduction of single cell technology further enables the identification of new macrophage subpopulations. Given the "big data" nature of high-throughput assays, the development of novel computational algorithms and their rigorous application is crucial to gain meaningful insights from the experiments. In this thesis, we demonstrated how integrative computational analysis of bulk and single cell RNA-seq data can improve our understanding of macrophage heterogeneity across organisms and tissues. We characterized macrophage subpopulations residing in murine synovium and human pediatric livers, while assessing changes to their phenotypes under pathological conditions. We further extended the concept of cellular heterogeneity beyond macrophages, where we uncovered the existence of murine synovial monocyte populations distinct from circulation. In the last part of the thesis, we identified a major weakness in the current analytical workflow for transcriptional data and developed a web application called MAGNET, which aims to improve functional enrichment analysis for macrophage-related genomic data.

ACKNOWLEDGEMENTS

Five years ago, I boarded a one-way flight bound for Chicago, an incredible city I have never thought I would set foot on, and embarked on my journey at Northwestern University for the PhD degree in biomedical informatics. It was an honor to be a part of the exciting research happening in this novel and ever-changing field. This incredible journey would not have been possible without the support from numerous wonderful people, which I dedicate this section of the thesis to.

First and foremost, I would like to express my highest gratitude to my advisor and mentor, Dr. Deborah Winter, who encouraged me to become a better computational biologist. Without your continuous support and guidance, it would have been impossible to overcome the numerous challenges encountered in my research and finish this thesis. I especially want to thank you for your mentorship in developing important skills including public speaking, independent critical thinking, and tackling a problem from multiple perspectives. I would also like to extend my appreciation towards the members from my thesis committee, Dr. Edward Thorp, Dr. Benjamin Singer, and Dr. Jiping Wang for their generous availability and valuable advice. I want to further thank Dr. Justin Starren and Dr. Cara Gottardi for managing and supporting the amazing biomedical informatics track within Driskill Graduate Program.

I am deeply grateful for the wonderful colleagues and collaborators I had the pleasure of working with. To the members of Winter Lab, especially Kishore, Anna, Gaurav, Monica, Blair, Yidan, Sam (not me), and Lucia, thank you all so much for being the friendliest, kindest, and supportive group of people I have ever been with. I will cherish our memories together in the lab dearly. To my colleagues in the former bioinformatics office, Ziyou, Marjorie, and Sam (also not

me), I will never forget the light-hearted atmosphere and our sense of comradeship. It was truly a pleasure to share the same office with you guys. I also want to thank my dear friends from our cohort Chang, Guangyuan, and Dylan for being awesome food and travel buddies. Thank you for accompanying me throughout the long journey, cheering me up when I was feeling down, and making our lives interesting through thought-stimulating conversations.

Last but not least, I cannot express enough love and gratitude towards my mother Annie, my father Carl, and my younger sister Judy. Thank you for being my greatest supporters and the best life mentors. I could not have completed this journey without your unconditional love and support. It is truly a blessing to be born your son and brother. Thank you all so much!

PREFACE

Chapter 1 consists of an overall survey on the current understanding of macrophage heterogeneity, and how bulk and single cell RNA sequencing assays can be applied to further our knowledge of the matter. A review of the current technology and computational analysis strategies for these assays is also included.

Chapter 2 is adapted from original research articles entitled "Synovial Macrophage Heterogeneity Confers Differential Response to Acute and Chronic Inflammatory Arthritis." (Chen & Montgomery et al., in preparation) and "Critical Role of Synovial Tissue-resident Macrophage Niche in Joint Homeostasis and Suppression of Chronic Inflammation (Huang et al. 2021, Science Advances, 3rd author). This chapter describes the elucidation of macrophage heterogeneity in synovial tissue in the context of arthritis by performing integrative analysis of bulk and scRNA-seq data.

Chapter 3 contains some results from an original research article entitled "A novel tissue-resident non-classical monocyte population forms a critical barrier to inflammation in the synovium" (Montgomery et al., in revision, 2nd author). This chapter extends the concept of synovial myeloid heterogeneity through the discovery and characterization of extra-vascular resident monocytes. The applications of bulk and scRNA-seq data analysis is again emphasized.

Chapter 4 is derived from an original research article entitled "Transcriptional Profiling of Pediatric Cholestatic Livers Identifies Three Distinct Macrophage Populations" (Taylor et al. 2021, PLoS One, 2nd author) and is reproduced here with the permission of the copyright holder. This

chapter discusses how scRNA-seq analysis can facilitate the identification of distinct macrophage subsets across human patients with pediatric cholestasis.

Chapter 5 is adapted from an original research article entitled "MAGNET: A Web-based Application for Gene Set Enrichment Analysis Using Macrophage Data Sets" (Chen et al., in revisions). This chapter introduces a novel, interactive web application for performing enrichment analyses on custom gene sets that are specifically relevant to macrophages.

Chapter 6 contains a brief summary of the previous chapters, concluding remarks, and future perspectives on high-throughput sequencing assays and macrophage heterogeneity.

LIST OF ABBREVIATIONS

DNA - deoxyribonucleic acid

RNA - ribonucleic acid

cDNA – complementary DNA

TF – transcription factor

NGS – next generation sequencing

FACS – flow-activated cell sorting

RNA-seq – RNA-sequencing

scRNA-seq – single cell RNA-sequencing

BCL – binary base call

DEG – differentially expressed gene

UMI – unique molecular identifier

PCA – principal component analysis

tSNE – t-distributed stochastic neighbor embedding

UMAP – uniform manifold approximation and projection

CCA – canonical correlation analysis

GO – gene ontology

GSEA – gene set enrichment analysis

RA – rheumatoid arthritis

OA – osteoarthritis

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Identity and plasticity of tissue resident macrophages

1.1.1 Functional diversity of macrophages

First described in the late 19th century by Elie Metchnikoff [1, 2], macrophages are a group of immune cells in mammals that specializes in phagocytosis, a process where they physically engulf and digest foreign particles such as bacteria. For decades, macrophages along with other members of the mononuclear phagocyte system (MPS) [3] have been regarded as first line defenders of the innate immune system. Common functions attributed to macrophages includes scavenging of foreign pathogens through phagocytosis [4, 5], activation of adapt immune responses via interaction with dendritic cells (DCs) and lymphocytes [6], and regulation of inflammation by secretion of various cytokines and chemokines [7].

Findings from recent studies have led to major paradigm shifts in our understanding of macrophage biology. It is now clear that macrophages reside in virtually every tissue and organ under homeostatic conditions, and they are capable of tissue-specific phenotypes and functions well-beyond innate immunity [8, 9]. For example, microglia, the resident macrophages of the brain, participate in synapse pruning during development [10, 11]. Kupffer cells of the liver are involved in breaking down erythrocytes and recycling of heme [12, 13]. Alveolar macrophages of the lung play major roles in removing foreign particles from the respiratory surfaces [14, 15]. A non-exhaustive list of known tissue-specific functions of resident macrophages are provided in **Table 1**. In addition to inter-tissue diversity, functional specialization of macrophages can exist within individual tissues as well. Macrophages in the lung, for example, have been classified into three major subpopulations – alveolar macrophages and two subsets of interstitial macrophages. They

occupy different anatomical compartments within lung, possess distinct morphologies, and participate in varied biological processes ranging from alveoli homeostasis to immune surveillance [15, 16]. Intra-tissue functional diversity has also been documented for macrophages in the liver [17], intestines [18, 19], brain [20, 21], and other organs. These studies, along with others, highlight the functional diversity of macrophages.

A complex network of biological factors underlies and contribute to the observed functional diversity of macrophages, which can be roughly categorized into three major themes: **ontogeny**, **local environment**, and **inflammation**. Each of them will be described in more details in the subsequent sections.

1.1.2 Ontogeny

It was the prevailing view that macrophages in tissues are replenished and replaced exclusively by circulating monocytes derived from bone marrow progenitors [3]. However, this classical model of hematopoietic differentiation failed to explain the presence of long-living tissue resident macrophages that are able to self-maintain through local proliferation [22-24]. This raised the possibility that some macrophages may owing their origins to embryonic precursors that predates bone marrow hematopoiesis, which have been experimentally confirmed thanks to the development of novel fate-mapping techniques [25-27].

Current consensus in the field indicates that there are at least three distinct embryonic origins for macrophages, arising in successive waves. The first wave, called primitive hematopoiesis, is derived from the yolk sac mesoderm at around embryonic day (E) 7.0 [25, 28]. This is followed

by the appearance of erythro-myeloid precursors (EMPs) from the hemogenic endothelium of yolk sac between E8.0-8.5, termed transient definitive hematopoiesis due to their inability to persist in immune-compromised animals [29]. EMPs migrate into the fetal liver after the initial establishment of blood circulation and give rise to primitive monocytes [30]. Definitive hematopoiesis, which represents the third wave, is characterized by the colonization of immature hematopoietic stem cells (HSCs) in fetal liver at E10.5 [31]. With the establishment of fetal HSCs, fetal liver become the primary hematopoietic organ in the embryo from E12.5 onwards. These precursor HSCs will also enter the fetal bone marrow and eventually develop into adult HSCs. Thus, the developmental origins of tissue resident macrophages can be traced to three sources: yolk sac, fetal liver, and bone marrow [29, 32, 33].

It is now widely accepted that macrophage populations are substantially derived from embryonic precursors, but with considerable variation across tissues. Microglia is the only known population to be exclusively derived from the yolk sac, with minimal changes over lifetime [25, 34]. Macrophages residing in lung [35] and liver [36], on the other hand, originated primarily from fetal liver monocytes, while also harboring smaller populations of bone marrow-derived populations. Cardiac macrophages are also primarily derived from fetal liver, but the proportion of bone marrow-derived cells increases with age [37]. Intestinal macrophages were long thought to be composed of exclusively bone marrow derived macrophages and is continuously replenished from circulation [38]. This view has however been recently disproved with the identification of long-living macrophages with embryonic origins, characterized by their surface expression of TIM-4 and CD4 [19]. Therefore, with the exception of microglia, the majority of tissues harbor multiple subsets of macrophage with distinct ontogenic histories (**Table 1**).

Table 1 Functional diversity and plasticity of tissue resident macrophages.

Tissue	Ontogeny	Environmental signals	Master TF(s)	Specialized Functions	Citations
Brain	Yolk sac	M-CSF	SALL1, SMAD2/3	Synapse pruning	[11]
Kupffer cells (liver)	Fetal liver	DLL4, NOTCH	LXRα, ID3	Erythrocyte metabolism	[13]
Alveolar space (lung)	Fetal liver	GM-CSF	PPARγ	Surfactant clearing	[15]
Peritoneal space	Yolk sac + BM	Retinoic acid	GATA6	Interaction with B1 lymphocytes	[39, 40]
Red pulp (spleen)	Fetal liver	Heme	SPIC	Blood homeostasis	[41]
Intestines	Fetal liver + BM	TGF-β, NOTCH	RUNX3	Interaction with microbiome	[18, 42]
Langerhans cells (epidermis)	Yolk sac + Fetal liver	TGF-β, IL34	ID2	Antigen uptake and transportation to lymph nodes	[27, 43]
Kidney	Fetal liver + BM	?	AHR, NFATC (?)	Renal homeostasis	[44]
Heart	Fetal liver + BM	?	?	Electrical conduction	[45]

1.1.3 Local environment

Experimental evidence indicated that ontogenic heterogeneity alone is unlikely to account for the functional plasticity observed in macrophages across tissues. van de Laar et al. [46] showed that upon transplantation into newborn Csf2rb-/- mice, which lacks alveolar macrophages, wild type embryonic progenitors regardless of origins, including yolk sac, fetal liver, and bone marrow,

are all able to develop into functioning alveolar macrophages with no discernable differences in transcriptional profiles. Lavin et al. [47] further demonstrated that adult peritoneal macrophages are able to acquire up to 70% of gene expression patterns of alveolar macrophages when adoptively transferred into lung. [47]. Collectively, these two studies implicate local tissue environment as another major factor that influence the phenotypes and functions of macrophages.

The development and establishment of macrophage identity is controlled by a number of master transcription factors (TFs), including PU.1 [48, 49] and MAFB [50]. Emerging evidence suggest environmental signals encountered in tissue by macrophages are also capable of inducing additional TFs that contribute to the distinct gene expression profiles observed across tissues [47, 51]. For example, retinoic acid is known to trigger the transcription of GATA6, which is an essential regulator in the development of peritoneal macrophages [40]. Heme induces the expression of SPI-C, which mediates the development of red pulp macrophages [52]. The induction of PPARG-γ by GM-CSF is required for establishment and maintenance of alveolar macrophage identity [53]. Other examples of tissue-specific environmental signals and induced regulatory TFs is provided in **Table 1**. The mechanism behind environmental regulation likely goes beyond transcriptional level. Macrophages exhibit distinct enhancer profiles across tissues, exemplified by the divergent patterns of histone modification (e.g., H3K4me1 & H3K27ac) and open chromatin landscapes. These differential enhancer sites are disproportionately enriched for bind motifs of the tissue-specific TFs, insinuating targeted remodeling of the epigenomic landscapes triggered by environmental signals [47, 54]. In summary, tissue-specific environmental stimuli play a major role in shaping the identity of resident macrophages through regulation on both transcriptional and epigenetic levels.

The influence of local environment is emphasized and summarized in the ***niche model*** proposed by Guilliams et al. [55]. By definition, a niche represents a physical foundation or scaffold for macrophage residence that directly regulate the sizes and identities of tissue-specific populations. The niche acts as a major source of environmental signals that confer their specialized functions and their ability to self-maintain through inducing transcription of key TFs. The macrophages would in turn also provide benefits to their niche [56]. The dependence of microglia to their surrounding environment is an example of the niche model in action. Neurons and glial cells secrete IL34 [57] and CSF1 [58], respectively. They both are non-alternative ligands for the CSF1R receptor, which is an essential factor for the survival of microglia [59]. Microglia in turn contribute to the development and homeostasis of neurons and glial cells [58, 60]. The concept of macrophage niche is not limited to the whole tissue, as evidence of distinct niches on the sub-tissue level have been extensively documented. For example, liver consists of multiple anatomical zones, which act as distinct niches for macrophage development by promoting their interaction with different cell types and environmental cues. Within the perisinusoidal space, hepatocytes, stellate cells, and endothelial cells collaboratively scaffold the Kupffer cells and imprint their identities through synergistic NOTCH-BMP signaling, which induces the lineage-specific TF LXR-α [61, 62]. On the other hand, the liver capsule harbors a distinct population of capsular macrophages. Their identities are conferred through direct interfacing with the hepatic parenchyma and peritoneal cavity, with specialized dendrites extending into the sinusoids [63]. Spleen is another organ that contains multiple niches, where the white pulp, marginal zone, and red pulp each give rise to distinct macrophage subpopulations owing to their very different local environments [64]. In conclusion, the local environment defined and sustained by niches within tissues is a primary

contributor to the functional plasticity of tissue resident macrophages. It is however important to note that niches are not static and can be rapidly altered when homeostasis is disrupted, such as inflammatory conditions.

1.1.4 Inflammation

Inflammation is the protective immune response that occurs when the body encounters external challenges, such as bacterial infection and tissue injury [65]. Macrophages play central roles in promoting, maintaining, and resolving inflammation [7]. The resulting changes to local tissue environment, however, can render the niche unsuitable for pre-existing resident macrophages, which often leads to their necroptosis [66, 67]. The loss of tissue resident macrophages results in the opening of niches, which are then filled up either through local proliferation of surviving resident macrophages or the recruitment of circulating bone marrow monocytes in a CCR2-dependent manner [68, 69]. With an altered local environment, the newly differentiated and surviving macrophages adopt distinct identities and functional programs compared to those under homeostatic conditions. There are conflicting reports on the ultimate fate of recruited macrophages, however. A study in liver suggest that cells recruited from circulation eventually become transcriptionally indistinguishable from embryonically derived Kupffer cells after the subsidence of inflammation [69], while another study of peritoneal inflammation indicated that recruited macrophages adopt a distinct immature transitory state that persisted long-term compared to resident cells [70]. It is possible that the degree of restoration to steady state from inflamed tissue environment is the deciding factor of recruited macrophage phenotype. The evidence of trained immunity, where monocyte-derived macrophages are seemingly able to “memorize” the initial insult and elicit stronger immune responses for future challenges [71, 72], further suggest that

changes to the local tissue environment due to inflammation have the potential of reprogramming the identities of macrophages.

Although their exact functional roles vary across tissues and disease models, macrophages exhibit a common pattern of phenotypic shift over the course of inflammation. At the onset of infection/injury, classical Ly6C+ (CD14+CD16- in humans) monocytes [73] are predominately recruited into tissue. Macrophages differentiated from this initial wave specialize in the secretion of various pro-inflammatory cytokines such as IL-1B and TNF that propagate the magnitude of inflammation and facilitate the recruitment of other immune cells [74]. They also actively participate in the removal of potential pathogens and tissue debris from dying cells through phagocytosis [75]. After the initial wave, monocyte infiltration gradually become of primarily non-classical Ly6C- (CD14+CD16+ in humans) phenotype [73]. This coincides with the shift from the effector phase of inflammation to resolution [76, 77]. Macrophage derived from non-classical monocytes promote the resolution process through secretion of anti-inflammatory cytokines including IL6 and IL10 [78]. They also directly partake in the remodeling and regeneration of tissue by promoting angiogenesis and fibrosis [79]. The pro-reparative functions of macrophages are not always beneficial, as it is also possible that the inflammation process be never fully resolved, at which point it become chronic rather than acute. The constant remodeling of extracellular matrix mediated by pro-reparative macrophages often results in permanent tissue scarring and impairment of function, as observed in pulmonary fibrosis [80] and cardiovascular diseases [81]. In summary, macrophages are critical to both the propagation and resolution of inflammation.

The term ***polarization*** is often used to describe the phenotypic changes to macrophages

encountering pathological conditions. Traditionally, macrophage polarization is classified into two major types: M1 (classical activation) and M2 (alternative activation). M1 polarization can be induced through stimulation by LPS, and is characterized as being pro-inflammatory. M1 macrophages possess enhanced ability to eliminate pathogens through secretion of chemicals like reactive oxygen species (ROS) and nitrogen radicals. In contrast, M2 polarization can be activated via IL-4 and is characterized as being anti-inflammatory. M2 macrophages are potent producers of growth factors such as TGF-β and VEGF, which promotes wound repair and tissue-remodeling processes [82, 83]. Although easy to understand, the concept of M1/M2 polarization is largely being viewed as being too simplistic. A study that profiled and analyzed macrophage transcriptomes after administration of 28 different stimuli revealed that there are at least 9 major dimensions of polarization [84]. Furthermore, our previous understanding of macrophage polarization is based solely on *in vitro* experiments, and there is no evidence that macrophages with the exact phenotypes of M1 or M2 polarization exist *in vivo* [85]. Examples of in vivo polarization responses that are specific to the tissue type and pathological condition is just starting to be examined. In the case of skeletal muscle injury, both recruited and resident macrophages aid in the regeneration of muscle fiber by stimulating the proliferation and differentiation of myogenic progenitor cells (MPCs). During the process, macrophages exhibit a wide range of transcriptional profiles spanning from pro-inflammatory to pro-reparative depending on the timepoint of injury progression and the cell types they interacted with [76, 86]. The multi-faceted response of macrophages under inflammatory conditions have also been documented in myocardial infarction [87, 88], chronic obstructive pulmonary disease (COPD) [89], non-alcoholic fatty liver disease (NASH) [90], and many other diseases [91]. Thus, the polarization of macrophages in vivo should be considered a multidimensional process rather than a black or white M1 versus M2 dichotomy

[92].

1.2 Combining FACS and bulk RNA-sequencing to study macrophage heterogeneity

1.2.1 FACS and macrophage biology

Flow cytometry is an experimental technique for measuring physical characteristics of individual cells or particles. In brief, a light source is focused onto the particles being examined, which are directed and guided by fluidic systems. The resulting light scatter upon striking the particles is then recording by optical instruments, converted to digital signals, and then analyzed computationally [93]. Fluorescence-activated cell sorting (FACS) is considered a specialized type of flow cytometry, possessing the capability of distinguishing and isolating multiple cell populations. This is achieved by pre-labeling the single cells with fluorescent-tagged antibodies specific to cell surface protein markers of choice. Cells with different surface markers will then exhibit differing fluorescent emittance when ran through the flow cytometer, enabling their separation from others [94, 95]. For example, macrophages can be isolated by targeting their surface expression of CD11b and F4/80, while T cells can be identified through CD3, CD4, and CD8 (BD Biosciences). FACS has therefore seen widespread use in the field of immuno-cell biology, where populations of different lineages are often examined separately and thereby requires purification from the heterogeneous blood or tissue samples beforehand.

In addition to the major immune lineages, FACS is also a powerful experimental technique for discovering and characterizing subpopulations within a specific lineage. In the case of macrophages, FACS has been utilized to define distinct macrophage and dendritic cell (DC) subsets in the lung, where surface markers including CD206, Siglec F, and MHCII are found to be

major indicators for distinguishing subsets [96]. Other notable examples of employing FACS in macrophage-related studies include mapping the developmental trajectory of tissue resident macrophages [26, 97, 98] and characterizing changes to subpopulations under diseased conditions [88, 99, 100]. In summary, FACS is an essential tool for studying the heterogeneity of macrophages and is ubiquitously utilized across macrophage-related studies.

1.2.2 RNA-sequencing technology

After isolation by FACS, examining the differences in their gene expression patterns is a very desirable way to characterize individual macrophage populations. The quantification of gene expression is usually done by measuring the abundance of messenger ribonucleic acid (mRNA). This is based on the central dogma of biology, where deoxyribonucleic acids (DNA) are first transcribed into mRNA, which are then translated into amino acids and eventually become functioning proteins [101]. Given the fact that the process of mRNA transcription is selective at active gene loci on DNA, the relative abundance of mRNA molecules is a good indicator of differing gene expressions and in extension protein functions between sorted cell populations.

Over the past few decades, the technology for deciphering nucleotide sequences like DNA and RNA has seen tremendous improvements. The traditional method for sequencing DNA was introduced by Frederick Sanger in 1977. The method involves the selective incorporation of four possible complementary dideoxynucleotides (ddNTPs – Adenine (A), Thymine (T), Cytosine (C), and Guanine (G)) to the DNA strand being sequenced. The ddNTPs are radioactively or fluorescently tagged and would terminate the chain elongation process by DNA polymerase, which enables the determination of individual DNA nucleotides through gel electrophoresis [102].

Sanger sequencing was the prevailing DNA sequencing method for over 40 years, but it suffers from the limitations of low scalability and being labor intensive. For instance, the human genome project, which utilizes Sanger sequencing, took 13 years and 3 billion dollars to deliver the first completed draft of our genome [103]. Thus, the demand for higher capacity and cost-effective nucleotide sequencing led to the introduction of many high throughput sequencing, also known as next-generation sequencing (NGS) in the last couple decades [104].

There are currently three major commercial NGS systems: Roche 454, Illumina Genome Analyzer, and SOLiD/Ion torrent from Life Sciences [105]. Although their underlying chemistry differs, all three systems employed the concepts of massive parallelization and multiplexing. Here, the procedures of the Illumina platform are briefly described as they are employed in the experiments in this thesis. First, sequencing libraries are prepared by extracting nucleotide sequences from cells, fragmenting into shorter reads, and ligating adapter sequences to the reads. Next, each read is replicated up to 1000 times in close proximity on the surface of flow cell channels through the process of "bridge amplification", forming dense clusters of double-stranded reads. Third, a sequencing cycle is initiated by administering fluorescently tagged dNTPs with reversible chain elongation terminators. Fluorescence is emitted when a single complementary dNTP is incorporated on each read, which are captured and converted into digital signals by optical instruments. By distinguishing the colors of the emitted fluorescence, the identities of nucleotides can be deciphered across large number of reads at once. Finally, the terminators are cleaved and the dNTPs are added again to sequence the next base of the reads. This "sequencing by synthesis (SBS)" process is repeated over multiple cycles until reaching the predetermined read length, which is typically between 50 and 300 bases [106]. Compared to Sanger sequencing, NGS methods

like Illumina are able to sequence multiple biological samples simultaneously while providing better coverage of genomic regions due to their ability to demultiplex reads and their much higher sequencing capacity. We have thus seen widespread adoption of NGS assays for genome-wide experiments in recent years.

The next-gen assays most commonly employed for gene expression profiling are microarray and RNA-seq. Microarray is the older technology out of the two, and it functions by utilizing a specialized chip spotted with complementary DNA (cDNA) probes representing individual genes or transcripts. Expression differences among genes then can be observed and quantified by hybridizing sample RNA fragments against the probes [107, 108]. Microarray is highly applicable in clinical settings, especially for diagnostic purposes due to its relatively standardized nature [109]. However, it is limited by our current knowledge of transcriptomes, and therefore cannot be used for the discovery of unannotated genes and splicing variants. RNA-seq, on the other hand, utilizes the aforementioned NGS technology to obtain both the nucleotide sequences and expression levels of mRNA fragments [110]. RNA-seq involves an extra step of reverse transcribing mRNA fragments into cDNA, but otherwise the downstream procedures are shared with DNA sequencing. In comparison to microarray, RNA-seq has the advantage of being able to discover novel transcripts and possessing lower background noises due to its lack of hybridization step. It is also able to quantify absolute and wider ranges of expression levels as opposed to microarray, where expression estimates are relative [111]. Given these advantages, combining FACS and RNA-seq have become an extremely popular approach to characterize and compare the gene expression patterns between inter- and intra-tissue resident macrophage subsets.

There are several important factors to consider when designing an RNA-seq experiment. First, an RNA-seq run can be configured to either output single reads or paired end reads, where sequencing is performed on both ends of a fragment in opposite direction. While being more costly, a paired end run can provide extra information of read positioning in the genome, making it more suitable for tasks such as de novo genome assembly and discovery of structural rearrangements. For simpler tasks like comparing gene expression levels between samples, the cheaper single read runs are usually sufficient [112]. Second, two major methods exist for RNA extraction and purification during library prep, which are ribosomal RNA (rRNA) depletion and poly(A) tail enrichment. Similar to the choice between single versus paired end reads, rRNA depletion cost more but require less input materials and is less prone to biases towards the 3' end of fragments. rRNA depletion is therefore a desirable choice when working with samples of lower quality and for detection of splicing isoforms, while poly(A) enrichment is more appropriate for expression profiling on the whole gene level [113]. Third, the target number of reads to be sequenced in a run, also known as the sequencing depth or library size, need to be determined before executing the assay. Higher sequencing depths provides the benefits of better transcriptome coverage and therefore higher statistical powers when testing for differentially expressed genes, but may also amplify the noises from contamination and off-target transcripts. It is therefore suggested to utilize pilot runs and saturation curves to gauge the expected improvements to transcriptome coverage from increasing sequencing depth [114]. Finally, the number of biological replicates is another essential design factor. Albeit 3 replicates for each experimental group is commonly seen in studies, the actual optimal number is heavily affected by the underlying biological variability of the system being studied. Therefore, power analysis calculations should always be carried out beforehand to estimate the number of replicates to be used [115]. To summarize, the optimal parameter choices

when planning an RNA-seq experiment ultimately depends on the balance between experimental goals, desired statistical power, and budget.

1.2.3 Computational analysis of RNA-seq data

After conversion from fluorescent intensities into digital signals by sequencing machines, it is up to the computational pipelines to process and interpret the generated RNA-seq data. There is currently no consensus on how to "optimally" analyze RNA-seq data, as bioinformatic applications are continuously introduced with various functions and purposes. It is therefore up to the researchers to decide which combinations of bioinformatic tools best fit their research scenarios and goals. The computational analysis of RNA-seq data can be roughly categorized into distinct steps, which are described below:

A. Demultiplexing

The raw data generated by Illumina sequencers is stored in the Binary Base Call (BCL) format, which contains nucleotide base call and quality information for each location on the flow cell in each sequencing cycle. Software like bcl2fastq [116] would sort and assign sequenced reads back to individual biological samples based on the distinct read barcodes added during library prep, known as the process of demultiplexing. The demultiplexed data are outputted as FASTQ files, which is the standard input format for most bioinformatic tools.

B. Quality control

Raw reads of low quality are removed in this step to prevent technical artifacts and contamination from confounding analysis downstream. Metrics often utilized to assess overall

read quality include Phred scoring (Q30), GC content, adaptor contamination, duplication levels, and sequence overrepresentation. Tools such as FastQC [117], Trimmomatic [118], and RSeQC [119] are commonly employed for quality control of RNA-seq data.

C. Genome alignment or assembly

The next major step is to align the filtered reads to a reference genome or transcriptome if it is available. Most aligner software developed is based on the classic Burrows-Wheeler transform [120] and Needleman-Wunsch [121] algorithms. Some well-known examples include Bowtie [122] and BWA [123]. Some applications designed specifically for aligning RNA reads are able to account for potential splicing junctions, either by leveraging existing annotations or detecting new ones. Examples of popular "splice-aware" aligners include STAR [124], HISAT2 [125], and TOPHAT2 [126]. If a reference genome is not available, de novo assembly of transcripts can be performed using specialized tools like Trinity [127] and Oases [128].

D. Expression quantification

After the reads are aligned to the reference genome, there are multiple methods to quantify the expressions of individual transcripts or genes. The most straightforward method is to assign the uniquely mapped reads to genes based on existing annotations of exon coordinates, and is implemented in tools such as HTSeq-count [129] and featureCounts [130]. With this simple approach, ambiguously mapped reads are discarded and splicing variations cannot be recovered. On the other hand, applications like Cufflinks [131], RSEM [132], and StringTie [133] utilizes expectation maximization or graph-based approaches to enable more accurate estimation of splicing variant abundances [134], at the cost of being more computationally

intensive. Finally, tools including Kallisto [135] and Salmon [136] introduced the concept of "pseudoalignment", where genome alignment and expression quantification are essentially merged into a single step by modelling the relationships of k-mer occurrences between sequenced reads and reference transcripts. These alignment free tools provide massive improvements in speed and memory usage while exhibiting comparable overall accuracy to alignment-based applications, but is limited in their power for isoform discovery and quantifying lowly expressed transcripts [137].

After performing the steps mentioned above, the outputted data usually come in the form of count matrices, where rows represent genes or transcripts, columns represent individual biological samples, and cell values constitute the number of sequenced reads for each gene-sample pair. The subsequent analyses from count matrices become a lot more open ended, as appropriate strategies largely depend on specific research scenarios and therefore varies from experiment to experiment. Here, I detailed some of the tasks most often undertaken for downstream RNA-seq analyses.

E. Normalization

Normalization of count matrices is considered an essential step in the workflow, as technical variance across samples can lead to batch effects and confound downstream interpretation. The most commonly used method is Counts Per Million (CPM), where the count values are divided by the library sizes/sequencing depths of each sample and scaled by a factor of 1000000. Other popular normalization methods include Fragments Per Kilobase of transcript per Million (FPKM), Reads Per Kilobase of transcript per Million (RPKM), and Transcripts Per Million (TPM), which also remove the bias from gene or transcript lengths in addition to library sizes

[138].

F. Filtering out noises

Genes with low normalized expression values should be excluded from further analysis, given that they have a higher likelihood of being technical noises. The low threshold for expression value can be visually determined by plotting the distributions of gene number across their range of expression levels for all samples [139].

G. Exploratory analysis

To assess sample variability within and between groups and identify potential outlier samples, correlation coefficients across samples can be calculated using normalized expression values and visualized in heatmaps, while scatter plots of individual gene expressions are often utilized to examine relationships between any two samples. Another commonly employed method is Principal Component Analysis (PCA), which is a classic dimensionality reduction algorithm that enables visualizing distribution of samples in a 2-dimensional plot [140].

H. Differential Expression Analysis

Traditional parametric statistical tests such as T-test and ANOVA are unsuitable for the identification of differentially expressed genes (DEGs) because gene expression counts usually do not conform to their assumption of data normality. More sophisticated statistical models designed specifically for RNA-seq data are implemented in the popular DESeq2 [141] and EdgeR [142] packages, which utilizes negative binomial distribution to model count variability. Other often used packages for DE analysis includes baySeq, NOISeq and limma, which utilizes

Bayesian, non-parametric, and linear statistical models respectively [143-145]. For identification of differential splicing variants, the pipeline of HISAT – StringTie – Ballgown is suggested to be an effective approach [146]. Differentially up- and downregulated genes between conditions are usually determined by thresholding on both p-values and expression fold changes.

I. Clustering

Clustering is the task to categorize genes with similar expression patterns into different groups. Similarity between genes can be quantified through distance metrics such as Euclidean, Manhattan, and correlation. There are two major algorithms for gene expression data clustering, hierarchical and k-means clustering. Hierarchical clustering works by first treating each sample as a cluster, then iteratively merge the two most similar samples together until all clusters are merged, generating a dendrogram in the process. The final clustering can be determined by drawing a cutoff line on the dendrogram. On the other hand, K-means clustering operates by initializing a *k* number of "centroids", or mean of cluster, at random locations and assigning individual samples to them. The sample assignments and centroid locations are updated iteratively until the sum of distances between centroids and their assigned samples is minimized. The value of *k* needs to be supplied by user, and can be roughly estimated through elbow and silhouette methods [139].

J. Functional Characterization

After obtaining DEGs or clusters, it is important to characterize and understand the biological relevance of these individual gene sets. This is typically achieved through executing statistical

enrichment tests, such as Fisher's exact and hypergeometric tests, against a standardized biological knowledge repository like Gene Ontology (GO) [147] and KEGG pathways [148] to attribute specific biological processes to each gene set. Popular applications for performing functional enrichment tests includes GOrilla [149], DAVID [150], and PANTHER [151]. Gene set enrichment analysis (GSEA) is another commonly used alternative, where a user-supplied gene list ranked by either p-values or fold changes is compared against a pre-defined gene set retrieved usually from Molecular Signatures Database (MSigDB) [152]. An enrichment score and its significance are calculated based whether genes in the pre-defined set is statistically overrepresented towards the top or bottom of the ranked gene list [153, 154].

1.2.4 Examples and limitations of employing FACS and bulk RNA-seq in macrophage-related studies

Given the advantages outlined above, it is a very popular experimental approach to first isolate individual macrophage populations using FACS, and then subject them to gene expression profiling to functionally characterize and compare across subpopulations. One of the earlier examples was performed as part of the Immunological Genome Project (Immgen) [155], in which researchers aim to reconstruct complete gene regulatory networks for immune cell lineages. Gautier et al. [51] extracted macrophages from murine spleen, lung, peritoneal, and brain tissues and performed microarray on them. They confirmed the diverse gene expression patterns among macrophages from different tissues and was able to infer TFs that drives the regulatory programs of individual populations using the Ontogenet algorithm [156]. A more comprehensive study performed later by Lavin et al. [47] isolated macrophages from eight different tissues and performed bulk RNA-seq instead, providing a higher-resolution catalogue of tissue resident

macrophage expression signatures. In addition, the authors were able to predict regulatory TFs more confidently by directly profiling their epigenomic landscapes using ATAC-seq and ChIP-seq. The combination of FACS followed by RNA-seq is also commonly utilized to interrogate macrophages within a specific tissue. For example, Matcovich-Natan et al. [157] and Thion et al. [158] extracted microglia precursors over multiple time points during fetal and post-natal development and profiled their gene expression patterns across the time course.

However, some limitations exist for this approach. First, prior knowledge of protein surface markers is required to separate cell populations. Therefore, FACS is inherently limited in its power for discovery of novel macrophage subpopulations. Furthermore, given that most FACS experiments utilize sorting panels of only a few dozen surface markers, one cannot exclude the possibility of rare and novel subpopulations being sorted together within larger populations even when using "gold standard" marker panels. When bulk RNA-seq assays are performed, this results in the potential transcriptional heterogeneity being masked and unrecoverable even by computational analyses since gene expressions of the sorted cells are averaged out early in the workflow. Fortunately, the introduction of single cell RNA-sequencing technologies (scRNA-seq) provides a direct answer to these limitations.

1.3 Using Single Cell RNA-sequencing to study macrophage heterogeneity

1.3.1 scRNA-seq technology

As a more recent technological breakthrough, expression profiling on the single-cell level was first introduced in 2009 by Tang et al., where a single mouse blastomere was analyzed [159]. Many more powerful scRNA-seq protocols have subsequently been introduced and can be roughly

characterized as either well-based or droplet-based depending on their cell capture approaches. As their names suggest, well-based methods (e.g., SMART-seq [160] & CEL-seq [161]) involve separating the extracted cells into "wells" on a microchip, providing potentially better sequencing coverage for each individual cell but is limited in the number of cells that can be profiled at once. In contrast, droplet-based methods (e.g., Drop-seq [162] & 10X Chromium [163]), which were introduced later, involve capturing cells into droplets through microfluidic devices and possess much higher throughput than well-based methods. scRNA-seq protocols also varies in their coverage of transcripts, where some would provide full-length information, like Smart-seq and Quartz-seq [164], while others are biased for 3' end of transcripts, including CEL-seq and MARS-seq [165]. Currently, the most widely adopted protocol by far is 10X Chromium, which is a commercialized droplet-based 3'-end method that is shown to exhibit overall better performance across multiple benchmarking metrics [166].

Compared to traditional bulk RNA-seq, library preparation for scRNA-seq protocols like 10X chromium involves largely parallel steps, including reverse transcription and PCR amplification. A key difference for 10X chromium library prep is the barcoding of single cells, which occurs when individual cells are captured and encapsulated in microfluidic droplets containing unique barcode sequences. Another major feature of 10X chromium protocol is the use of unique molecular identifiers (UMIs), which is another kind of barcode sequence for tagging individual transcripts [167]. PCR duplication bias is a major obstacle for scRNA-seq protocols due to the much lower input RNA available for each library (individual cells vs biological replicates), which can be significantly ameliorated through incorporation of UMIs before the amplification step. These modifications to the protocol enable the profiling of single-cell transcriptomes at very high-

throughput – a single run of 10X chromium is able to report the gene expressions of up to 80000 cells.

scRNA-seq enjoy the major advantage of being able to reveal previously underappreciated heterogeneity within cell populations, and have therefore experience widespread applications in macrophage-related studies. It is commonly utilized as an exploratory tool to uncover novel macrophage subpopulations within tissue, especially in the context of pathological conditions. For example, scRNA-seq aided in the identification of a novel aortic macrophage subset associated with atherosclerosis, characterized by the expression of Trem2 and a potential functional role in lesion calcification [168]. Other examples of scRNA-seq application includes the examination of microglia heterogeneity in brain tissues with glioma [169] and the identification of two previously unknown lung interstitial macrophage subsets that are functionally distinct [170]. scRNA-seq is also especially useful for the characterization of macrophage environmental niches, since all cell types that contribute to a niche can be profiled unbiasedly. This was exemplified in Joshi et al., where scRNA-seq and ligand-receptor interaction analysis was performed to confirm the roles of monocyte-derived alveolar macrophages and M-CSF signaling in the establishment of fibrotic niche [171]. The introduction of single cell spatial transcriptomics methods provides further potential for accurate delineation of macrophage niches [172]. In summary, scRNA-seq is a highly potent experimental method for unbiased investigation of macrophage heterogeneity that has quickly gained popularity in recent years.

1.3.2 Computational analysis of scRNA-seq data

The initial computational processing of scRNA-seq data is largely similar to traditional bulk RNA-seq. Usually handled through automated pipelines like Cell Ranger [163], the sequenced

reads are first demultiplexed to their individual cells of origins. The reads of low quality are trimmed and then aligned to the reference genome. UMIs, rather than individual reads, are aggregated into a count matrix with gene annotations as rows and cell barcodes as columns. Finally, the count matrix is filtered by a specialized algorithm to remove barcodes with low RNA content, which likely represent technical artifacts rather than functional, intact cells [173].

After obtaining the filtered gene by cell UMI matrix, there are many specialized packages/tools available for its downstream analysis, such as Seurat [174], scater [175], and Scanpy [176]. These packages implemented specialized modifications to their workflow to account for features specific to single cell data. Namely, scRNA-seq data usually exhibits higher technical variance, data dimensions, and count sparsity. relative to their bulk counterpart. Here, I will briefly describe the major steps in a typical Seurat analytical workflow, which is currently the most popular tool for downstream analysis, and highlight the differences to the analysis for traditional bulk RNA-seq data:

A. Quality control

In contrast to bulk RNA-seq, another round of quality control is usually performed in the downstream analysis for scRNA-seq data because they are more prone to higher level of technical noises. Specifically for droplet-based methods, the cell capturing process and failed amplification can lead to the missed detection, or dropout, of certain transcripts. It is also possible that broken cells or even multiple cells (doublets) can be encapsulated in droplets, resulting in low quality data. Therefore, Seurat suggest removing the low-quality cells by thresholding on the number of expressed genes and UMI counts. The proportion of

mitochondrial reads is another useful indicator for low-quality cells, as mitochondrial RNA fragments are usually disproportionately retained in broken or dead cells.

B. Normalization

Seurat implemented a very straightforward approach of normalizing scRNA-seq data by correcting for total UMI counts in each cell, log-transforming, and multiplying by a scaling factor of 10000. Other more sophisticated normalization methods that aim to remove the effects of technical drop out events have been introduced and implemented in scran [177] and sctransform [178] packages.

C. Dimensionality reduction

scRNA-seq data are high dimensional in nature, as they contain information for thousands of genes and cells at once. To facilitate user interpretation, it is preferable to computationally project the data into a lower dimensional space to enable visualization of distinct cell populations while maximizing the information retained (Figure 1.3.2). In practice, only the highly variable genes are used as input to focus on the strongest biological signals. There are three popular algorithms for reducing the dimensions of scRNA-seq data, which are Principal Component Analysis (PCA) [140], t-Distributed Stochastic Neighbor Embedding (t-SNE) [179], and Uniform Manifold Approximation and Projection (UMAP) [180].

PCA possesses the advantage of scaling up well to large datasets while being easily interpretable, due to the fact that each individual gene's contribution to a principal component (PC) can be manually inspected. However, a major drawback of PCA is that it can only detect

linear relationships, which can be inadequate to describe the inherently nonlinear single cell data. Nowadays, PCA is often performed as a preprocessing step in scRNA-seq workflows including Seurat to reduce overhead computational burdens and technical noises, where the top-ranking (10-30) PCs are selected as the input for the other two dimensionality reduction algorithms. t-SNE and UMAP, on the other hand, are both designed to resolve non-linear relationships and are therefore well suited for scRNA-seq data. UMAP, which was developed particularly for visualization of single cell data, is generally better at recapitulating continuous biological transitions compared to t-SNE due to its improved handling of both local and global distances. It also possesses the advantages of being deterministic and having much better runtimes and scalability than t-SNE [181]. UMAP has therefore increasingly replaced t-SNE as the prime choice for dimensional reduction and visualization of scRNA-seq data.

D. Clustering

After the visualization of individual cells on a two- or three-dimensions plot is achieved, unsupervised clustering can be applied to group cells with similar expression profiles together. Seurat and many other packages like Scanpy perform the clustering of single cells using the Louvain algorithm [182]. It is a modularity optimization algorithm that operates on a k-nearest neighbor (KNN) graph with edges drawn between single cells with similar gene expression patterns. The graph is then iteratively partitioned into clusters of highly interconnected regions. The number of resulting clusters can be indirectly controlled by a “resolution” parameter supplied by user. While other single cell clustering methods that utilizes other algorithms like K-means [183] and self-organizing map (SOM) [184] also exist, a systematic evaluation study has shown that Louvain clustering consistently outperform all other methods when applied on

scRNA-seq data.

E. Cell annotation

Annotation of cell identities is a crucial step in the workflow after obtaining clusters, but there is currently no consensus of what represents the best approach. This is mostly because the definition of a "cell type" is ambiguous and is largely a question of resolution. For example, while lung macrophage subpopulations possess distinct expression signatures, they would look like a homogenous population when compared to other immune cells such as T cells. Seurat suggest simply using a couple of previously published gene markers to manually annotate clusters. This is obviously subjective and prone to biases, but many published studies still rely solely on this marker-based approach. Methods that perform automated cluster annotation is just beginning to be introduced [185]. Some examples include the marker databased-based CellAssign [186], correlation-based SingleR [187], and supervised classification-based Garnett [188]. It is recommended that both manual and automated approaches be carried out to increase the confidence of final annotation [189].

F. Differential expression analysis

Differential expression analysis in the context of single cell data usually involves identifying genes with distinct expressions in one cluster compared to others. Seurat uses a simple non-parametric Wilcoxon Rank Sum Test by default, but many other tools designed specifically for scRNA-seq data are also available. Some notable examples include MAST [190], SCDE [191], and DESingle [192], which utilize different statistical models like Poisson [191] and Zero-Inflated Negative Binomial [192] distributions to deal with specific characteristics of single

cell data, such as dropout events [193]. Applications originally designed for bulk RNA-seq analysis, including DESeq2 [141] and EdgeR [142], have also seen extensive use in scRNA-seq studies. A benchmarking study showed that the resulting DEGs can differ considerably when using different tools on the same datasets, but there is no major performance loss with bulk RNA-seq tools compared to those dedicated for scRNA-seq data [194].

In addition to core workflow, there are other optional tasks for scRNA-seq analysis that are often performed depending on the scenarios and needs of individual studies, which are described below in no particular order:

G. Count imputation

Owing to the high dropout rates in single cell data, imputation methods that aim to "recover" some of missing gene expression values are developed and can be a viable approach prior to data normalization. Commonly used tools for this task includes SAVER [195], MAGIC [196], and AutoImpute [197], each utilizing very different statistical approaches.

H. Regressing out biological and technical effects

Seurat and Scanpy both implemented the function to remove known biological or technical covariates specified by user via simple linear regression [189]. It is especially common to regress out the effects from cell cycle genes since they often mask the signals from other underlying biological processes. Other more specialized methods that rely on mixture models is implemented in f-scLVM package [198].

I. Dataset integration

As scRNA-seq experiments become more affordable, it may be desirable to perform multiple assays across different biological conditions (e.g., healthy versus diseased tissue) and compare the resulting datasets directly. This can be achieved through numerous single cell data integration algorithms such as Canonical Correlation Analysis (CCA) [199], Mutual Nearest Neighbors (MNN) [200], and Harmony [201]. Besides direct integration of multiple datasets, algorithms like Cell Anchoring [174] also enables the projection of cell annotations from one dataset to another. These single cell specific methods differ from typical batch correction approaches such as ComBat [202] in that they strive to preserve the cell compositional variations that may exist between datasets.

J. Trajectory inference

Being able to examine transcriptomes on the resolution of single cells enables the inference of cell state transitions and differentiation branching processes. In brief, a trajectory line is drawn over cells by modeling the continuous gradient of gene expression changes. Each cell is assigned a value based on its location on the inferred trajectory, termed “pseudotime”. Popular tools that perform trajectory inference includes Monocle3 [203], TSCAN [204], SlingShot [205], and PAGA [206]. Trajectory inference can be performed to supplement clustering analysis as the latter separate cells into discrete groups, and is therefore unable to capture subtle transitional states across distinct populations.

K. Gene regulatory network (GRN) reconstruction

By surveying for gene co-expression patterns across cells, the complex regulatory interactions

between genes and environmental stimuli can potentially be recapitulated. The method development for GRN reconstruction is still in its infancy, however, with research efforts ongoing for both bulk and single cell RNA-seq data. Tools currently available for single cell data include SCENIC [207], which functions by leveraging and integrating co-expressing TFs and cis-regulatory motif information. On the other hand, SingleCellSignalR [208] focuses on the identification of ligand-receptor (LR) interactions by utilizing a curated database while accounting for dropout events using a regularized prediction score.

In summary, there have been tremendous progress in the development of analytical tools for scRNA-seq data analysis since the technology's introduction merely a decade ago. The development of other single cell protocols such as CITE-seq [209] and scATAC-seq [210], which profiles cell surface protein markers and open chromatin regions respectively, further broadens the horizon for future tool development. New analytical methods being developed for the integrative analysis of multimodal single cell data [211, 212] holds great potential for more thorough investigation of macrophage heterogeneity and functional plasticity.

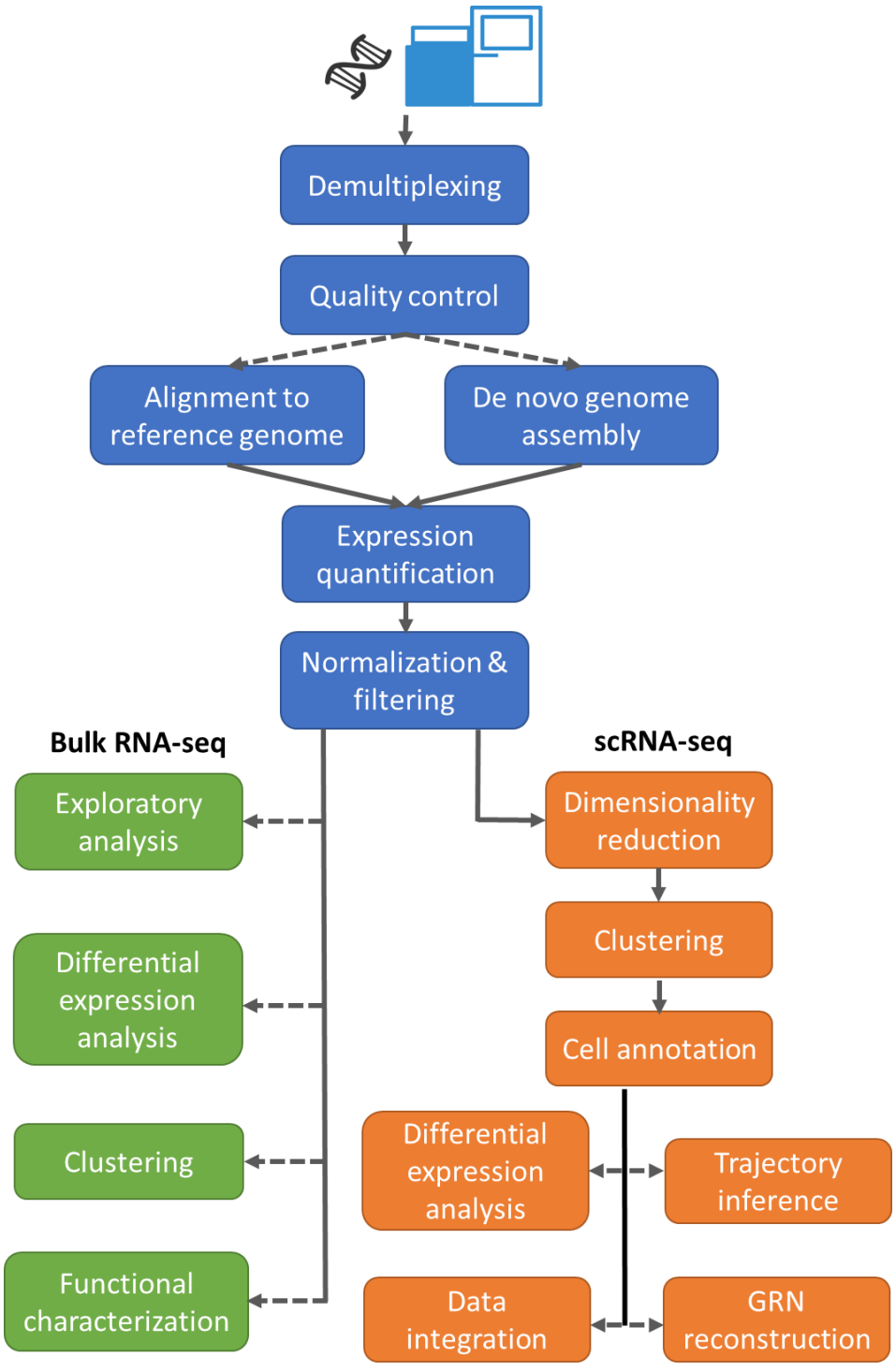


Figure 1. Common tasks for computational analysis of bulk and single cell RNA-seq experiments.

1.4 Summary and Objectives

Macrophages are a group of immune cells that display remarkable functional diversity and plasticity, which are vastly underappreciated until recently. The introduction of next-generation sequencing technology facilitated deeper interrogation of macrophage transcriptomes between specific subpopulations isolated by FACS. The rising popularity of scRNA-seq assays further enabled the characterization of macrophage heterogeneity and plasticity at an unprecedented resolution. In this thesis, I will demonstrate how integrative computational analysis of bulk and single cell RNA-seq data can be utilized to further our understanding of macrophage heterogeneity across multiple tissue and disease models in both mouse and humans. First, we characterized and defined subpopulations of resident joint macrophages and then assessed their transcriptional changes under arthritic conditions. In the next chapter, we extended the concept of resident macrophage diversity to myeloid cells in general by confirming the presence of long-living non-classical monocytes residing in synovial space, which are functionally and ontogenically distinct from circulating ones. In chapter 4, the focus was shifted to hepatic macrophages, where we uncovered the existence of three distinct subpopulations associated with human pediatric cholestasis. Finally, after identifying a major shortcoming with current analytical workflows, I developed a novel web application that aims to improve functional enrichment analysis for macrophage-related genomic data.

CHAPTER 2

Characterizing the heterogeneity of synovial macrophages

Chapter 2 is adapted from original research articles entitled “Synovial Macrophage Heterogeneity Confers Differential Response to Acute and Chronic Inflammatory Arthritis.” (Chen & Montgomery et al., in preparation) and “Critical Role of Synovial Tissue-resident Macrophage Niche in Joint Homeostasis and Suppression of Chronic Inflammation (Huang et al. 2021, Science Advances). They are reproduced here with the permission of the copyright holders.

2.1 Introduction

Macrophages are innate immune cells that participate in phagocytosis and maintenance of homeostasis [5, 213, 214]. It is now well established that macrophages reside in nearly every tissue in steady state and readily adapt to local microenvironmental signals, leading to specialized tissue functions and phenotypes [47, 54, 215]. For example, microglia, the resident macrophages of the brain, participate in synapse pruning during development [10, 216]. Kupffer cells of the liver are involved in breaking down erythrocytes and recycling of heme [12, 13]. Alveolar macrophages of the lung play major roles in removing foreign particles from the respiratory surfaces [14, 15]. These studies, along with others, highlight the plasticity and functional diversity of macrophages. Within the joint tissue, macrophage resides in both the synovial lining and interstitial or sub-lining spaces [217]. Their specialized function remains to be formally elucidated, but they are commonly thought to participate in the homeostatic maintenance of joint function by clearance of cartilage and bone debris from the synovial fluid [218, 219]. In addition to homeostatic and anti-inflammatory properties, emerging evidence suggest that macrophages residing in the synovial lining form a physical barrier that shields the synovial cavity, protecting it from external perturbation and potentially regulate the chemical composition of synovial fluid [219, 220].

It was traditionally thought that tissue macrophages were exclusively derived from circulating monocytes through hematopoiesis in the adult bone marrow [221]. More recent studies have challenged this paradigm, reporting the existence of tissue-resident macrophage subsets that are derived from embryonic progenitors, populate tissues during early fetal development, and are long-lived with minimal replenishment from circulation [8, 26, 29, 222, 223]. Thus, considerable ontogenic and functional heterogeneity for macrophages can exist within tissues. Macrophages in

the lung, for example, have been classified into tissue-resident alveolar macrophages and two subtypes of monocyte-derived interstitial macrophages [224]. Recent studies that utilized fate-mapping and single cell profiling techniques demonstrated that synovial macrophages in both mice and humans are also composed of distinct subsets with embryonic or bone marrow origins [220, 225-228]. Whether finer subdivisions and subset-specialized functions exist for synovial macrophages at steady state is incompletely understood. A better characterization of synovial macrophage subpopulations is a prerequisite for a more detailed assessment of macrophage phenotypes under diseased conditions, as different subsets may exhibit distinct responses.

As the most abundant resident immune cells in the joint [229], synovial macrophages are central players in the pathogenesis of RA, which remains one of the most prevalent autoimmune disorders [230, 231]. RA is associated with chronic inflammation of the synovial membrane, which eventually leads to degradation of cartilage and bone structures as the disease progress [232]. Macrophages drive RA-associated synovial inflammation through secretion of pro-inflammatory cytokines, degradation of extracellular matrix, and recruitment of other immune cells, including neutrophils and T cells [233-235]. Synovial macrophages also mediate the destruction of cartilage and bone structures through signaling interaction with fibroblasts and osteoclasts [236, 237]. Inflammation in the joint is accompanied by the expansion of activated and proliferating synovial macrophages [219, 238]. We have previously utilized chimera models and selective depletion approaches to demonstrate that resident synovial macrophages are required for the attenuation of arthritis. In contrast, recruitment of peripheral monocytes, which differentiate into macrophages, is required for both the propagation and resolution phases of arthritis, highlighting the time-dependent plasticity of monocyte-derived macrophages [225]. How monocyte infiltration alters

the preexisting synovial macrophage heterogeneity and their functions during disease progression remain to be elucidated. A better understanding of synovial macrophage dynamics over the course of inflammatory arthritis could potentially help pinpoint pathways and therapeutic targets for treating RA that are specific to particular subpopulations.

In this study, we employed an integrative approach of fate mapping and transcriptional profiling to define four distinct subsets of synovial macrophages at steady state. We utilized two mouse models, serum transfer induced arthritis (STIA) and collagenase-induced arthritis (CIA), which represent acute and chronic arthritis respectively, to assess the function of identified subsets in response to joint inflammation. We uncovered evidence of more diverse transcriptional responses during chronic inflammatory arthritis compared to acute. Finally, we used our definition of murine subpopulations to characterize human synovial macrophage heterogeneity in RA patients. We observed significant associations between proportions of macrophage subsets with RA disease severity and treatment responsiveness.

In Chen et al., we employed an integrative approach of bulk and single cell transcriptional profiling to characterize the heterogeneity of synovial macrophages at steady state, defining four distinct subpopulations based on the expression of Cx3cr1 and MHCII. We utilized two mouse models, serum transfer induced arthritis (STIA) and collagenase-induced arthritis (CIA), which represent acute and chronic arthritis respectively, to assess the function of identified subpopulations in response to joint inflammation. Finally, we used our definition of murine sub-populations to characterize human synovial macrophage heterogeneity in RA patients. Collectively, our results indicate that heterogeneity of synovial macrophages confers differential response to

acute and chronic inflammatory arthritis.

In Huang et al., we utilized a novel mouse model developed by our group called HUPO, where the RA risk locus Flip (FLICE-like inhibitory protein) [239] was deleted through mediation of Cre recombinase under the control of a CD11c promoter. HUPO mice spontaneously develop chronic arthritis that is dependent on macrophages for its initiation, with the incidence increasing with age to 80% by ~22 weeks after birth [240]. We identified F4/80^{hi}MHCII− (FH1) macrophages, in which Flip was highly expressed, as the dominant tissue-resident macrophage subset, which was essential during homeostasis for maintaining niche integrity. In contrast, Flip was reduced in F4/80^{hi} macrophages in HUPO, associated with reduction of the FH1 subset, opening a niche and permitting the influx of circulating proinflammatory monocytes, which differentiated into the F4/80^{hi}MHCII+ (FH2) subset. Genes important in macrophage tissue residency are reduced in monocytes from inflammatory RA synovium, suggesting a role for the disruption of the synovial resident macrophage niche in the transition from preclinical to clinical RA and the progression of disease.

2.2 Synovial Macrophage Heterogeneity Confers Differential Response to Acute and Chronic Inflammatory Arthritis

2.2.1 Materials and Methods

Mice (Performed by Anna Montgomery)

C57Bl/6 and CX3CR1f/wtCreER.zsGFP mice were purchased from Jackson Laboratory. Mice were housed in specific pathogen-free facilities at Northwestern University. All experimental

procedures were carried out on female mice aged 8-10 weeks unless otherwise stated. The mice were bred and maintained in the Northwestern University barrier animal facility, and all procedures followed ethical guidelines and approved by Northwestern IACUC.

Arthritis induction (Performed by Anna Montgomery)

Serum transfer arthritis (STIA) was induced by the injection of K/BxN serum obtained at 8 weeks from the progeny of the KRN X NOD [72]. Collagen-induced arthritis (CIA) was induced by injecting an emulsion of complete Freund's adjuvant and type II collagen [65]. Disease was assessed using clinical scoring 3 times per week for 21 days for STIA and 62 days for CIA.

Processing of murine synovial tissue (Performed by Anna Montgomery)

Joints were removed from hindpaw in pairs following euthanasia and perfusion. Skin and toes were removed from each paw and bone marrow were flushed from exposed tibia with sterile HBSS. Synovial tissue was then infused with 1.5ml/joint of ankle digestion buffer (2.4mg/mL dispase II, 2mg/mL collagenase D, 0.2mg/mL DNAse I in HBSS pH 7.2-7.6) before incubation at 37C for 1h with shaking. Cells were then agitated through a 40-um mesh filter. Erythrocyte lysis was performed using 1x PharmLyse. Dead cells were stained with eFluor 506 viability dye (eBioscience) (1:1000 dilution). Cells were incubated with FcBlock (BD Bioscience) and stained with antibodies (see table) for fluorescence-activated cell sorting (FACS). 23count eBeads (Thermo Fisher) were utilized to calculate cell counts. Flow cytometry analysis was performed using FlowJo software. Statistical analysis of flow-cytometry data was carried out in GraphPad Prism. Data shown is mean minimum N=6 ±SEM. P values were calculated using unpaired T Test, and were considered statistically significant if $P<0.05$.

Mouse single cell RNA-seq library preparation and analysis

For murine steady-state (day 0) and peak (day 7) STIA samples, CD45+CD11b+Ly6G-SiglecF- cells were sorted. RNA libraries were prepared using droplet-based 10x Chromium Single Cell 3' Solution v2. The libraries were sequenced on Illumina Nextseq machines. Reads were processed and aligned to mm10 reference genome using *mkfastq* and *count* commands of cellranger v2 pipeline [163]. Subsequent analyses, including quality control, unsupervised clustering, and visualization of individual gene expressionwere carried out using Seurat v2 package [199] in R. Low quality cells that have < 500 detected genes or > 30% mitochondrial reads were removed. tSNE dimensionality reduction and Leiden graph-based unsupervised clustering were performed with top principal components (PC) chosen based on elbow plots (Steady state: 15 PCs; D7 STIA: 16 PCs) and resolution parameter of 0.8. SingleR package v1.0.5 [187] was used to annotate cells with Immgen reference cell types and flow-sorted macrophage subsets from bulk RNA-seq data (described in next section). Identification of marker genes was done by running Wilcoxon tests implemented in FindAllMarkers with 0.25 min.pct and 0.25 logFC cutoff. Cell cycle scoring was performed using G2/M and S phase gene sets provided in Seurat, converted to orthologous mouse genes using BioMart R package. Steady state was compared against the steady state subsets defined in this study by running Seurat AddModulesScore function with the reported Culemann gene markers as input. Trajectory analysis was performed on steady state sample using reduceDimension and orderCells functions from Monocle2 package [241]. Genes differentially expressed across the trajectory were identified through differentialGeneTest function and three gene modules were obtained by executing plot_pseudotime_heatmap. Projection of steady state subset annotations onto D7 STIA sample was performed using label transfer method through FindTransferAnchors and Transfer data functions[174] with top 15 PCs

from steady state data as reference data. Pearson correlations between steady state and D7 STIA macrophage subsets was calculated on the averaged expression profiles across single cells. The relative expression signatures of the bulk steady state and STIA time course RNA-seq clusters were computed and visualized in D7 STIA sample using Seurat AddModuleScore and FeaturePlot functions. The clustered heatmap of select marker gene expressions across Day 7 single cells was generated using ComplexHeatmap package.

Bulk RNA-seq library preparation and analysis

For each of the four macrophage subpopulations, RNA was extracted using PicoPure RNA Isolation kit (ThermoFisher) as per manufacturer's instructions. Library prep was performed using QuantSeq 3' mRNA sequencing kit (Lexogen) and sequenced on Illumina NextSeq. The resulting BCL sequencing files were demultiplexed using bcl2fastq into FASTQ format. The reads were trimmed using BBDuk v37.22 (http://jgi.doe.gov/data-and-tools/bb-tools/), aligned to mm10 genome with STAR [124], and mapped to reference gene exons using HTseq [129] to generate a matrix of gene expression counts. Raw counts are normalized to counts per million (CPM) to account for differing read depth across samples. Expressed genes were defined as those with expression greater than 16 CPM in at least 4 samples. This results in 7668 expressed genes for steady state, 8513 genes for STIA, and 9388 genes for CIA datasets. Principal component analysis (PCA) was performed using the prcomp function with data scaling and centering. Forcalculating the expression foldchange relative to Day 0 of STIA and CIA time course experiments, CPM values lower than 16 were adjusted to 16 to minimize confounding noises from lowly expressed genes. Temporally differential genes across the STIA and CIA time courses were defined as those with at least 2-fold change in expression between day 0 and any subsequent time points in at least one

macrophage subsets. K-means clustering and heatmap visualization on CPM values (steady state) or log expression foldchanges relative to day 0 (STIA and CIA time course) was performed using Morpheus web app (https://software.broadinstitute.org/morpheus). For each steady state K-means cluster, we quantified the percentage of temporally differentially genes using the prior definition across the STIA time course and macrophage subsets. MAGNET web app (https://magnet-winterlab.herokuapp.com/magnet) was utilized to determine hypergeometric enrichment of the steady state clusters against published gene sets from Lavin et al. [47]. Gene ontology (GO) analyses of biological processes were performed using GOrilla webtool [149] with 8513 (STIA) and 9388 (CIA) expressed genes as background. Transcription factor binding motif enrichment analysis was carried out with HOMER [49]using each k-means cluster as input and expressed genes as background. GSVA R package [242] was utilized to compute the combined relative expression scores for putative downstream target genes of select TFs. The downstream target genes for the select TFs were obtained from Dorothea database [243] by filtering for positive regulatory direction (mor=1) and using all levels of confidence (A-E). To assess the similarity between STIA and CIA clusters, log enrichment of overlapping genes ((size of overlap/ # of genes in STIA cluster)/(# of genes in CIA cluster/ size of background)) cluster was calculated, with significance determined through hypergeometric test using 8168 shared expressed genes as background. All computational analyses were performed using R v3.6.3, with figures generated via ggplot2 package.

Human synovial biopsy data processing and analysis

RA patients were recruited as part of the RhEumatoid Arthritis SynOvial tissue Network (REASON) consortium. Minimally invasive ultrasound-guided joint biopsies were performed as

previously described [244]. In brief, ultrasound scanning over wrist joint was used to locate the region of greatest synovitis. Two lidocaine injections were performed into the target joint to ensure anesthesia and a clear path for the biopsy device. A Quick Core Biospy needle was then used to retrieve synovial tissue with the aid of continuous real-time ultrasound imaging. This procedure was repeated to obtain a total of 12 samples for each patient.

Mechanical disaggregation of collected tissue samples was performed on a GentleMACS dissociator (Miltenyi Biotec). Tissues were then infused with a digestion buffer [RPMI 1640 (Sigma), Liberase TL (Roche, 0.1mg/mL) and DNase (Roche, 0.1mg/mL)] and minced with scissors. Tissue suspensions were incubated for one hour at 37°C with aggressive disaggregation pre- and post- incubation. The digestion reaction was quenched with MACS buffer (Miltenyi Biotec) and the tissue suspension was filtered over a 40-micron filter. Red blood cells were lysed (BD PharmLyse) and then washed twice with HBSS (Thermo Fisher). Cells were counted (Invitrogen Countess) and stained with a viability dye (Supplemental Table 1; 0.5 μL/mL, 15 min., 25°C, dark). Cells were then washed twice with MACS buffer, incubated with Fc block (BD Bioscience), stained with antibodies, washed twice and re-suspended in MACS buffer and kept on ice until sorting. Synovial immune cells were sorted on a BD FACSAria SORP instrument (BD Bioscience).

Sorted CD45+ synovial immune cells, with the exception of sample 1022, which was additionally enriched for CD11B+ myeloid cells, were utilized as input for 10x Genomics Chromium Next GEM 3’ v2 or v3 library preparation protocol. Reads were aligned to hg19 reference genome and processed with cellranger v3 pipeline. To account for variability among

samples, quality control was performed by filtering for cells that possess < 10% mitochondrial reads and have > 0.4*median and < 3*median read counts for each sample. The human samples were integrated using Seurat anchoring method (FindIntegrationAnchors and IntegrateData) with 30 CCA dimensions and visualized by tSNE using top 10 PCs. CITE-seq was performed on two of the samples (hs5, hs6), with antibody intensities normalized using CLR method in NormalizeData function. To distinguish myeloid cells from other lineages, a scoring system was devised using top myeloid marker genes identified in mouse samples, which were converted to their human orthologous counterparts using BioMart package [245]. Briefly, this was achieved by leveraging the hidden ***AddModuleScore*** function within the Seurat package. Originally introduced by Tirosh et al. [246], the algorithm consists of the following steps:

1) Assign all genes into bins (n=25 by default) based on their average expressions across all cells.
2) Assuming that we are calculating the combined module score for 10 query genes, generate a control gene set by randomly selecting 10 genes from the matching expression bins for each query gene.
3) Compute scores for query and control gene sets by calculating mean expression of all genes for each cell.
4) Subtract control scores from query scores, and the resulting number for each cell is the enrichment module score for all query genes combined.

The ***AddModuleScore*** function was implemented in Seurat as an internal function for the calculation of cell cycle scores and is not directly mentioned in the user vignettes. This leads to missed potential for more innovative analytical strategies due to its flexibility of deriving module scores for gene sets of any context and sizes. Here, we computed myeloid signature scores using the top 6 marker genes (Ctss, Fcer1g, Psap, Crip1, Lyz1, Ftl1, Npc2, Fxyd5) and observed a clear

bimodal distribution across the cells (**Figure 2.2.6**), enabling in silico separation of myeloid cells with a simple cutoff.

To annotate the four synovial macrophage subpopulations defined in mice onto identified human myeloid cells, we developed a gene signature scoring approach that again utilizes the AddModuleScore function:

1) Top 10 marker genes for each of the four macrophage subpopulations from D7 STIA sample, ordered by fold changes, were used to calculate gene module scores for each human cell (**Table 2**).
2) The scores were normalized separately by each patient into the range of -1 and 1 to account for variability across samples.
3) The minimum scoring cutoff for annotating a cell was set to the equivalent of the pre-normalization score of 0 (**Figure 2.2.1**).
4) Each cell was assigned to the macrophage population with the highest normalized score above the minimum cutoffs. If normalized scores for all populations is below the minimum cutoffs, the cell is annotated as N/A.

The overall quality of the annotations was assessed by comparing the overlaps between annotated subsets and cluster labels from de novo clustering (resolution = 0.2). The same approach was applied for the analysis of the AMP [227] dataset. Similarity of our defined macrophage subsets with published clusters in Alivernini et al [17, 18] was determined by calculating the fraction of their top 20 reported markers that overlapped with the list of differentially expressed genes for each of our subpopulations. Significance of overlap was determined by hypergeometric

distribution with the expressed genes as background, defined as those that were present in more than 25% of cells for at least one subpopulation (n = 6063). A FWER cutoff of 0.05 was applied based on Bonferroni correction.

Table 2 Top 10 marker genes of the four macrophage subpopulations from D7 STIA sample used for calculation of gene module scores

MA	MB	MC	MD
CTSB	PLAC8	PF4	IL1B
LGMN	HP	SEPP1	HLA-DQB1
SYNGR1	TMSB10	MRC1	CD74
MS4A7	IFITM3	C1QC	PTGS2
PSAP	LYZ	MAF	BTG1
CD63	MGST1	CTSB	TGFB1
APOE	MSRB1	GAS6	SRGN
C1QC	IFITM2	CCL7	NFKBIA
GRN	TYROBP	ITM2B	CXCL2
FABP5	FAU	LGMN	CLEC4E

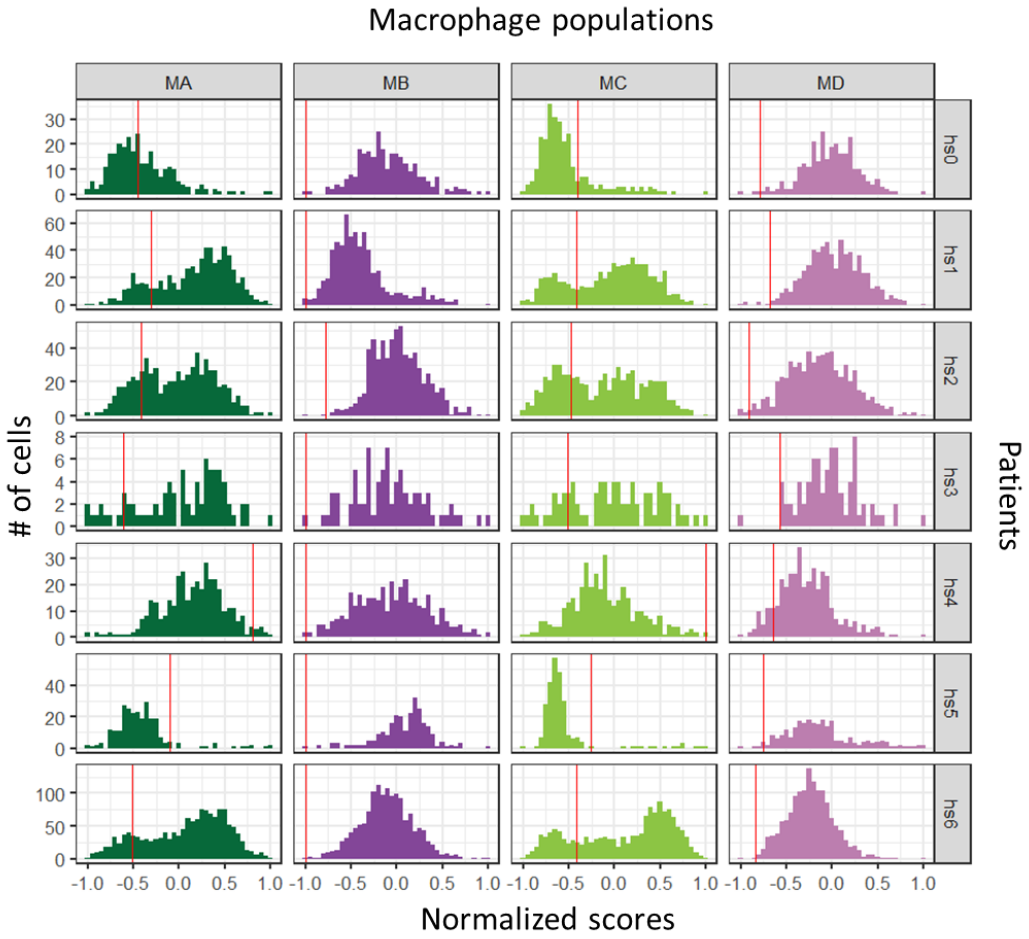


Figure 2.2.1 Histogram of normalized gene module scores across patients for the four macrophage populations. The red lines represent the minimum scoring cutoffs for annotating a cell.

To assess the association between macrophage subset proportions and RA severity, we examined multiple RA clinical parameters, including clinical disease activity index (CDAI), swollen joint count (SJC), tender joint count (TJC), and Disease Activity Score-28 with C-reactive protein (DAS28-CRP). Shannon index was computed to summarize the proportional diversity of macrophage subsets. Using available data from follow up clinic visits (6 week and > 12 week after

initial visit), patients were categorized as either low severity (HS6), responsive to treatment (HS0, HS1, HS5), non-responsive to treatment (HS3, HS4), or no follow-up (HS2) based on the trends of CDAI over the visits. Differentially expressed genes between responsive and non-responsive patients were obtained by down sampling each patient to 80 cells to minimize bias stemming from individual patients with high cell numbers, and then performing Seurat Wilcoxon signed-rank test using FindMarkers function with their averaged expression profiles. GO analysis was executed with GOrilla with the DE genes as input and expressed genes (same definition as above) as background.

2.2.2 Results

Single cell RNA-seq reveals 4 populations of synovial macrophages in mice.

To investigate the heterogeneity of mononuclear cells in murine synovium, we performed single-cell RNA-sequencing (scRNA-seq) on sorted CD45+CD11b+Ly6G-SiglecF- cells from ankle joints of healthy mice. A total of 6444 cells passed quality thresholds and 11 cell clusters were identified (Figure 2.2.2A, S2.2.2A-B). Clusters 0-7 were determined to be monocytes or macrophages, encompassing 98% of all profiled cells using canonical markers and SingleR algorithm, which assigns cell identities based on correlation with Immgen reference data (Figure S2.2.2C-D). Cluster 8-10 were assigned identities of fibroblasts, neutrophils, and migratory DCs based on documented gene markers, respectively (Figure S2.2.2E). These data confirm that the majority of the single cells profiled in this experiment are monocytes/macrophages.

At least two subpopulations of synovial macrophages of differing ontogenic origins, embryonic and monocytic, were previously documented [218]. In agreement, we also observed

two distinct groups of clusters with correlated gene expression signatures (Figure 2.2.2B). A closer look at individual differential markers in turn revealed four major expression patterns across the clusters. Clusters 0-3 specifically expressed markers associated with tissue residency and homeostasis, including Pf4, Cd163, Lyve1 and Sepp1 [8, 51, 168]. Conversely, clusters 4-7 preferentially expressed antigen presentation genes such as H2-Eb1 and Cd74, as well as those associated with arthritic inflammation, including Il1b, Ccl17, and Tnip3 [247-249]. Within the two correlated cluster groups, cluster 3, in addition to upregulating tissue resident genes, also had high expression of genes implicated in the macrophage population that forms the synovial lining, including Cx3cr1, Vsig4, Pmepa1, and Sparc [219, 220]. On the other hand, cluster 7 expressed higher number of transcripts typical of classical monocytes in addition to antigen presentation genes, including Ly6c2, Plac8, and Thbs1 [250, 251] (Figure 2.2.2C-D). Taken together, our data suggest four subgroups of synovial macrophages exists in the synovium based on distinct gene expression signatures. These subsets are initially named as MA (cluster 3), MB (cluster 7), MC (cluster 0-2), and MD (cluster 4-6) (Figure 2.2.2E). We then compared the expression signatures of the subsets to those reported in Culemann et al. [220], where six myeloid populations were defined (Figure S2.2.2F). We observed that MA subset is enriched for the signatures of Cx3cr1+ lining macrophages, while MC is similar to Retnla+ and Aqp1+ interstitial macrophages. MB and MD subsets are in turn enriched for the signatures of MHCII+ interstitial macrophages. The differing enrichment patterns of gene signatures from an independent dataset support the transcriptional and potentially functional distinctness of the defined subsets.

To explore the potential developmental relationships between the four defined subsets, we performed trajectory inference as implemented in Monocle2 package [241], where a potential cell

differentiation path was reconstructed by modelling the gradients of gene expression changes. We set MB as the starting point due to their expression of classical monocytic markers and inferred a trajectory that goes through MD cells, coinciding with the gradual downregulation of classical monocytic and infiltration markers (Figure 2.2.2F, S2.2.2G). These data implicate MD as a possible differentiation product from MB cells. A portion of MD exhibited heightened cell cycle gene signatures compared to other subtypes (Figure S1H), insinuating their potential expansion after differentiation. The trajectory then branched off, with one proceeding into MA and another ending in MC, suggesting that MA and MC subsets likely represent distinct terminal populations rather than being on the same differentiation path (Figure 2.2.2F). Differentially expressed genes along the trajectory were then clustered into three modules based on their relative expressions as the function of pseudotime, corresponding to the three observed branches (Figure S2.2.2G, I). In summary, we uncovered 4 distinct synovial macrophage subpopulations where two of the populations appear to be terminally differentiated, indicative of tissue resident cells using scRNA-seq.

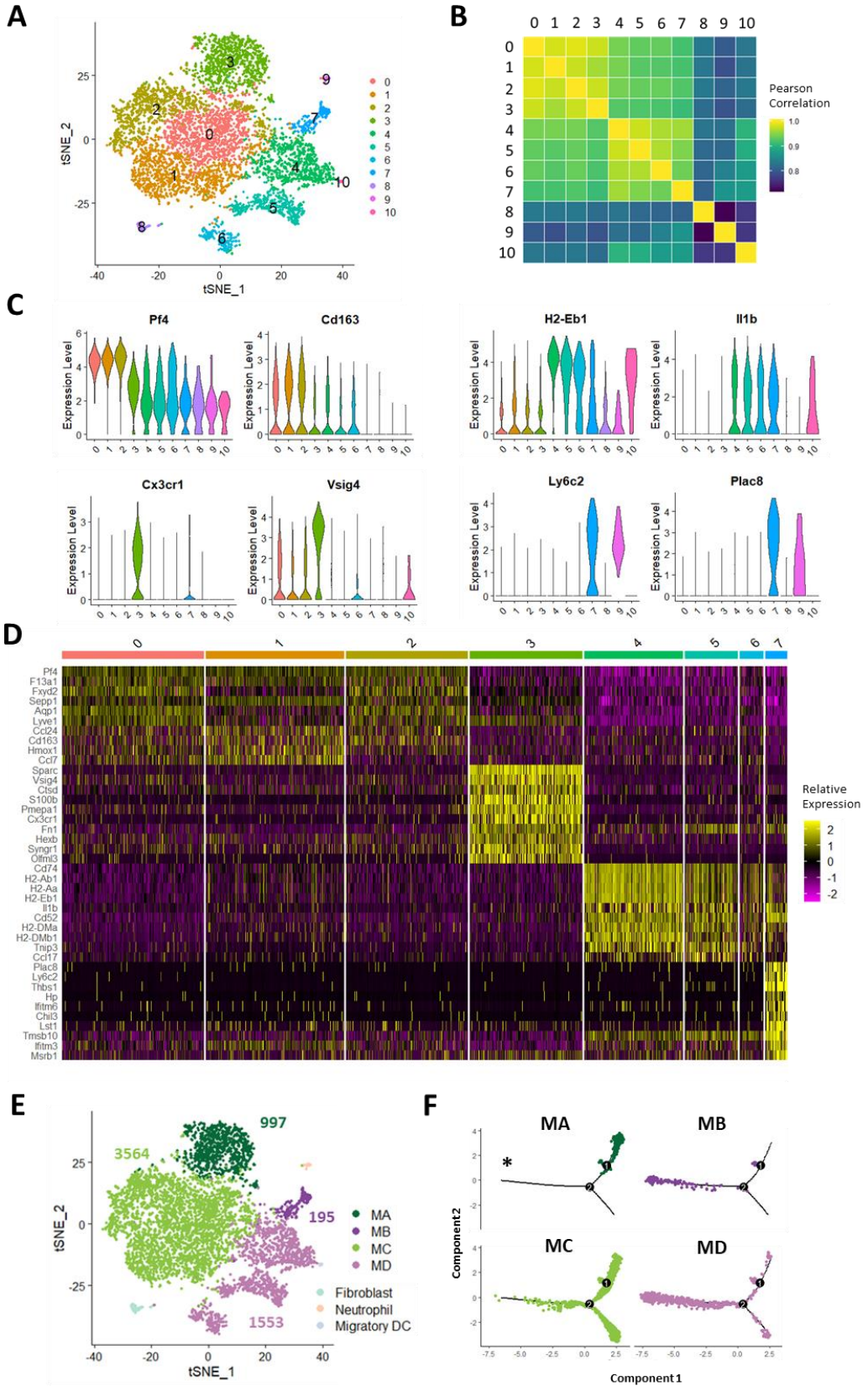

A
tSNE_2
tSNE_1
B
Pearson Correlation
C
Pf4
Cd163
H2-Eb1
Il1b
Cx3cr1
Vsig4
Ly6c2
Plac8
Expression Level
D
Relative Expression
E
997
3564
195
1553
MA
MB
MC
MD
Fibroblast
Neutrophil
Migratory DC
F
MA
MB
MC
MD
Component2
Component1

Figure 2.2.2 Single cell RNA-seq reveals 4 populations of synovial macrophages in mice.

(A) TSNE of 11 clusters from 6444 steady state CD45+CD11b+Ly6G-SiglecF- cells. (B) Correlation heatmap of averaged expression profiles for single cell clusters. (C) Violin plots representing log-normalized expression values of select cluster markers. (D) Heatmap of scaled expression values for top 10 marker genes ranked by foldchange from each single cell cluster. (E) TSNE visualization of defined synovial macrophage subsets. (F) Inferred pseudotime trajectory of single cells visualized on DDRTree dimensional reduced space, separated by subset annotations. The starting point of the trajectory is on the left.

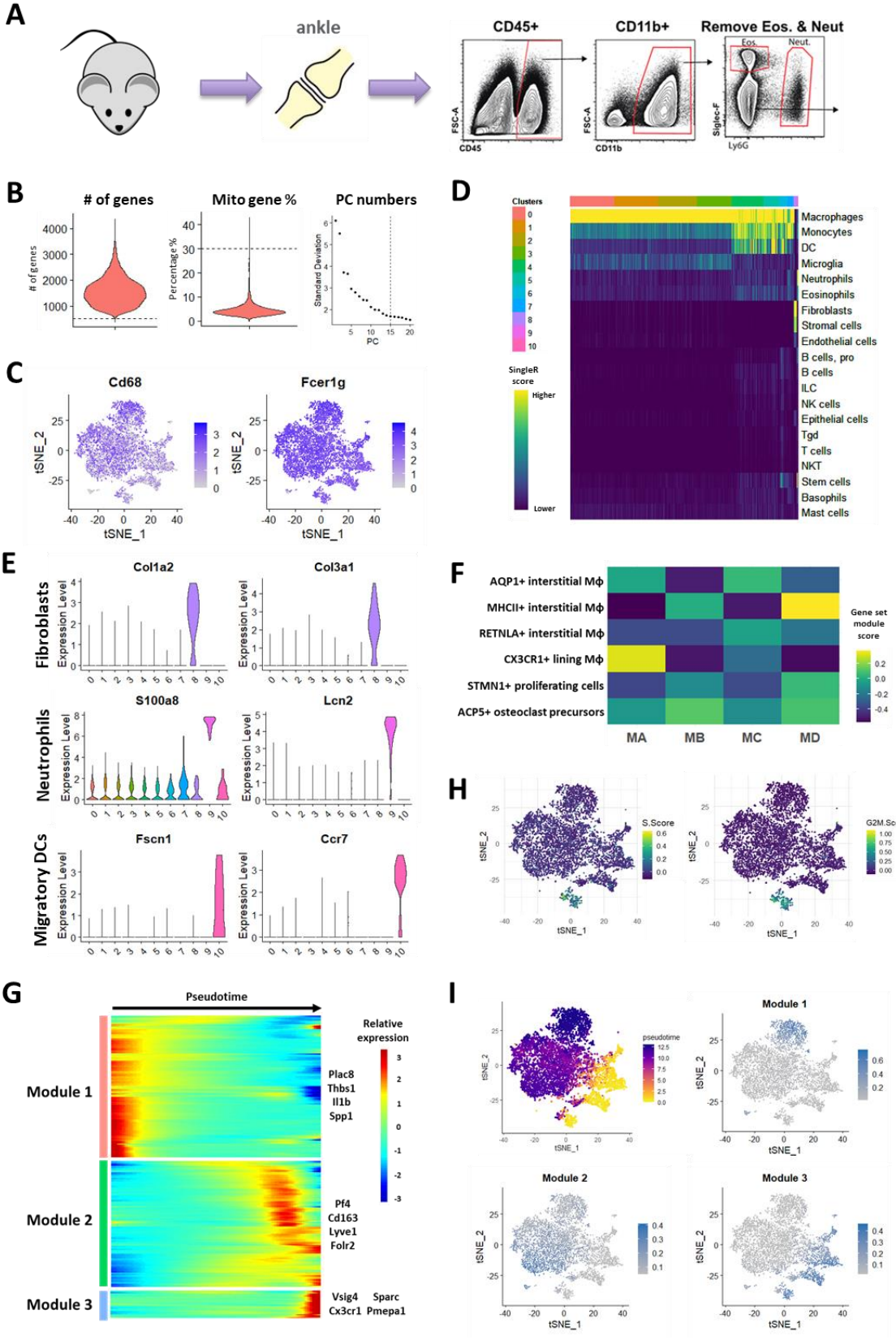

A
ankle
CD45+
CD11b+
Remove Eos. & Neut
Eos.
Neut.
FSC-A
CD45
CD11b
Siglec-F
Ly6G
B
# of genes
Mito gene %
PC numbers
Percentage %
Standard Deviation
PC
C
Cd68
Fcer1g
tSNE_2
tSNE_1
D
Clusters
SingleR score
Higher
Lower
Macrophages
Monocytes
DC
Microglia
Neutrophils
Eosinophils
Fibroblasts
Stromal cells
Endothelial cells
B cells, pro
B cells
ILC
NK cells
Epithelial cells
Tgd
T cells
NKT
Stem cells
Basophils
Mast cells
E
Col1a2
Col3a1
S100a8
Lcn2
Fscn1
Ccr7
Fibroblasts
Neutrophils
Migratory DCs
Expression Level
F
AQP1+ interstitial Mϕ
MHCII+ interstitial Mϕ
RETNLA+ interstitial Mϕ
CX3CR1+ lining Mϕ
STMN1+ proliferating cells
ACP5+ osteoclast precursors
MA
MB
MC
MD
Gene set module score
H
S.Score
G2M.Score
G
Pseudotime
Relative expression
Module 1
Module 2
Module 3
Plac8
Thbs1
Il1b
Spp1
Pf4
Cd163
Lyve1
Folr2
Vsig4
Cx3cr1
Sparc
Pmepa1
I
pseudotime
Module 1
Module 2
Module 3

Figure S2.2.2 (A) Schematic showing the collection and FACSorting of synovial myeloid cells. (B) Quality control cutoffs for the scRNA-seq experiment. The metrics used includes the number of expressed genes in each cell, % of mitochondrial reads in each cell, and number of principal components (PCs) for dimensionality reduction. (C) Expressions of select myeloid marker genes. (D) Heatmap illustrating SingleR assignment of Immgen reference cell types. (E) Expressions of select non-myeloid marker genes, including fibroblasts, neutrophils, and migratory DCs. (F) Mean expression module scores for gene signatures of myeloid subpopulations reported in Culemann et al. in MA-MD subsets. (G) Heatmap illustrating genes differentially expressed as a function of the trajectory. The genes are clustered into three gene modules based on their relative expression patterns along the trajectory. Select genes for each module are indicated on the right. (H) Cell cycle scoring of single cells using S phase (left) and G2/M phase genes (right). The scores are superimposed on tSNE plots. (I) Pseudotime values inferred by Monocle 2 (top left) and combined expression module scores for the three gene modules superimposed on tSNE plot.

Transcriptional profiles of 4 synovial macrophage populations supports differing functions and ontogeny

We next ask whether the four defined synovial macrophage subsets can be isolated definitively using antibodies to cell surface proteins. CITE-seq was performed on $CD45^{+}CD11b^{+}Ly6G^{-}SiglecF^{-}CD64^{+}Ly6C^{-}$ synovial cells. The surface expression pattern of CX3CR1 and MHCII identified the 4 macrophage subsets (Figure 2.2.3A, S2.2.3A), which was confirmed by flow cytometry (Figure 2B). Concordant relative proportions of cells were observed between the subsets defined by scRNA-seq and flow (Figure S2.2.3B), indicating that our sampling of single cells was representative of true myeloid heterogeneity in the joint. Given that

traditional bulk RNA-seq provides higher data resolution than scRNA-seq experiments, we performed bulk RNA-seq on the four sorted subsets to extensively characterize their transcriptional profiles. We confirmed that the 4 bulk subsets are transcriptionally distinct from each other (Figure S2.2.3C) and similar to their corresponding scRNA-seq subsets by assigning single cell identities with SingleR using the bulk profiles as reference (Figure 2.2.3C). The expression patterns of Cx3cr1 and H2-Eb1 (MHCII) were also consistent with cell surface expression pattern observed via CITE seq and by flow cytometry (Figure S2.2.3D). Collectively, these document the presence of four distinct macrophage subpopulations in the synovial tissue.

We then utilized unsupervised k-means clustering to identify differential gene expression patterns among the four macrophage subsets, along with FACsorted blood classical (CM) and non-classical (NCM) monocytes. Monocyte samples were included in the clustering to highlight gene signatures that are specific to individual synovial macrophage subpopulations. Previously, we were able to establish concordance between the MA-MD subsets and the subpopulations defined in Culemann et al [220] (Figure S2.2.2F). Similarly, here we defined six bulk gene expression clusters that defined MA (lining), MC (interstitial), MB and MD (monocyte-derived), and blood monocytes (pan-monocyte to NCM) (Figure 2.2.3D-E, S2.2.3D-E). The lining cluster (MA) includes genes linked to the formation of synovial lining, such as Vsig4 and Trem2 [220] and associated with GO processes including ossification and positive regulation of bone mineralization (Figure S2.2.3F). We further observed the enrichment of MEF2 TF binding motifs within the gene promoters of lining cluster, which are known to govern the development and differentiation of numerous cell lineages [252]. Mef2a, in particular, has been shown to promote terminal differentiation of macrophages [252, 253]. Mef2c has also been implicated in formation of bone and cartilage

structures [254], reflecting the cluster's enrichment of ossification processes (Figure 2.2.3F, S2.2.3H). The interstitial cluster (MC) consisted of previously reported key markers of macrophage homeostasis and localization, including Sepp1, Fxyd2, Mafb, and Aqp1 [9, 168, 220]. This cluster was enriched for the motifs of KLF TF family, which are well-characterized regulators of macrophage polarization [255]. Members of the KLF TFs specifically expressed in MC subset include Klf3, 4, and 9, which promote anti-inflammatory programs [256-258] while Klf3 promotes cartilage maintenance and repair [259]. On the other hand, the monocyte-derived cluster (MB and MD) consisted of leukocyte activation genes such as Cxcl1 and Spp1 [260, 261], while also expressing genes associated with antigen presentation (e.g., Cd74, H2-Eb1) (Figure S2.2.3F). This cluster exhibited enrichment of NFkB-p65 motif, a master regulator of innate inflammation processes through initiation of early response genes such as Tnf and Il1b after translocation into nucleus [262] (Figure 2.2.3, S2.2.3H). Notably, the gene signatures of the monocyte-derived cluster do not overlap with those of circulating monocytes (Figure 2D), with similar patterns observed in the correlations between blood monocytes and MB & MD subsets (Figure S2.2.3C). Furthermore, the motifs for IRF TF family, another major initiator of acute innate inflammation, were enriched specifically in classical monocytes (CM) but not monocyte-derived cluster (Figure S2.2.3H). NFKB and IRF TFs are reported to cross-regulate during inflammatory signaling owing to the presence of TF binding sites of both families within each other's downstream targets [263]. The distinct enrichment patterns for NFKB and IRF binding motifs suggest that MB and MD subsets may interact closely with classical monocytes in a complementary fashion when generating an immune response. These observations indicate that although MB and MD subsets express gene markers indicating monocytic origins, they are transcriptionally distinct from circulating monocytes and may possess synovial-specific functions.

To further understand how synovial macrophage subsets relate transcriptionally to resident macrophages from other tissues, we compared the k-means clusters with published gene sets from Lavin et al. [47] using hypergeometric tests implemented in MAGNET web tool (Figure S2.2.3G). Significant overlaps were observed between the lining cluster and microglia, the interstitial cluster and lung alveolar macrophages, and the monocyte-derived cluster and intestinal macrophages. Microglia and alveolar macrophages are of exclusively embryonic origins, whereas intestinal macrophages are derived predominately from bone marrow monocytes [29], indicating that synovial macrophage subsets may possess differing ontogenic histories. To more definitively examine their origins, we crossed a tamoxifen-inducible CX3CR1-Cre ($CX3CR1^{CreER}$) mouse with a GFP reporter mouse (zsGFP) to generate a mouse ($CX3CR1^{CreER}$.zsGFP) in which $CX3CR1^{+}$ cells express GFP after administration of tamoxifen (TMX). We then utilized $CX3CR1^{CreER}$.zsGFP mice for fate-mapping studies of synovial macrophages. TMX was delivered at E15 to pregnant mothers from the cross of $CX3CR1^{CreER}$ with zsGFP mice, which will allow for identification of embryonically derived cells that are not derived from the yolk sac. TMX-naïve $CX3CR1^{CreER}$.zsGFP mice provided control to estimate background levels of GFP. The MHC-synovial population regardless of the expression of Cx3Cr1 were positive for GFP (62 ±4.6% MHCII-CX3CR1+; 56 ±3.9% MHCII-CX3CR1-), while MHCII+Cx3CR1+ and MHCII+Cx3CR1- synovial macrophage were just above the negative control (Figure 2.2.3G). These data indicate the majority of MHCII- synovial macrophage population is derived from the fetal liver monocytes, consistent with a previous report [225].

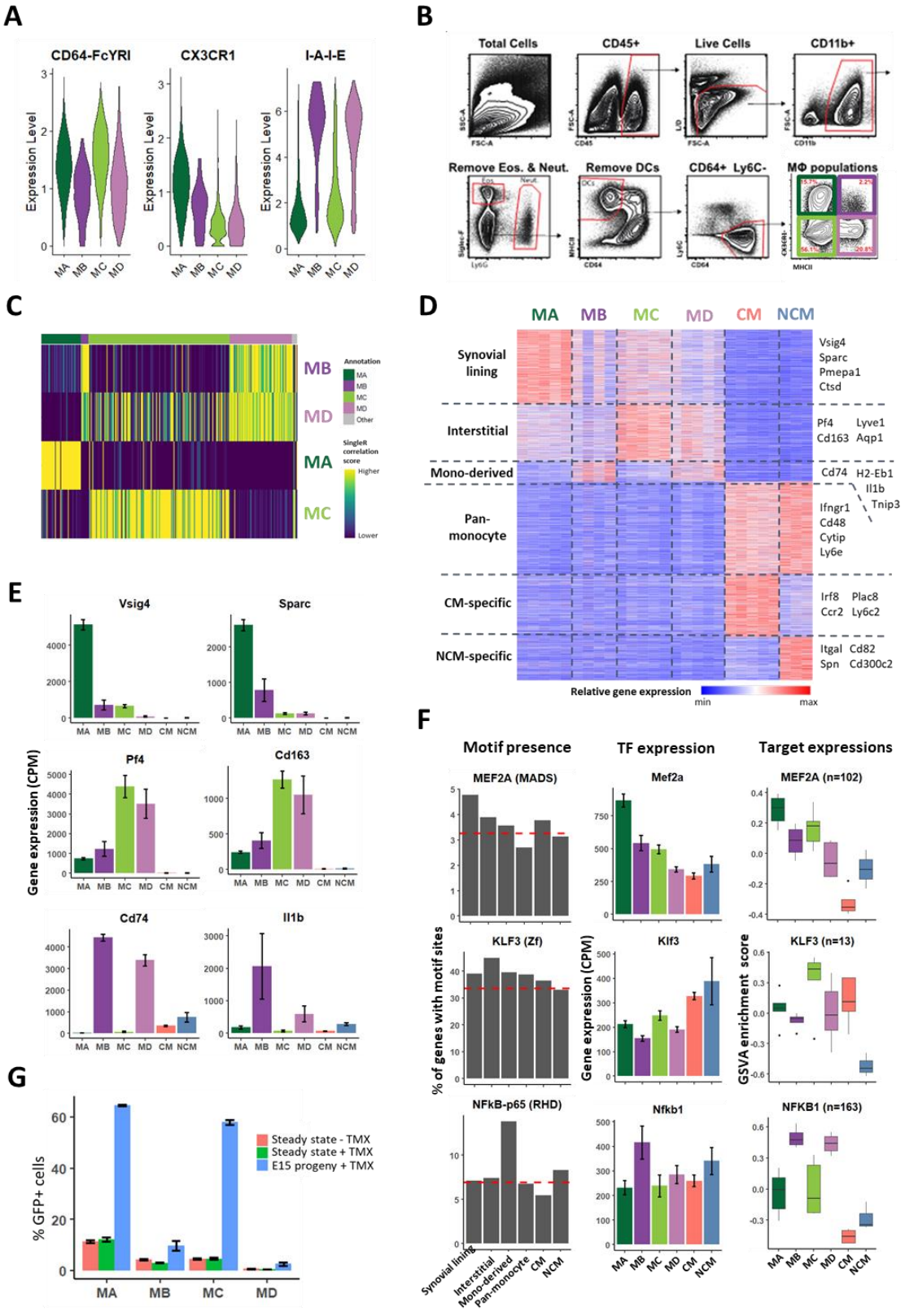

A
CD64-FcYRI
CX3CR1
I-A-I-E
Expression Level
MA MB MC MD
B
Total Cells
CD45+
Live Cells
CD11b+
Remove Eos. & Neut.
Remove DCs
CD64+ Ly6C-
MΦ populations
C
MB
MD
MA
MC
Annotation
MA
MB
MC
MD
Other
SingleR correlation score
Higher
Lower
D
MA MB MC MD CM NCM
Synovial lining
Vsig4
Sparc
Pmepa1
Ctsd
Interstitial
Pf4 Lyve1
Cd163 Aqp1
Mono-derived
Cd74 H2-Eb1
Il1b
Tnip3
Pan-monocyte
Ifngr1
Cd48
Cytip
Ly6e
CM-specific
Irf8 Plac8
Ccr2 Ly6c2
NCM-specific
Itgal Cd82
Spn Cd300c2
Relative gene expression
min
max
E
Vsig4
Sparc
Pf4
Cd163
Cd74
Il1b
Gene expression (CPM)
MA MB MC MD CM NCM
F
Motif presence
TF expression
Target expressions
MEF2A (MADS)
Mef2a
MEF2A (n=102)
KLF3 (Zf)
Klf3
KLF3 (n=13)
% of genes with motif sites
Gene expression (CPM)
GSVA enrichment score
NFkB-p65 (RHD)
Nfkb1
NFKB1 (n=163)
Synovial lining
Interstitial
Mono-derived
Pan-monocyte
CM
NCM
G
% GFP+ cells
Steady state - TMX
Steady state + TMX
E15 progeny + TMX
MA MB MC MD

Figure 2.2.3. Transcriptional profiles of 4 synovial macrophage populations supports differing functions and ontogeny

(A) Normalized single cell surface expressions of CD64 (CD64-FcgRI), CX3CR1, and MHCII (I-A-I-E), as measured by CITE-seq. (B) Flow cytometry classification of synovial macrophage subpopulations: CX3CR1+MHCII- (MA), CX3CR1+MHCII+ (MB), CX3CR1-MHCII- (MC), and CX3CR1-MHCII+ (MD). (C) Heatmap illustrating SingleR assignment of synovial macrophage subpopulations from bulk RNA-seq data to steady state single cell. (D) K-means clustering (K = 6) of 4525 differentially expressed genes among the four sorted synovial macrophage and two blood monocyte subsets. (E) Gene expression of select marker genes from synovial lining, interstitial, and mono-derived K-means clusters. (F) Percent presence of DNA binding motifs for select TFs in the promoters of K-mean cluster genes (left), gene expression of the TFs as measured by bulk RNA-seq (center), and GSVA-inferred expression scores for downstream target genes of the TFs from Dorothea database (right). (G) Percent of GFP+ cells across synovial macrophage subsets in E15 progeny mice with tamoxifen administered and control steady state mice with and without tamoxifen.

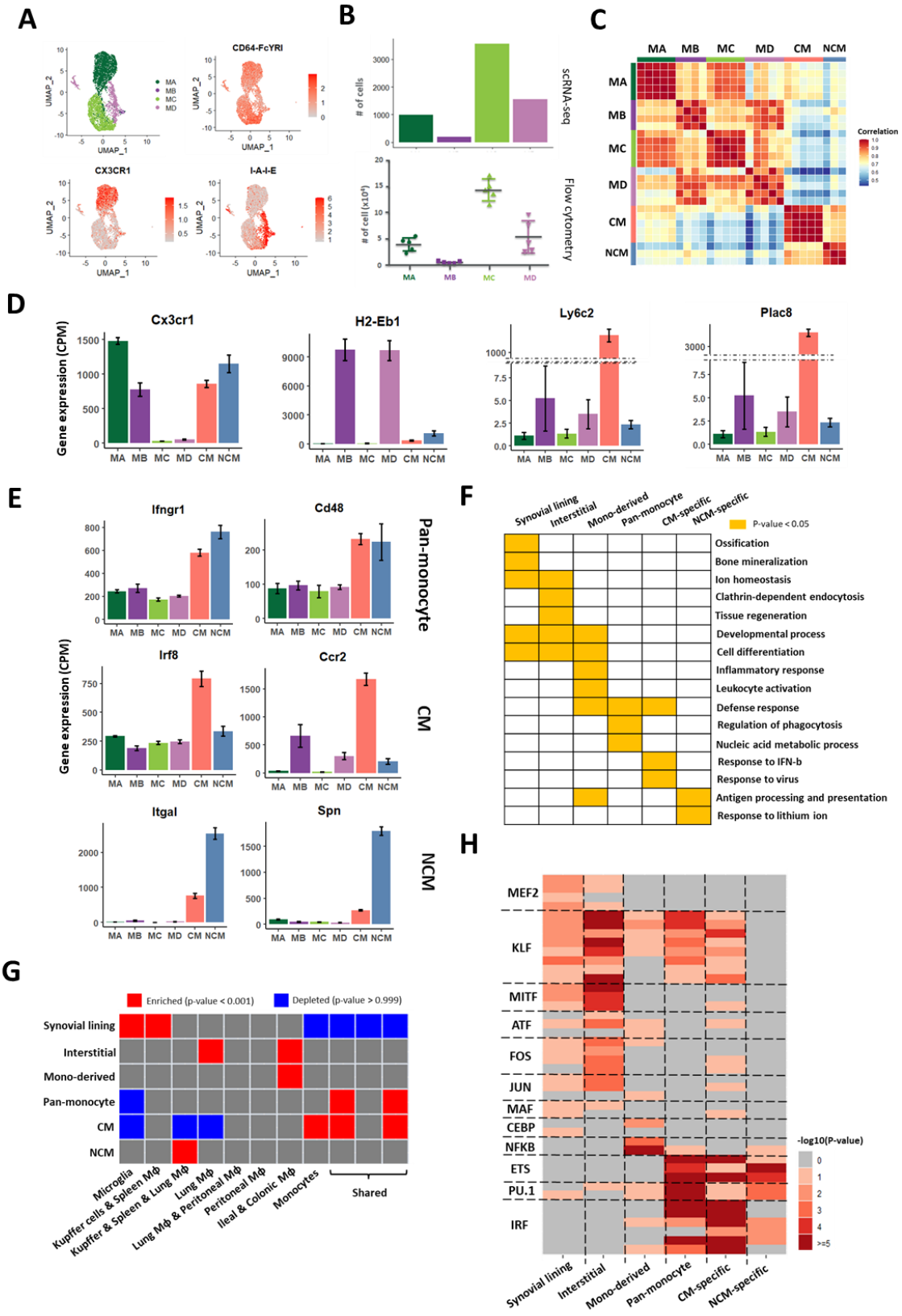

A
B
C
D
E
F
G
H
CD64-FcYRI
CX3CR1
I-A-I-E
UMAP_1
UMAP_2
MA
MB
MC
MD
# of cells
scRNA-seq
# of cell (x10⁴)
Flow cytometry
Correlation
CM
NCM
Cx3cr1
H2-Eb1
Ly6c2
Plac8
Gene expression (CPM)
Ifngr1
Cd48
Irf8
Ccr2
Itgal
Spn
Pan-monocyte
Synovial lining
Interstitial
Mono-derived
Pan-monocyte
CM-specific
NCM-specific
P-value < 0.05
Ossification
Bone mineralization
Ion homeostasis
Clathrin-dependent endocytosis
Tissue regeneration
Developmental process
Cell differentiation
Inflammatory response
Leukocyte activation
Defense response
Regulation of phagocytosis
Nucleic acid metabolic process
Response to IFN-b
Response to virus
Antigen processing and presentation
Response to lithium ion
Enriched (p-value < 0.001)
Depleted (p-value > 0.999)
Microglia
Kupffer cells & Spleen Mϕ
Kupffer & Spleen & Lung Mϕ
Lung Mϕ
Lung Mϕ & Peritoneal Mϕ
Peritoneal Mϕ
Ileal & Colonic Mϕ
Monocytes
Shared
MEF2
KLF
MITF
ATF
FOS
JUN
MAF
CEBP
NFKB
ETS
PU.1
IRF
-log10(P-value)

Figure S2.2.3 (A) Annotations of the four synovial macrophage subsets in the CITE-seq experiment and surface expressions of CD64, CX3CR1, and I-A-I-E visualized in UMAP. (B) Quantification of the 4 subsets in scRNA-seq (top) and flow cytometry (bottom). (C) Pairwise correlations of gene expression profiles between replicates as measured by bulk RNA-seq data. (D) Gene expression of Cx3cr1, H2-Eb1, Ly6c2, and Plac8. (E) Gene expression of select marker genes from pan-monocyte, classical monocyte, and non-classical monocyte K-means clusters. (F) Select gene ontology (GO) pathways significantly enriched across K-means cluster. (G) Comparison of k-means gene clusters to gene signatures of tissue resident macrophages from Lavin et al. 2014. Red cells indicate hypergeometric p-value < 0.05 whereas blue cells indicate p-value > 0.95. (H) Significance of DNA binding motif enrichments for select TF families across k-means clusters. The color scale represents -log10(p-value) as computed by HOMER.

Acute inflammatory arthritis alters the transcriptional landscapes of synovial macrophages subsets

Infiltration of circulating monocytes and their subsequent differentiation into activated macrophages is a hallmark of inflammation [68]. To investigate how the four subsets of synovial macrophages are altered during acute inflammatory arthritis, we utilized the mouse model of serum transfer induced arthritis (STIA), which represents the effector phase and requires monocytes and macrophages [264]. The clinical arthritic scores peaked at day 7 and largely returned to steady state levels by day 21 after the induction of arthritis (Figure 2.2.4A) consistent with previous studies [225, 265, 266]. Similarly, the height of synovial macrophage subset expansion occurred at the nadir of inflammatory arthritis development i.e., day 7 (Figure 2.2.4B-C). We then performed bulk RNA-seq on the sorted synovial macrophage subsets at day 7, 13 and 21 post serum transfer

to distinguish the functions of synovial macrophage subsets in response to STIA. All macrophage subpopulations were transcriptionally more dissimilar at day 7 and 13 compared with day 0, while macrophages from day 0 compared to day 21 were similar based on Principal Component Analysis (PCA) (Figure S2.2.4A). MB exhibited the largest number of differentially expressed genes relative to day 0 for all subsequent timepoints (Figure S2.2.4B), suggesting major contributions to the pathogenesis of STIA from this particular subset.

We defined 1772 genes that were differentially expressed between day 0 and any subsequent time point in at least one macrophage subset. We obtained 6 clusters of temporal gene expression patterns using unsupervised K-means clustering (Figure 2.2.4D-E). The first three clusters encompassed genes upregulated over the course of STIA. Comparison to steady state gene clusters (Figure 2.2.4D) revealed that genes preferentially expressed in blood monocytes (pan-monocyte, CM, and NCM) were elevated across all subsets during STIA (Figure S2.2.4D) and exhibited strong overlap with these 3 clusters (Figure S2.2.4C). Cluster Up I included genes upregulated in all four subsets that peaked at days 7 and 13 post-serum transfer. These genes were associated with leukocyte activation and cell chemotaxis (e.g., Thbs1, Ccr2, Il1b, Malt1), further supporting that all subpopulations partake in the promotion of inflammatory reactions and recruitment of immune cells [234]. Cluster Up I also contained genes involved in collagen catabolic and tissue remodeling pathways (e.g., Mmp9, Mmp14, Ctsk, Spp1), and were disproportionately elevated in MC and MD subsets (Figure 2.2.4F). This specific set of genes have been associated with RA-related cartilage destruction [267-269]. Cluster Up II consisted of upregulated genes peaking at day 3 and were the most pronounced in MB and MD subsets. These genes (e.g., Ifitm3, Isg15, Rsad2, Ifit1, Mnda) are commonly associated with acute inflammatory processes, such as production and response to Type

I interferons. Early activation of Type I interferon pathways have been observed in rheumatic patients and is murine models of arthritis [270-272]. Motif enrichment analysis on this cluster revealed significant enrichments for the IRF TF binding motif, which was previously attributed to classical monocytes in steady state (Figure S2.2.3H, S2.2.4E). Cluster Up III also included genes upregulated specifically in MB cells, but remain elevated throughout 21 days after serum transfer. Genes associated with adaptive immune response are overrepresented in the cluster, including Ctsc [273], Cd48 [274, 275], and Bach2 [276, 277], insinuating a potential MB-specific role in facilitating the activation of CD4+ T cell subsets during RA [278]. This cluster also included additional genes linked to RA pathogenesis and susceptibility, such as Itgal [279], Cd82 [280], and Irak4 [281]. Given that STIA represented an acute model of arthritis, this observation raises the question of whether MB subset specifically remained altered transcriptionally after subsidence of inflammation. To investigate, we computed and tabulated the changes in expression at day 7 and 21 relative to day 0 to distinguish genes involved in acute inflammation and those that remained altered after resolution of STIA (Figure S2.2.4G). We found that compared to the other three subsets, MB possessed the highest number of genes that were elevated at both day 7 and 21 post serum-transfer (Figure S2.2.4H). In summary, these findings indicate that while all four macrophage subsets exhibited evidence of monocytic recruitment on the gene expression level, biological processes and individual gene markers associated with inflammatory arthritis were upregulated unevenly across the subsets during the time course, suggesting divergent roles in STIA pathogenesis.

Clusters Down I-III encompassed genes downregulated in response to STIA (Figure 2.2.4D). Cluster Down I consisted of genes downregulated specifically in MC and MD subsets and were

enriched for multiple signal transduction pathways, such as MAPK cascade and cAMP-mediated signaling (e.g., Atf3, Jun, Lpar1, Dusp6) (Figure 2.2.4D-E, Figure S2.2.4D). While previous studies have shown that both MAPK [282] and cAMP [283] pathways are involved in the pathogenesis of RA, decreased expression of these genes may not be associated with reduced activity at the protein level. Conversely, a recent mouse study demonstrated that impaired differentiation of synovial macrophages contributes directly to the development of inflammatory arthritis by opening a niche that permits influx of activated monocytes [284]. Thus, the temporary downregulation of MAPK- and cAMP-associated genes during STIA perhaps reflects this phenomenon as both pathways are central participants in macrophage differentiation processes [285, 286]. Cluster Down II included genes associated with cell cycle processes (e.g., Mki67, Ccnd2, Top2a, Kif23, and Cdkn2d) and are downregulated specifically in MB subset. Inflammatory responses are known to suppress proliferation in macrophages by shifting their metabolic programs from a Myc-dependent to Hif1a-dependent manner [287]. Given that Myc is a downstream target of MAPK pathway [288, 289], the suppression of cell cycle processes is consistent with the observations in Cluster Down I. Genes in Cluster Down III are in turn primarily downregulated in MA cells and included those previously reported in lining macrophages (e.g., Sparc, Pmepa1) [219, 220], as well as those associated with ossification (e.g., Gas6, Bmp2) [290, 291] and cell adhesion (e.g., Lyve1, Csf1r, Axl) [292-294] GO processes. Comparison to steady state gene clusters (Figure 2.2.3F) revealed the overall downregulation of lining and interstitial genes across all subsets. The lining signature disproportionately overlapped with Cluster VI whereas the interstitial signature overlapped more with Cluster IV, supporting the notion that the tissue-resident MA and MC subsets temporarily lost their distinct phenotypes over the course of STIA (Figure S2.2.4C). This is further corroborated by the overall downregulation of downstream

target genes for MEF2A and KLF3 (Figure S2.2.4F), which were previously attributed to synovial resident subsets during steady state. Collectively, these data suggest that synovial macrophage subsets during STIA underwent transcriptional changes indicative of monocyte influx from circulation and diminishment of their steady state identities.

We next utilized the CX3CR1CreER.zsGFP reporting system to definitively assess the contribution of circulating monocytes during STIA (Figure S2.2.4I). TMX was administered at Day -1 and Day 0 of STIA time course to adult CX3CR1CreER.zsGFP mice, which would specifically label CX3CR1-expressing already present in the synovium (MA and MB subsets). In healthy mice that serve as negative control (no STIA TMX D-1 & D0), all subsets exhibited subtle increases in GFP positivity over a 21-day time course (MA: 32.85% MB: 23.57% MC: 18.19% MD:14.11%), indicating local proliferation and minimal contribution from circulation, consistent ith previous reports [220, 225]. In contrast, the mice that underwent serum-transfer (STIA TMX D-1 & D0) displayed large decreases in GFP positivity for the two CX3CR1+ subsets (MA: -51.00% MB: -41.62%), suggesting significant cell replacements by circulation in at least these populations. To better quantify the monocytic contributions across the four macrophage compartments, we performed an additional experiment by administering TMX at Day 3 after serum-transfer, which would label cells that recently entered the synovium instead of those already residing within. All four subsets exhibited increases in GFP positivity over the 21-day time course (MA: 23.08% MB: 30.74% MC: 23.14% MD: 32.78%). These results from fate-mapping experiments suggest that circulating monocytes are key contributors to all four compartments of synovial macrophages during inflammatory arthritis.

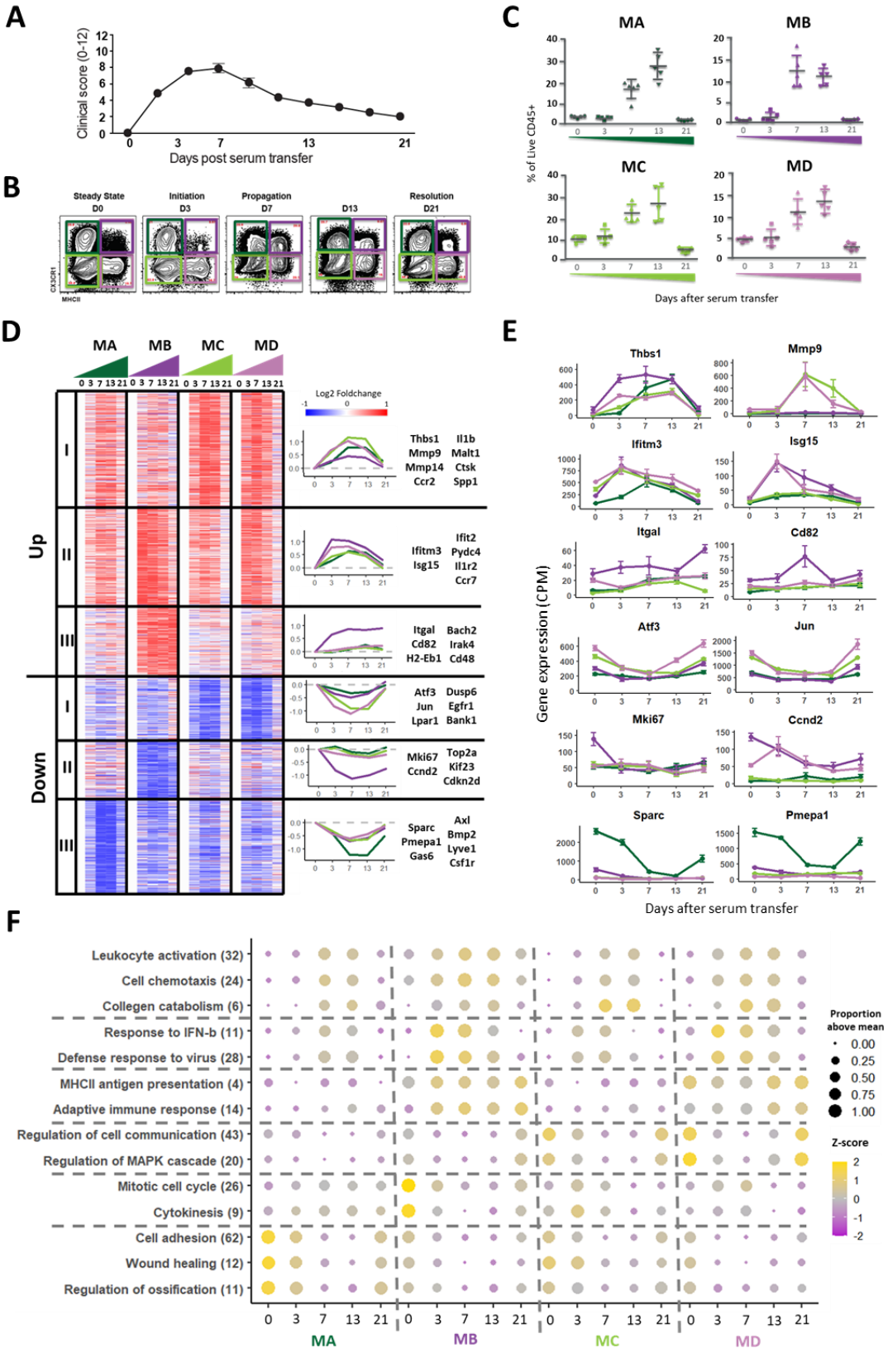

A
Clinical score (0-12)
Days post serum transfer
B
Steady State D0
Initiation D3
Propagation D7
D13
Resolution D21
CX3CR1
MHCII
C
MA
MB
MC
MD
% of Live CD45+
Days after serum transfer
D
MA MB MC MD
Log2 Foldchange
Up
Down
I
II
III
Thbs1 Il1b Mmp9 Malt1 Mmp14 Ctsk Ccr2 Spp1
Ifitm3 Isg15 Ifit2 Pydc4 Il1r2 Ccr7
Itgal Cd82 H2-Eb1 Bach2 Irak4 Cd48
Atf3 Jun Lpar1 Dusp6 Egfr1 Bank1
Mki67 Ccnd2 Top2a Kif23 Cdkn2d
Sparc Pmepa1 Gas6 Axl Bmp2 Lyve1 Csf1r
E
Thbs1
Mmp9
Ifitm3
Isg15
Itgal
Cd82
Atf3
Jun
Mki67
Ccnd2
Sparc
Pmepa1
Gene expression (CPM)
Days after serum transfer
F
Leukocyte activation (32)
Cell chemotaxis (24)
Collegen catabolism (6)
Response to IFN-b (11)
Defense response to virus (28)
MHCII antigen presentation (4)
Adaptive immune response (14)
Regulation of cell communication (43)
Regulation of MAPK cascade (20)
Mitotic cell cycle (26)
Cytokinesis (9)
Cell adhesion (62)
Wound healing (12)
Regulation of ossification (11)
MA
MB
MC
MD
Proportion above mean
Z-score

Figure 2.2.4. Acute inflammatory arthritis alters the transcriptional landscapes of synovial macrophages subsets

(A) Clinical arthritic scores over the course of STIA. (B-C) Proportions of synovial macrophages subsets over time. (D) K-means clustering (K = 6) of 1772 differentially expressed genes across the STIA time course, visualized as fold-changes between the mean expressions of each timepoint and day 0 expressions. Expression trend lines are indicated on the right. (E) Gene expressions of representative genes for each K-mean cluster. (F) Relative gene expressions of select Gene Ontology processes across macrophage subsets and STIA time course. Size of circles indicate proportion of genes expressed above the average of all samples while color represents mean expression Z-scores across all samples.

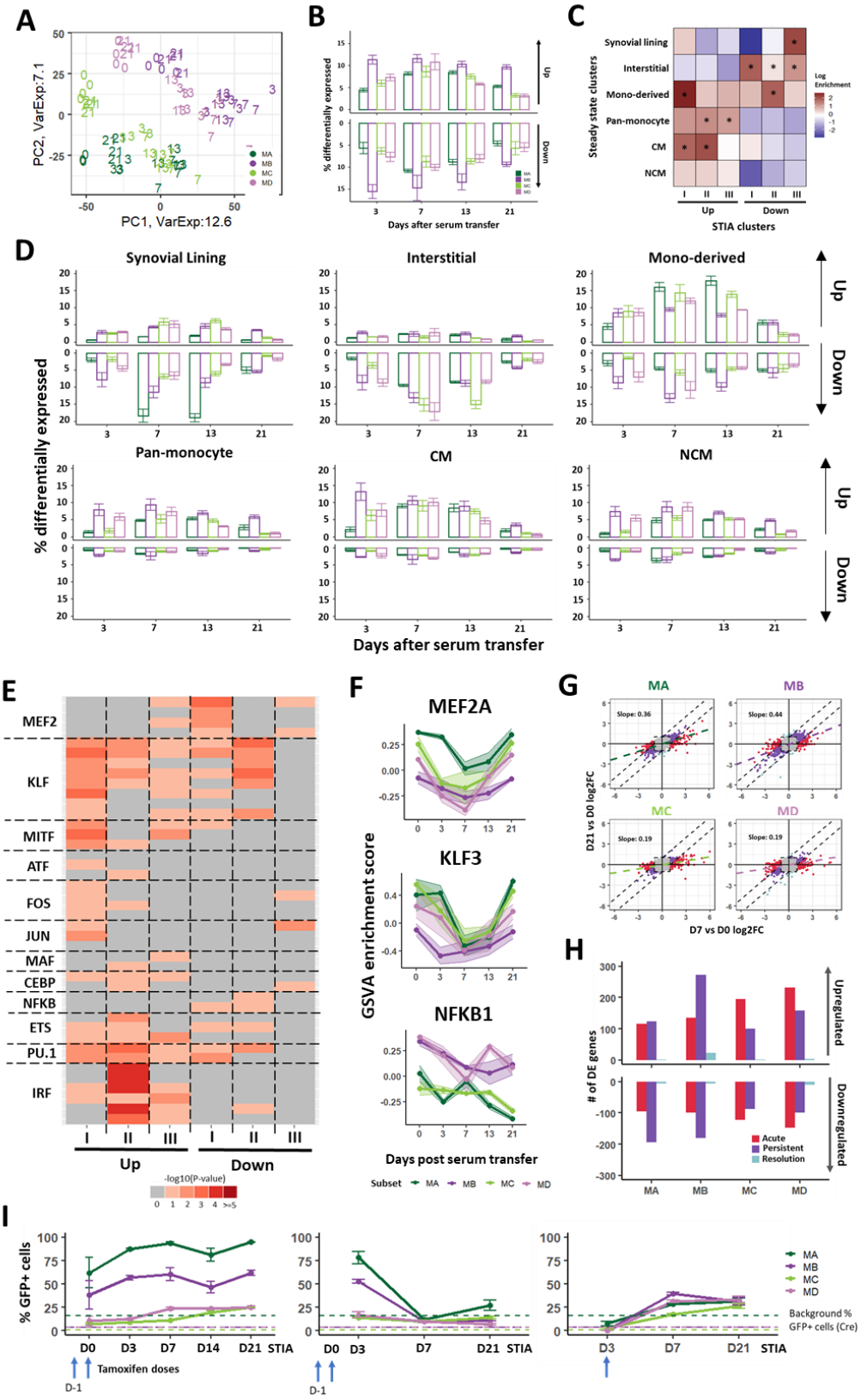

A
PC2, VarExp:7.1
PC1, VarExp:12.6
MA
MB
MC
MD
B
% differentially expressed
Up
Down
Days after serum transfer
C
Synovial lining
Interstitial
Mono-derived
Pan-monocyte
CM
NCM
Steady state clusters
Log Enrichment
Up
Down
STIA clusters
D
Synovial Lining
Interstitial
Mono-derived
Pan-monocyte
CM
NCM
% differentially expressed
Days after serum transfer
E
MEF2
KLF
MITF
ATF
FOS
JUN
MAF
CEBP
NFKB
ETS
PU.1
IRF
Up
Down
-log10(P-value)
F
MEF2A
KLF3
NFKB1
GSVA enrichment score
Days post serum transfer
Subset MA MB MC MD
G
MA
MB
MC
MD
Slope: 0.36
Slope: 0.44
Slope: 0.19
Slope: 0.19
D21 vs D0 log2FC
D7 vs D0 log2FC
H
# of DE genes
Upregulated
Downregulated
Acute
Persistent
Resolution
MA
MB
MC
MD
I
% GFP+ cells
D0
D3
D7
D14
D21
STIA
Tamoxifen doses
D-1
MA
MB
MC
MD
Background % GFP+ cells (Cre)

Figure S2.2.4 (A) PCA of gene expression profiles between replicates from STIA time course. Text indicates number of days after serum transfer. (B) Quantification of up- and down-regulated genes (2-fold) for each time point relative to day 0 of STIA. (C) Log enrichment of gene overlap between steady state and STIA K-means clusters. (D) Percent of steady state K-mean cluster genes (Figure 2D) expressed differentially (2-fold relative to day 0) across the STIA time course and synovial macrophage subsets. (E) Significance of DNA binding motif enrichments for select TF families across STIA k-means clusters. The color scale represents -log10(p-value) as computed by HOMER. (F) GSVA-inferred expression scores for downstream target genes of MEF2A, KLF3, and NFKB-p65 from Dorothea database across macrophage subsets and STIA time course. (G) Scatter plots of expression fold changes relative to Day 0 STIA for Day 7 and Day 21. (H) Quantification of acute (Day 7 only; red), persistent (Days 7 & 21; purple), and resolution (Day 21 only; light blue). (I) Percent of GFP+ cells across macrophage subsets and STIA time course with different experimental conditions: healthy mice with tamoxifen doses at Day 0 and Day 0-1 (left), STIA mice with tamoxifen doses at Day 0 and Day 0-1 (middle), and STIA mice with a tamoxifen dose at Day 3 of the time course. The background % of GFP expressions (Cre) are indicated in dashed lines.

The induction of STIA give rise to macrophages with specialized transcriptional profiles that extend beyond their steady state definitions

We previously identified 4 populations of synovial macrophages via scRNA seq at steady state. We then isolated and performed scRNA-seq on synovial $CD45^{+}CD11b^{+}Ly6G^{-}SiglecF^{-}$ cells from mice 7 days after K/BxN serum transfer to confirm the identity of these populations during

inflammatory arthritis. A total of 10 clusters consisting of 3916 cells were identified (Figure S2.2.5A-C) with the majority of them confirmed as monocytes/macrophages via singleR (Figure S2.2.5D). we annotated the identities of Day 7 STIA macrophages using the Seurat label transfer algorithm with the steady state data as reference [199] to assess the changes in transcriptional composition of synovial macrophage subsets in response to STIA. The relative proportions of annotated MB cells expanded during STIA compared to steady state (3.09% to 64.91%), consistent with the observed patterns using flow cytometry. On the other hand, the tissue resident subsets (MA and MC), exhibited a marked reduction in their proportion at day 7 STIA (MA: 15.80% to 0.78%; MC: 56.49% to 19.04%) (Figure 2.2.5A-B). Macrophages from Day 7 STIA displayed a higher correlation with monocyte-derived MB and MD compared to tissue-resident MA and MC subsets from steady state, potentially reflecting the influx of circulating monocytes (Figure S2.2.5E). This is corroborated by the expansions of previously defined blood monocyte and monocyte-derived signatures (Figure 2.2.3D) and the near-complete loss of lining and interstitial signatures in single cells from Day 7 STIA compared to steady state (Figure S2.2.5F-H). We then investigated the expressions of pseudotime gene modules identified previously (Figure S2.2.2I) in single cells from both steady state and Day 7 STIA samples to confirm the relative expansion of monocyte-derived gene signatures from another angle. The expression scores of the three modules were distinctly enriched in steady state MB/MD, MC, and MA subsets (Figure 2.2.5C), respectively, which are consistent with their inferred locations on the trajectory (Figure 2.2.2F). In contrast, the combined expression of Middle and Late modules were relatively diminished in macrophages from Day 7 STIA, whereas Early module expression remained specifically elevated in MB and MD cells (Figure 2.2.5D). These data suggest that steady state differentiation process for synovial macrophages may be disrupted or delayed during inflammatory arthritis.

We next investigated whether the four annotated macrophage subsets under STIA conditions retained a distinct transcriptional profile from steady state despite the decrease in heterogeneity. The gene expression pattern of Cx3cr1 and MHCII (H2-Eb1) was retained among the four subsets, with the exception of MC population, which exhibited elevated MHCII gene expression at day 7 STIA compared to steady state (Figure 2.2.5E). All four macrophage subsets also preferentially expressed their characteristic gene markers as identified during steady state, albeit the markers for monocyte-derived MB and MD subsets are also upregulated across the tissue-resident MA and MC cells (Figure 2.2.5F). These observations suggest that synovial macrophage subsets at D7 STIA remained transcriptional distinct from each other, but with decreased heterogeneity compared to their steady state counterparts.

The analysis of the STIA time course using bulk RNA-seq indicated that STIA-induced transcriptional responses maybe subset-specific. We examined the expression of K-means clusters obtained with bulk RNA-seq at Day 7 STIA (Figure 2.2.4D) with sc-RNA seq at day 7 to gain a clearer picture on the single cell level (Figure 2.2.5G, S2.2.5H). The cluster expression signatures were not enriched evenly in single cells, even within the same annotated macrophage subset (Figure 2.2.5H, S2.2.5F-G). Specifically, STIA Up I was predominately upregulated in MC/MD subsets and was associated with leukocyte activation and collagen catabolic processes. While genes linked to leukocyte activation (e.g., Spp1, Fn1, Il1b) exhibited relatively uniform expression across the MD cells at Day 7 STIA, those associated with collagen catabolism (e.g., Mmp9, Acp5, Ctsk) were upregulated only in a small subset of cells annotated as MC. Likewise, STIA Up II cluster, which was upregulated specifically in MB subset and attributed to Type I interferon

response, was also enriched only in a subset of labeled MB cells at Day 7 STIA. STIA Up III cluster was another MB-specific cluster containing genes associated with antigen presentation and adaptive immune response. Antigen presentation genes (e.g., H2-Eb1, H2-Aa) was expressed highly but unevenly across all Day 7 STIA macrophage subsets, even among MHCII-positive MB and MD subsets defined at steady state. Conversely, multiple adaptive immune response genes (e.g., Ctsc, Anxa1, Cd48) exhibited relatively uniform enrichments across Day 7 STIA macrophages. These findings imply that despite the overall decrease of transcriptional heterogeneity due to monocyte recruitment, the induction of STIA potentially give rise to macrophages with specialized transcriptional profiles that extend beyond their steady state definitions.

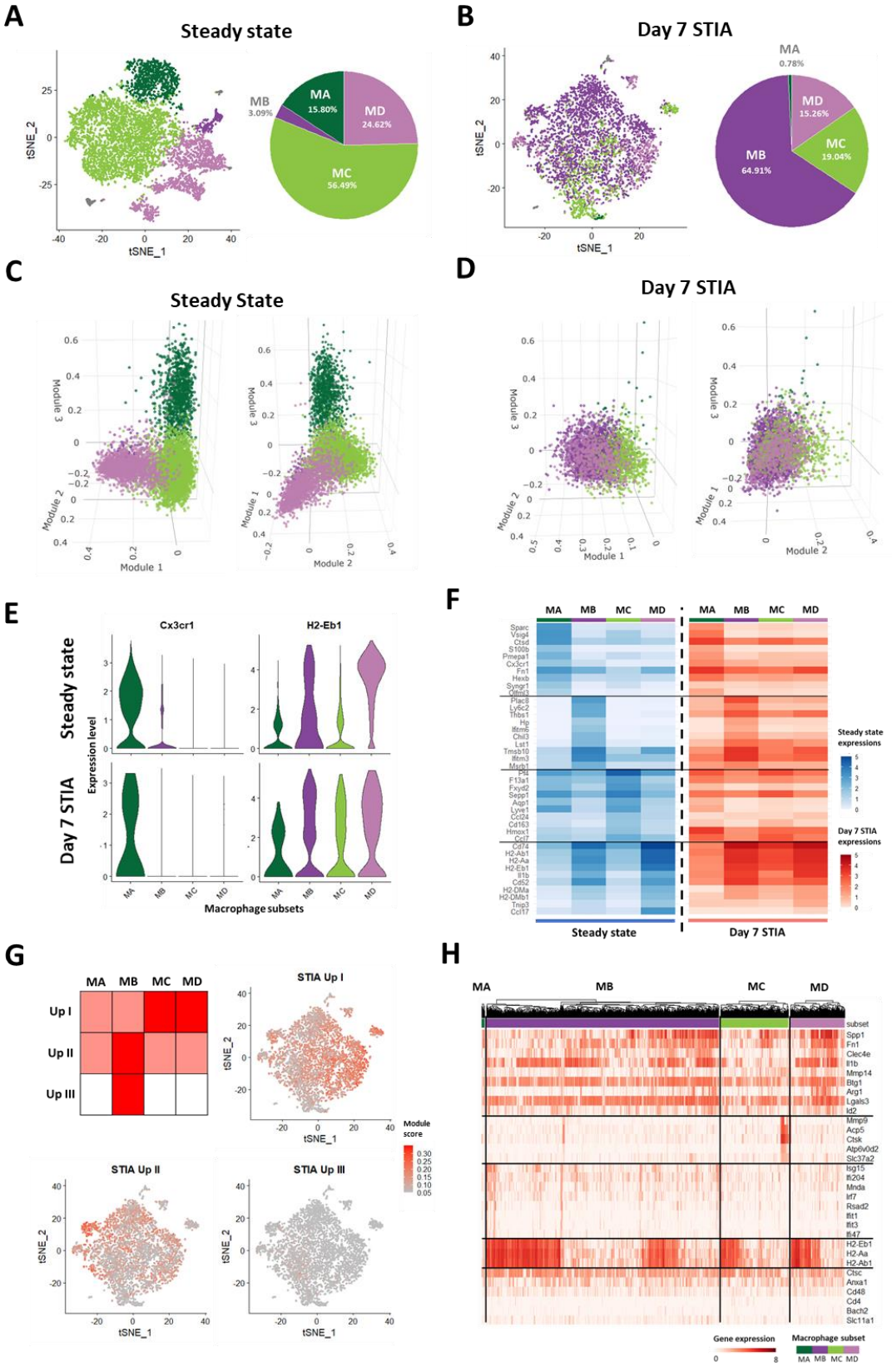

A
Steady state
tSNE_2
tSNE_1
MB
3.09%
MA
15.80%
MD
24.62%
MC
56.49%
B
Day 7 STIA
MA
0.78%
MD
15.26%
MC
19.04%
MB
64.91%
C
Steady State
Module 1
Module 2
Module 3
D
Day 7 STIA
E
Cx3cr1
H2-Eb1
Steady state
Day 7 STIA
Expression level
MA
MB
MC
MD
Macrophage subsets
F
Steady state expressions
Day 7 STIA expressions
Steady state
Day 7 STIA
G
Up I
Up II
Up III
STIA Up I
STIA Up II
STIA Up III
Module score
H
subset
Gene expression
Macrophage subset

Figure 2.2.5. The induction of STIA give rise to macrophages with specialized transcriptional profiles that extend beyond their steady state definitions

(A-B) tSNE visualization and proportions of annotated macrophage subsets for (A) steady state and (B) day 7 STIA after label transfer. (C-D) 3D scatterplot visualizing the expression module scores of three pseudotime gene modules identified in Figure 1 in (C) steady state and (D) day 7 STIA single cells. (E) Log-normalized gene expressions for Cx3cr1 and H2-Eb1 in steady state and day 7 STIA single cells. (F) Averaged log expressions of top 10 steady state subset gene markers (Figure 1D) in steady state (blue) and day 7 STIA (red) macrophage subsets. (G) tSNE plots visualizing the expression module scores for the Up I-III clusters from STIA time course K-means (Figure 3D) in day 7 STIA single cells and a schematic (top left) showing their relative enrichments across macrophage subsets. (H) Heatmap of log-normalized expressions values for select STIA Up I-III cluster markers in day 7 single cells. The cells are grouped by their subset annotations and hierarchically clustered.

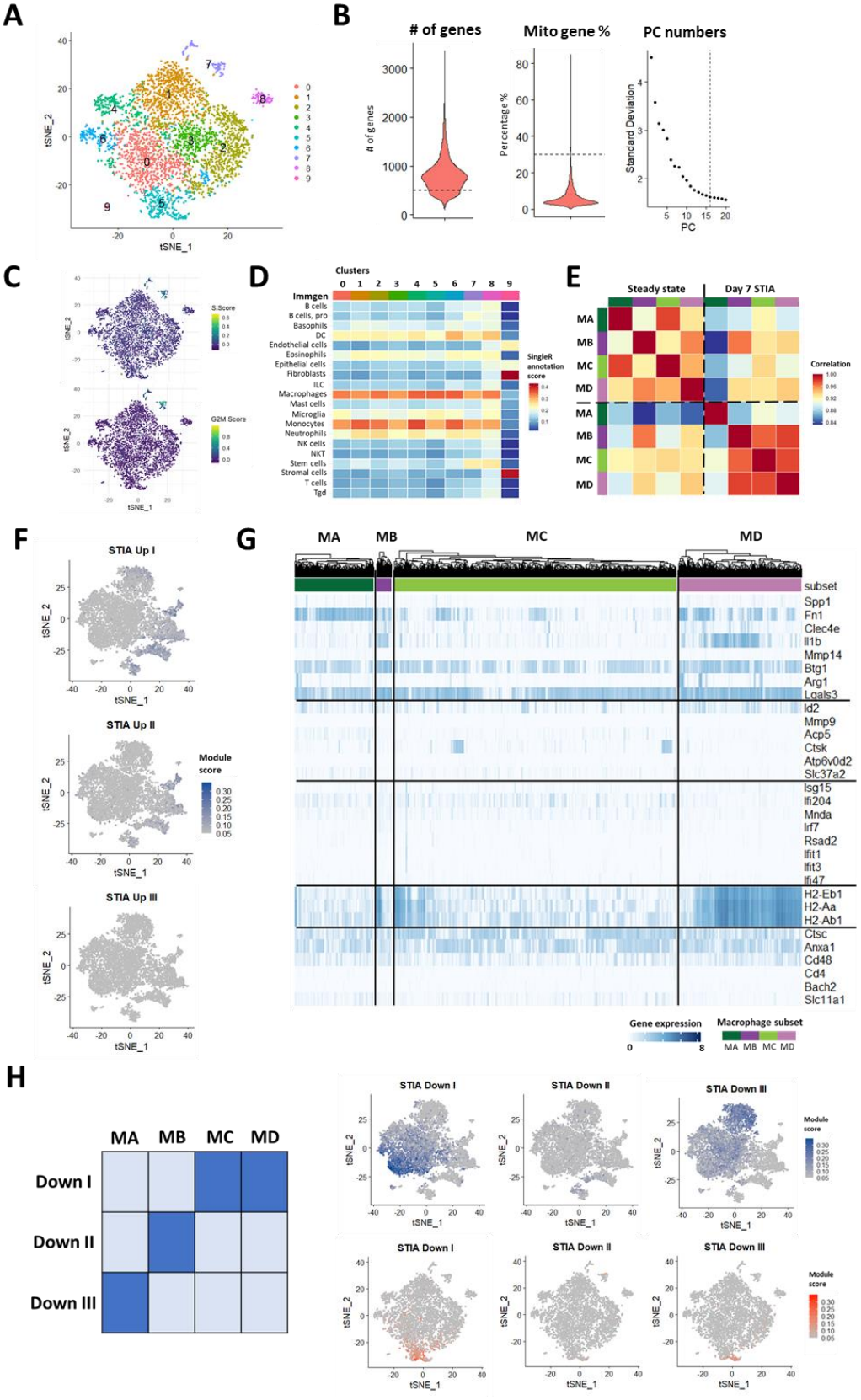

A
B
# of genes
Mito gene %
PC numbers
C
D
E
Steady state
Day 7 STIA
F
STIA Up I
STIA Up II
STIA Up III
G
MA
MB
MC
MD
subset
Spp1
Fn1
Clec4e
Il1b
Mmp14
Btg1
Arg1
Lgals3
Id2
Mmp9
Acp5
Ctsk
Atp6v0d2
Slc37a2
Isg15
Ifi204
Mnda
Irf7
Rsad2
Ifit1
Ifit3
Ifi47
H2-Eb1
H2-Aa
H2-Ab1
Ctsc
Anxa1
Cd48
Cd4
Bach2
Slc11a1
Gene expression
Macrophage subset
H
Down I
Down II
Down III
STIA Down I
STIA Down II
STIA Down III

Figure S2.2.5 (A) TSNE of 10 clusters from 3916 Day 7 STIA CD45+CD11b+Ly6G-SiglecF- cells. (B) Distributions of # of genes expressed, UMIs, and percentage of mitochondrial gene expression with the cutoffs for quality control. (C) Cell cycle scoring of Day 7 STIA single cells using S phase (top) and G2/M phase genes (down), visualized on tSNE. (D) Heatmap illustrating SingleR assignment of Immgen reference cell types. (E) Correlation heatmap of averaged expression profiles for steady state and Day 7 STIA macrophage subsets. (F) tSNE plots visualizing the expression module scores for the Up I-III clusters from STIA time course K-means (Figure 2.2.4D) in steady state single cells. (G) Heatmap of log-normalized expressions values for select STIA Up I-III cluster markers in steady state single cells. The cells are grouped by their subset annotations and hierarchically clustered. (H) tSNE plots visualizing the expression module scores for the Down I-III clusters from STIA K-means in steady state (top) and Day 7 STIA (bottom) single cells and a schematic (left) showing their relative enrichments across macrophage subsets.

Genes associated with acute inflammatory processes in synovial macrophage subsets exhibit persistent dysregulation during chronic inflammatory arthritis

Collagen induced arthritis (CIA) is an inflammatory arthritic model that emulates chronic synovial inflammation rather than acute inflammation as in STIA. We therefore utilized the CIA model to investigate how synovial macrophage subsets respond during chronic inflammatory arthritis [295]. We performed RNA-seq on the four macrophage subsets on day 0, 27, 41, and 62 post 2nd collagen injection. In contrast to STIA, the clinical severity scores do not return to steady state levels 62 days post injection (Figure 2.2.6A) and proportions of all four macrophage

subpopulations are expanded throughout the CIA time course (Figure 2.2.6B-C). The transcriptional profiles exhibited similar patterns in CIA and STIA, where replicates from day 62 remained distinct from day 0 replicates (Figure S2.2.6A). We computed the expression foldchanges of temporally variable genes (N = 3348) between Days 41 and 62 relative to Day 0 of CIA time course. Day 41 was chosen as the reference time point for peak inflammation in CIA due to its overall high correlations with Day 7 transcriptional profiles from STIA model (Figure S2.2.6B) to further investigate how chronic inflammation manifest in CIA model as opposed to STIA. Genes were classified as Acute (absolute 2-foldchanges at Day 41 only), Persistent (Both Day 41 and 62), or Chronic (Day 62 only) (Figure S2.2.6C). We found that the majority of DE genes from CIA experiment were classified as Persistent in all subsets (Figure S2.2.6C), as opposed to STIA (Figure S2.2.4H). These data are consistent with overall increase in slopes of the fold changes between the two reference timepoints (D62 vs D41 for CIA, D21 vs D7 or STIA) across all macrophage subsets (MA: 0.36 to 0.57; MB: 0.44 to 0.61; MC: 0.19 to 0.48; MD: 0.19 to 0.46). Furthermore, the majority of Acute genes previously defined in STIA model, which are those altered at Day 7 but not Day 21 post serum transfer, remained altered at Day 62 of CIA time course (Figure 2.2.6D). In addition, all subsets during CIA exhibit similar expression patterns compared to STIA time course on the levels of K-means clustering (Figure 2.2.6E), biological processes (Figure 2.2.6F), and individual genes (Figure 2.2.6G). This indicate that persistent dysregulation of genes previously associated with STIA is the hallmark of chronic inflammatory arthritis.

To compare the transcriptional signatures between CIA and STIA models more closely, we clustered the temporally variable genes into six clusters by their fold-changes in gene expression relative to day 0 using K-means (Figure S2.2.6D-E). Genes in cluster I were upregulated over the

CIA time course in all subsets, and included genes associated with cell chemotaxis (e.g., Spp1, Lgals3) [261, 296], lipid storage (e.g., Cav1, Soat1) [297, 298], and collagen catabolic (e.g., Ctsk, Mmp14) [299, 300] processes . These processes were also observed in the STIA experiment, evident by the significant number of genes shared between CIA cluster I and STIA Up I-III clusters (Figure S2.2.6E). While genes linked to leukocyte activation and chemotaxis processes displayed similar dynamics over both STIA and CIA time courses, those involved in osteoclastic process remain elevated at Day 62 CIA specifically in MC subset. Interestingly, the upregulation of Type I interferon genes, a major characteristic of MB subset during STIA time course, is seemingly absent in CIA (Figure S2.2.6F). On the other hand, cluster II consisted of genes downregulated over the course of CIA in all subsets, which is supported by their significant overlap with genes from STIA Down I-III clusters (Figure S2.2.6E). Genes associated with ossification processes (e.g., Fgfr1, Fermt2) exhibited persistent downregulation in CIA compared to STIA, with MA and MB subsets showing the most pronounced patterns. Additionally, we observed the relative absence of downregulation for MAPK cascade and cell communication pathways in MD subset during CIA compared to STIA time course (Figure S2.2.6F). Patterns of sustained downregulation were also observed for downstream target genes of TFs MEF2A and KLF3 (Figure S2.2.6G-H), which was previously attributed to maintenance of steady state synovial macrophage functions (Figure 2.2.3). Clusters IV and V in turn represent up and down-regulated genes specific to MB subset during CIA. These two clusters overlapped significantly with their counterpart MB-specific clusters in STIA (Up III and Down II) (Figure S2.2.6E). MB-specific pathways previously characterized in STIA time course therefore displayed similar patterns during CIA, with the persistent upregulation of adaptive immune response (e.g., Cd48, Bach2) and downregulation of cell cycle genes (e.g., Top2a, Mki67) observed in MB cells. Cluster IV additionally consisted of genes linked to positive

regulation of RNA metabolic processes, including Nfkb1 and Nlrp3. The enrichment of response to growth factors process (e.g., Xbp1, Lgmn) was also observed in this cluster, but not in STIA. These findings suggest that MB subsets may assume a more transcriptionally activated phenotype in response to CIA. Collectively, these data suggest that persistent alterations of specific biological processes previously characterized in STIA model, especially those associated with acute inflammation, contribute substantially to the chronic profile of the CIA model.

We also observed clusters that were specific to particular subsets. For instance, clusters II and III consisted of genes upregulated specifically in MA/MB and MC/MD subsets, respectively (Figure S2.2.6D). Genes in cluster II were enriched for protein deSUMOylation process, including Uspl1 and Senp7 (Figure S2.2.6D). Small ubiquitin-like modifiers (SUMOs) are a subgroup of post-translational epigenetic modifiers [301] and their dysregulation have been linked to apoptosis, cell migration, and chronic cartilage degradation in RA through modifications of MMP proteins [302]. Cluster II also contained genes downregulated specifically in MC and MD subsets, such as Jun and Lpar1. These genes were also present in the MC/MD-specific STIA Cluster Down I (Figure S2.2.6E). Cluster III included genes associated with dendritic cell (DC) differentiation, such as Ltbr and Tnfsf9 [303, 304] (Figure S5D). Evidence of DC differentiation from myeloid cells have been reported in RA, which are facilitated through growth factors present in synovial fluid and may contribute to the activation of inflammatory-type Th1 responses [305, 306]. In summary, the variations in gene expression patterns observed between CIA MA/MB and MC/MD compartments, which were not present in STIA, provide insights into the differing response of synovial macrophage subsets that potentially underlie the chronic pathology of CIA model.

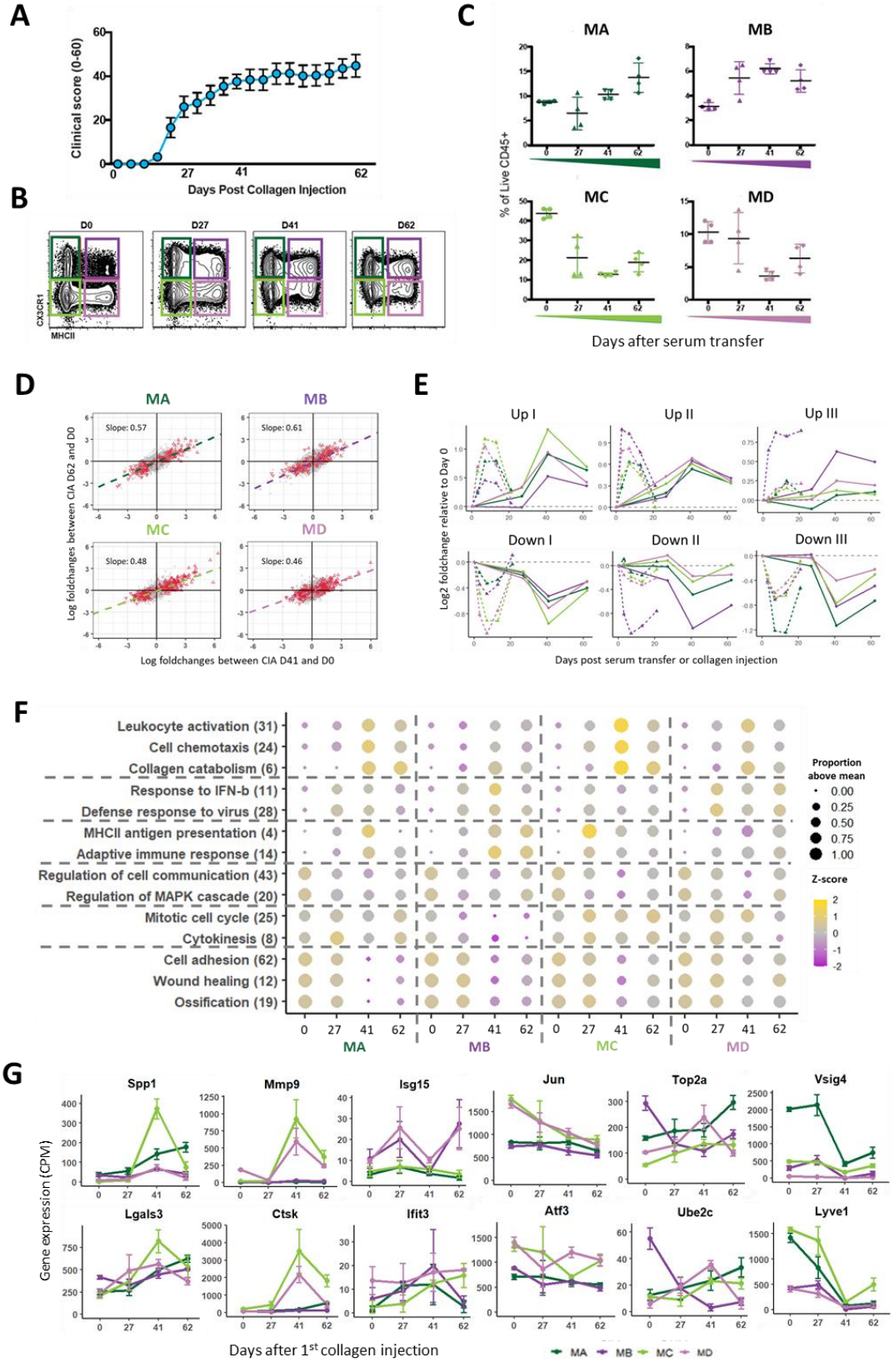

A
Clinical score (0-60)
Days Post Collagen Injection
B
D0
D27
D41
D62
CX3CR1
MHCII
C
MA
MB
MC
MD
% of Live CD45+
Days after serum transfer
D
MA
MB
MC
MD
Slope: 0.57
Slope: 0.61
Slope: 0.48
Slope: 0.46
Log foldchanges between CIA D62 and D0
Log foldchanges between CIA D41 and D0
E
Up I
Up II
Up III
Down I
Down II
Down III
Log2 foldchange relative to Day 0
Days post serum transfer or collagen injection
F
Leukocyte activation (31)
Cell chemotaxis (24)
Collagen catabolism (6)
Response to IFN-b (11)
Defense response to virus (28)
MHCII antigen presentation (4)
Adaptive immune response (14)
Regulation of cell communication (43)
Regulation of MAPK cascade (20)
Mitotic cell cycle (25)
Cytokinesis (8)
Cell adhesion (62)
Wound healing (12)
Ossification (19)
Proportion above mean
0.00
0.25
0.50
0.75
1.00
Z-score
MA
MB
MC
MD
G
Spp1
Mmp9
Isg15
Jun
Top2a
Vsig4
Lgals3
Ctsk
Ifit3
Atf3
Ube2c
Lyve1
Gene expression (CPM)
Days after 1st collagen injection
MA
MB
MC
MD

Figure 5. Genes associated with acute inflammatory processes in synovial macrophage subsets exhibit persistent dysregulation during chronic inflammatory arthritis

(A) Clinical arthritic scores over the course of CIA. (B-C) Proportions of synovial macrophage subsets over time. (D) Scatter plots of CIA Day 41 and Day 62 expression fold change relative to Day 0. Red dots represent STIA acute genes defined in Figure S2.2.4. (E) Average log expression fold change relative to Day 0 for select biological processes across STIA and CIA time courses. (F) Relative gene expression of STIA-related biological processes (Figure 2.2.4) across macrophage subsets during CIA time course. Size of circles indicate proportion of genes expressed above the average of all samples while color represents mean expression Z-scores across all samples. (G) Gene expressions of select genes for highlighted biological processes.

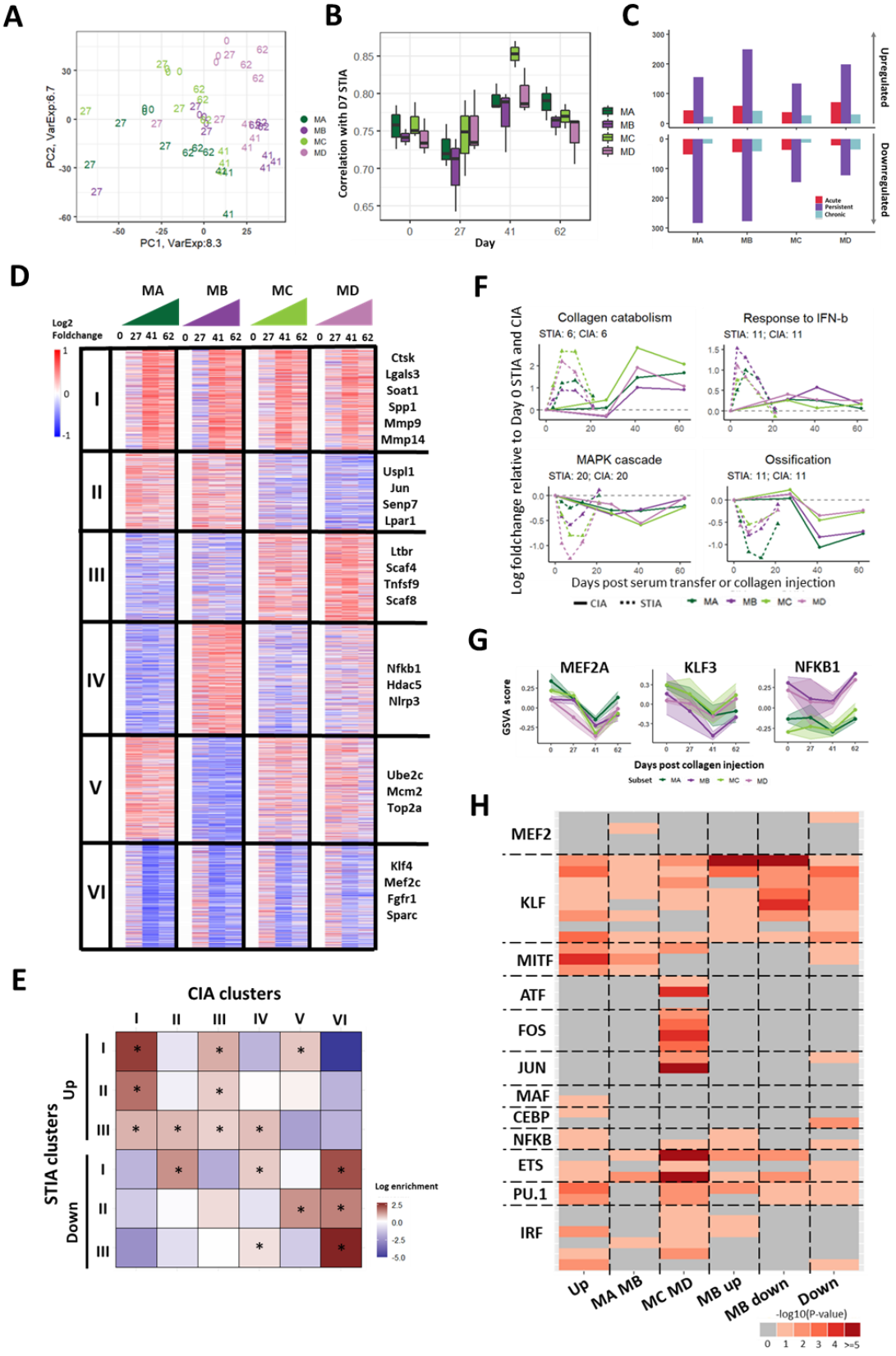

A
PC2, VarExp:6.7
PC1, VarExp:8.3
MA
MB
MC
MD
B
Correlation with D7 STIA
Day
C
Upregulated
Downregulated
Acute
Persistent
Chronic
D
MA
MB
MC
MD
Log2 Foldchange
0 27 41 62
I
II
III
IV
V
VI
Ctsk
Lgals3
Soat1
Spp1
Mmp9
Mmp14
Uspl1
Jun
Senp7
Lpar1
Ltbr
Scaf4
Tnfsf9
Scaf8
Nfkb1
Hdac5
Nlrp3
Ube2c
Mcm2
Top2a
Klf4
Mef2c
Fgfr1
Sparc
E
CIA clusters
STIA clusters
Up
Down
Log enrichment
F
Collagen catabolism
STIA: 6; CIA: 6
Response to IFN-b
STIA: 11; CIA: 11
MAPK cascade
STIA: 20; CIA: 20
Ossification
STIA: 11; CIA: 11
Log foldchange relative to Day 0 STIA and CIA
Days post serum transfer or collagen injection
CIA
STIA
G
MEF2A
KLF3
NFKB1
GSVA score
Days post collagen injection
Subset
H
MEF2
KLF
MITF
ATF
FOS
JUN
MAF
CEBP
NFKB
ETS
PU.1
IRF
Up
MA MB
MC MD
MB up
MB down
Down
-log10(P-value)

Figure S2.2.6 (A) PCA of gene expression profiles between replicates from CIA time course. Text indicates number of days after serum transfer. (B) Correlations of gene expression profiles between day 7 STIA subpopulations and CIA subpopulations over time. (C) Quantification of acute (Day 41 only; red), persistent (Days 41 & 62; purple), and chronic (Day 62 only; light blue). (D) K-means clustering (K = 6) of the mean fold-change of 3428 differentially expressed genes across the CIA time course in at least one macrophage subset. (E) Log2 enrichments of gene overlap between STIA and CIA K-means clusters. Asterisks indicate adjusted hypergeometric p-value < 0.05. (F) Average log expression fold change relative to Day 0 for select biological processes across STIA and CIA time courses. Numbers of genes expressed in STIA and CIA models are indicated at top. (G) GSVA-inferred expression scores for downstream target genes of MEF2A, KLF3, and NFKB-p65 from Dorothea database across macrophage subsets and STIA time course. (H) Significance of DNA binding motif enrichments for select TF families across CIA k-means clusters. The color scale represents -log10(p-value) as computed by HOMER.

Human RA patients exhibit patterns of synovial macrophage heterogeneity similar to mice

To investigate whether our characterization of synovial macrophage heterogeneity in mice can be translated into humans, we performed scRNA-seq on sorted CD45+ synovial immune cells of 7 RA patients, which were collected through the RhEumatoid Arthritis SynOvial tissue Network (REASON) (Table 3-4) [307]. Integrating patient samples yielded a total of 18718 cells across multiple immune cell lineages, including B, T/NK, neutrophils, and myeloid cells (Figure S2.2.7A). We computationally isolated 3727 myeloid cells by creating a gene module score using human orthologs of 6 signature macrophage genes derived from our murine single cell data (Figure 2.2.7A). We further verified the myeloid population using individual gene and surface protein

expression (Figure S2.2.7B-C).

We adopted a similar approach based on gene module scoring to determine whether human synovium exhibit macrophage heterogeneity similar the murine synovium. We utilized the top 10 orthologous genes ranked by log fold change from the murine single cell data to compute module scores for each mouse macrophage subset in individual human synovial cells. We then normalized the score and ranked to annotate specific mouse macrophage subsets onto single cells from RA patients (see method section for details) (Figure 2.2.7B, S2.2.7D). The resulting annotation allowed for distinct separation of the four subsets especially the tissue-resident (MA and MC) /monocyte-derived (MB and MD) axis (Figure 2.2.7C). The transcriptional distinctiveness among the four macrophage subsets in humans was supported by their prominent overlap with individual clusters obtained through unbiased graph-based clustering (Figure S2.2.7E-F) and by surface protein expression as measured by CITE-seq (Figure S2.2.7G). These data suggest that human myeloid cells are also composed of subsets that parallel mouse synovial macrophage subsets.

Recently, two studies investigated the heterogeneity of human synovial macrophages in RA patients. One of them profiled over 32000 synovial macrophages across RA patients with different clinical states and identified four major subpopulations, categorizable by surface expression of MERTK [228]. We observed similarities between MA-MD subsets and their four populations by assessing the overlap of significant marker genes. Specifically, MA and MC subsets corresponded to TREM2^{hi} and LYVE1+ subpopulations, respectively. These two populations were classified as MERTK^{+} by flow cytometry, which were shown to possess a unique transcriptional profile associated with remission maintenance and were enriched with numerous anti-inflammatory

regulators. On the other hand, the marker genes of MB and MD subsets overlapped with those reported in MERTK^{-} subpopulations, which were also CLEC10A^{+} and S100A12^{+}. The MERTK^{-} macrophages were associated with pro-inflammatory phenotypes and were predominately enriched in treatment-naïve or -resistant RA patients (Figure S2.2.7H). Another recent study is from the Accelerating Medicines Partnership (AMP) consortium, which profiled synovial myeloid cells from patients with OA and RA at varying levels of leukocyte-infiltration [227]. We annotated MA-MD synovial macrophage subsets onto single cells from the AMP study using top 10 marker genes ranked by fold change from our human data (Figure S2.2.7I-J). We found that cells from OA and leukocyte-poor RA patients, which exhibited lower Krenn inflammation scores, were predominately labeled as tissue-resident MA and MC subsets. In contrast, cells from leukocyte-rich RA patients were mostly labeled monocyte-derived MB and MD subsets (Figure S2.2.7K). Collectively, these findings indicate that our findings from mouse models and projection onto human patients are congruent with previous published classification schemes for human synovial macrophages.

Considerable variation in proportions of macrophage subsets was observed across patients (Figure 2.2.7D). Given that comparison with other published studies suggest that subset proportions may be linked to clinical properties, we next investigated whether such variations are associated with disease severity among patients included in this study. The seven patients were classified as responsive to treatment, non-responsive to treatment, low disease severity, or no follow up according to reported Clinical Disease Activity Index (CDAI) from initial and follow up (if available) clinic visits up to 18 weeks after synovial biopsy procedure and subsequent change in therapy (Figure 2.2.7E). We identified that the distribution of the macrophage subsets, which

can be quantified via Shannon diversity index, were significantly associated with less severe RA activity based on CDAI values from the initial clinic visit (Figure 2.2.7F). We further observed that the proportions of MA and MC exhibited negative correlation with disease severity across multiple metrics from the initial visit, including CDAI, Disease Activity Score-28 with C-reactive protein (DAS28-CRP), Disease Activity Score-28 with erythrocyte sedition rate (DAS28-ESR), swollen joint count (SJC), and tender joint count (TJC), whereas positive correlations was observed for MB subset proportions. On the other hand, no clear patterns were found for the MD subset (Figure S2.2.7L). These data suggest that the relative proportions of the synovial macrophage subsets can potentially be indicative of disease severity in RA patients.

Lastly, we sought to determine whether the transcriptional profiles of synovial macrophages differ between RA patients that were responsive to treatment and those that were not. GO enrichment analysis revealed that the responsive patients exhibit relative upregulation of genes associated with antigen presentation (e.g., HLA-DQA1, HLA-DRB1) and macromolecule biosynthetic process (e.g., APOE, NR4A2) based on differential gene expression of macrophages from the responsive and non-responsive groups of patients. Numerous genes attributed to $CLEC10A^+$ subset in Alivernini et al. [228], including RPS proteins, REL, and BTG1 were elevated in responsive patients. On the other hand, patients that did fail to respond to treatment preferentially upregulated genes linked to glycolytic process (e.g., GAPDH, ALDOA) and cell adhesion (e.g., FLNA, ANXA1). Non-responsive patients also shared multiple genes with the $SPP1^+$ subset from Alivernini et al., such as MIF, S100A10, and CD36 (Figure 2.2.7G-H). The $CLEC10A^+$ and $SPP1^+$ subpopulations were associated with RA patients who have active disease. These results shed light on the potential functional diversity in synovial macrophages that may

underlie the variation in treatment responsiveness among RA patients.

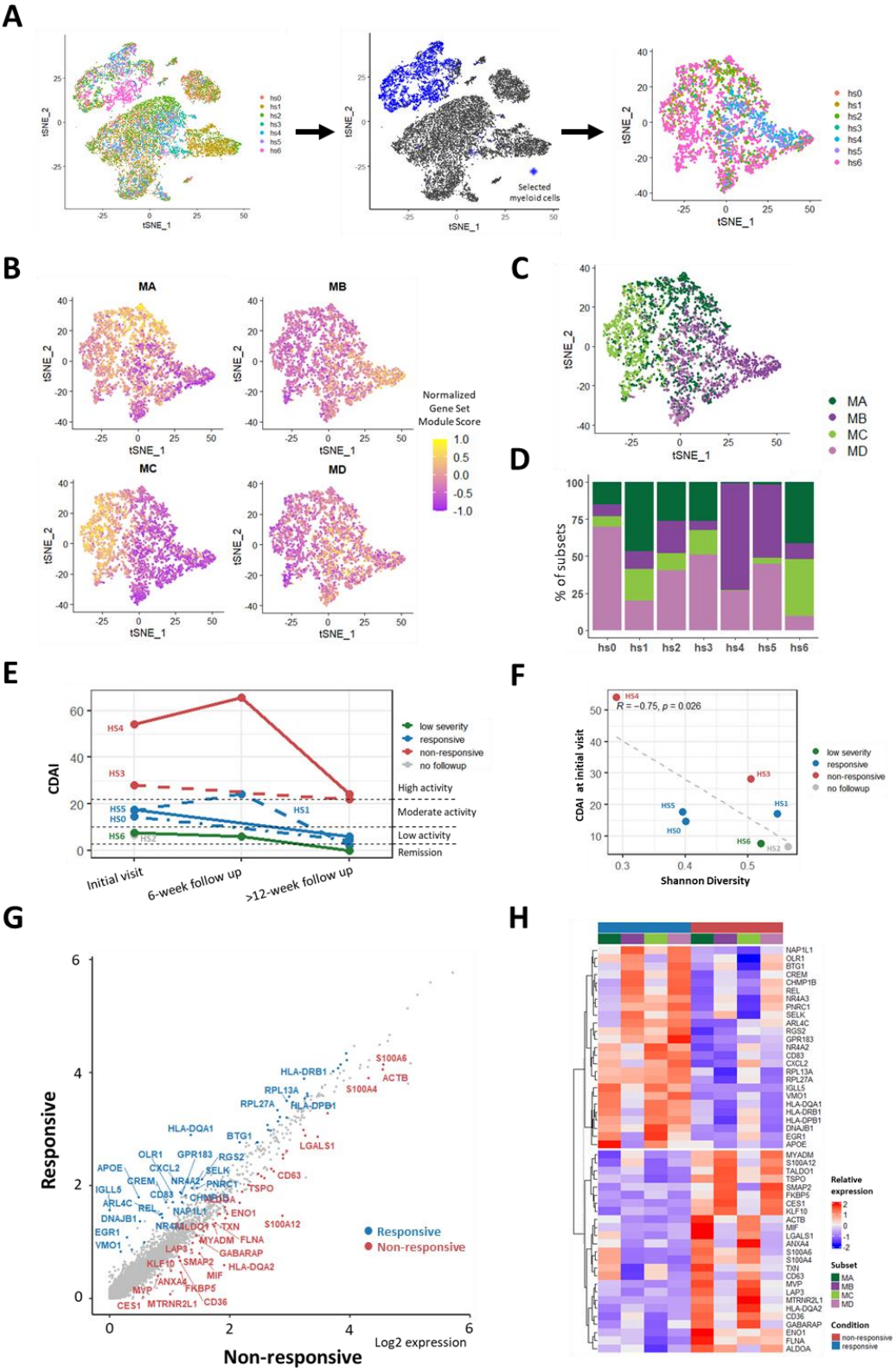


Figure 2.2.7. Human RA patients exhibit patterns of synovial macrophage heterogeneity similar to mice

(A) TSNE visualization of 18718 CD45+ synovial immune cells from biopsy samples of 7 RA patients. 3727 mononuclear phagocytes were isolated through module scoring and filtering (> 0.3) of myeloid signature genes. (B) Enrichment of module scores in human synovial mononuclear phagocytes derived from top 10 gene markers from murine steady state macrophage subsets. (C) Annotation of murine synovial macrophage subsets on human cells and (D) breakdown of proportions by patients. (E) Clinical Disease Activity Index (CDAI) of the RA patients at initial clinical visit, 6-week follow-up, and >12-week follow-up. Patients were categorized as low severity, responsive, non-responsive, and no follow up based on their CDAI trends across the clinic visits. (F) Scatterplot of CDAI values at initial visit and Shannon diversity index computed from the relative proportions of macrophage subsets for each patient. The correlation coefficient and p-value is indicated on the top left. (G) Scatterplot of averaged log expression values between responsive and non-responsive patients. The top 25 DE genes are colored. (H) Heatmap of relative gene expressions for top 25 responsive and non-responsive DE genes grouped by patient responsiveness and synovial macrophage subsets. The genes are ordered by hierarchical clustering.

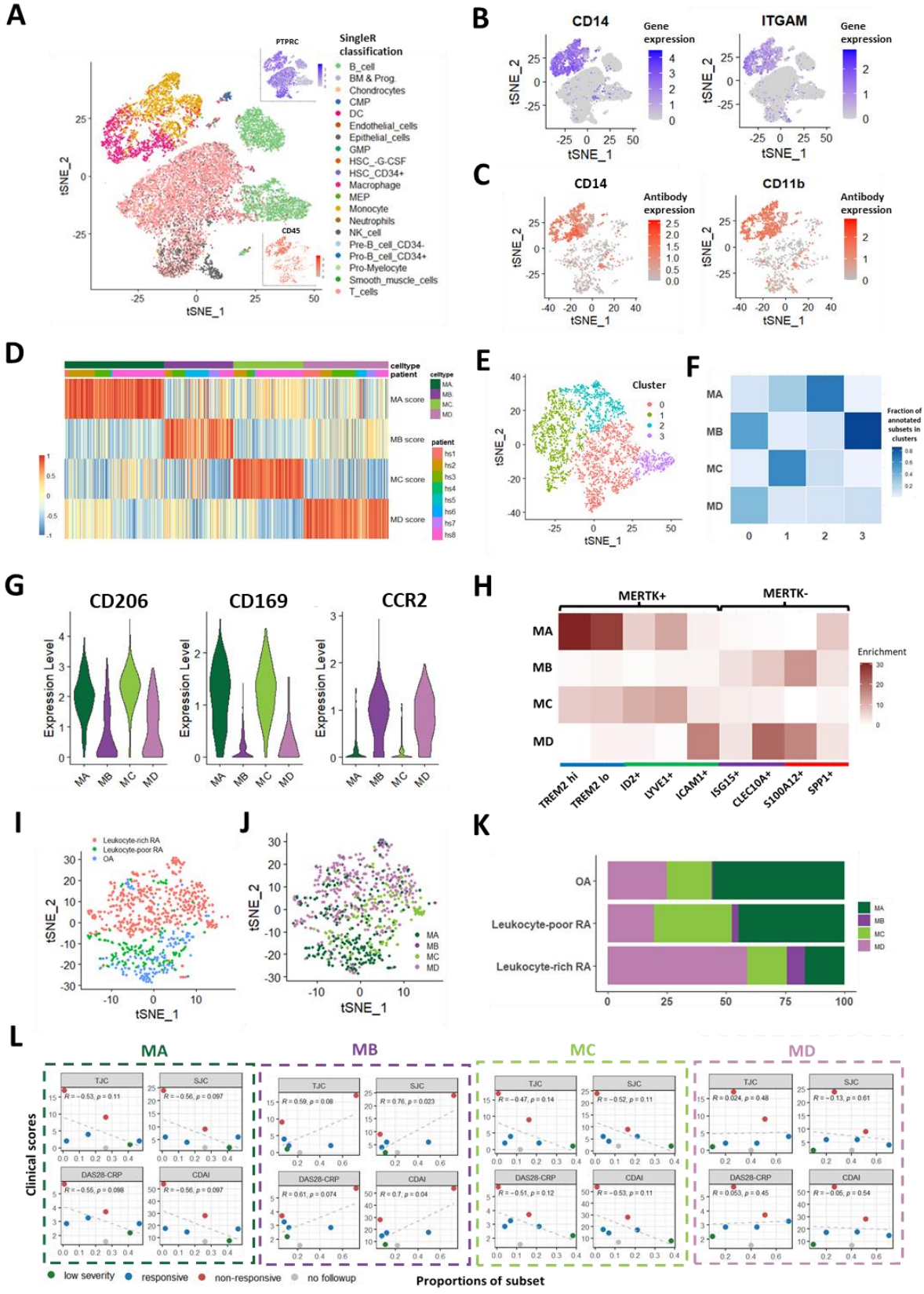

A
PTPRC
CD45
SingleR classification
B_cell
BM & Prog.
Chondrocytes
CMP
DC
Endothelial_cells
Epithelial_cells
GMP
HSC_-G-CSF
HSC_CD34+
Macrophage
MEP
Monocyte
Neutrophils
NK_cell
Pre-B_cell_CD34-
Pro-B_cell_CD34+
Pro-Myelocyte
Smooth_muscle_cells
T_cells
tSNE_1
tSNE_2
B
CD14
ITGAM
Gene expression
C
CD14
CD11b
Antibody expression
D
celltype
patient
MA score
MB score
MC score
MD score
E
Cluster
F
MA
MB
MC
MD
Fraction of annotated subsets in clusters
G
CD206
CD169
CCR2
Expression Level
H
MERTK+
MERTK-
Enrichment
TREM2 hi
TREM2 lo
ID2+
LYVE1+
ICAM1+
ISG15+
CLEC10A+
S100A12+
SPP1+
I
Leukocyte-rich RA
Leukocyte-poor RA
OA
J
K
L
MA
MB
MC
MD
Clinical scores
TJC
SJC
DAS28-CRP
CDAI
low severity
responsive
non-responsive
no followup
Proportions of subset

Figure S2.2.7 (A) SingleR assignment of human BluePrint reference cell types. (B) Gene expressions and (C) surface expressions of select myeloid marker genes/proteins. (D) Heatmap indicating module scores of the 4 murine macrophage subpopulations in human cells grouped by annotation results. (E) UMAP clustering of human synovial myeloid cells and (F) the proportional distribution of the four macrophage subpopulations among the clusters. (G) Violin plots showing surface expressions of select proteins. (H) Enrichments of overlaps between significant marker genes of MA-MD macrophage subsets and 9 clusters reported in Alivernini et al. 2020 [76] (I-J) tSNE of single cells from AMP study [77] colored by (I) RA disease subtype and (J) annotations of MA-MD subsets. (K) Quantification of annotated MA-MD subsets among cells from different RA disease subtypes. (L) Scatterplots showing relationships of various clinical variables against proportions of each macrophage subset. TJC: Tender Joint Count, SJC: Swollen Joint Count, DAS28-CRP: Disease Activity Score-28 with C-reactive protein, CDAI: Clinical Disease Activity Index.

Table 2.2 Clinical characteristics of REASON RA patients

	Initial Visit (N = 7)	6 week follow up	> 12 week follow up
Age (years)	44.29 ± 5.79		
Sex (Male or Female)	M = 1; F = 6		
Ethnicity (W = White, H = Hispanic, A = African American)	W = 3; H = 3; A = 1		
Disease duration (months)	76.59 ± 28.23		
CDAI	20.71 ± 6.17	31.83 ± 17,62 (N = 3)	9.75 ± 4.28 (N = 6)
DAS28-ESR	3.97 ± 0.64	4.26 ± 1.47 (N = 3)	2.42 ± 0.70 (N = 5)
DAS28-CRP	3.15 ± 0.50	3.19 ± 1.44 (N = 3)	2.50 ± 0.37 (N = 5)
ESR	32.29 ± 8.55	39.67 ± 20.67 (N = 3)	17.20 ± 4.77 (N = 5)
CRP	3.97 ± 0.64	4.73 ± 4.49 (N = 3)	3.33 ± 2.15 (N = 6)
SJC	7.57 ± 2.89	10.75 ± 6.13 (N = 4)	3.00 ± 1.21 (N = 6)
TJC	5.00 ± 2.29	6.50 ± 4.72 (N = 4)	4.50 ± 1.73 (N = 6)

Table 2.3 Sequencing statistics of REASON RA patients

	HS0	HS1	HS2	HS3	HS4	HS5	HS6
# of cells	5162	5713	3897	1515	1810	1370	2754
Reads Mapped Confidently to Genome	78.30%	83.80%	88.10%	90.30%	91.60%	87.50%	91.80%
Fraction Reads in Cell	91.40%	91.60%	80.30%	94.30%	73.10%	87.10%	92.80%
Mean Reads per Cell	104297	60325	92644	218554	196286	256806	123268
Median Genes per Cell	1062	954	651	785	1888	2192	4270
Median UMI Counts per Cell	3452	2776	1697	3108	6004	7113	20758
Median % mitochondrial reads per Cell	2.71%	3.67%	4.00%	0.91%	4.71%	7.23%	7.56%
Median % ribosomal reads per Cell	38.34%	28.75%	31.14%	48.54%	23.48%	20.70%	8.45%
# of Cells after QC	4623	5254	3420	1391	1476	940	1614

2.2.3 Discussion

In this study, we demonstrate the existence of four macrophage subpopulations in healthy murine joint synovium that are categorizable by surface expression of CX3CR1 and MHCII. The MA and MC subsets, which are MHCII-, would typically be considered tissue-resident macrophages based on their transcriptional profile and lineage-tracing results suggesting that they are derived from embryonic progenitors. This is in line with our prior studies suggesting that MHCII- synovial macrophages were long-lived and radio-resistant [225]. The two sub-populations are further distinguished by the unique expression of genes, such as Vsig4, Sparc, and Cx3cr1, in CX3CR1+ MA. The CX3CR1+ synovial lining population described in Culemann et al. express the same marker genes [220]. This is in contrast to the interstitial cells presented in that study which represent the majority of synovial macrophages. Thus, we conclude that MA form the synovial lining, while MC localize to the interstitium. The distinct gene expression signatures and localizations of the MA and MC subpopulations suggest they exhibit specialized function in maintaining joint homeostasis.

In contrast, we propose the MHCII+ MB and MD subsets are monocyte-derived cells. Both

populations highly express genes associated with antigen presentation and inflammation, but MB cells have higher expression of genes associated with circulating monocytes, such as Ly6c2 and Plac8. Given that our single-cell gating does not explicitly exclude monocytes, the MB subset may not represent a permanent state, but instead include cells that have recently infiltrated into the synovium and have yet to fully shed their patrolling phenotype. Based on our approaches, it is not possible to determine whether these cells are truly embedded in synovial tissue or associated with the vasculature. The transient state of these cells is supported by the considerable expansion and subsequent contraction of MB in the peak of inflammatory arthritis. We also observe that these cells are on the extreme end of the pseudotime trajectory opposite MA and MB. On the other hand, MD macrophages exhibit more similarities to the MC subset and it is likely that the MC-MD axis aligns to our previous results showing that MHCII+ transition into MHCII- cells [225]. We also observe the highest expression of genes associated with proliferation in MD macrophages, which is consistent with the MHCII+ proliferating cells in the Culemann et al study [220]. Taken together, the MB and MD sub-populations may not exhibit discrete functions in steady state but rather reflect the dynamics of macrophage heterogeneity in the joint synovium.

Other studies have presented a classification of synovial macrophages in the human joint [227, 228]. In the published AMP study, Zhang et al. used single-cell RNA-seq to categorize four sub-populations of CD14+ myeloid cells from synovial tissue in RA and OA patients as IL1B+ pro-inflammatory, NUPR1+, C1QA+, and IFN-activated [227]. In is not entirely clear whether all synovial macrophages would be identified through this method. Nevertheless, we find that the expression profile of NUPR1+ is consistent with our synovial lining (MA) macrophages while the other subpopulations are less obviously matched with our subsets. However, using our

classification system, we are able to independently assign human cells to MA-MD in both AMP data and our own novel single-cell data set of CD45+ immune cells from synovial biopsy of RA patients with a range of clinical presentations. We are able to demonstrate that an expanded MB subset is associated with RA, particularly in patients with high disease severity, while MA exhibits the opposite trend. Alivernini et al. reported to 9 single-cell clusters among synovial macrophages from RA patients which they group into TREM2+FOLR2+, TREM2-FOLR2+, HLA+CD48+ and CD48+ subpopulations [228]. Although there was variability across clusters, the populations associated with healthy and/or RA remission in their study were generally similar to our MA and MC populations, while those associated with active RA linked to MB and MD. The general consensus of synovial macrophage populations across studies and species suggests that our results truly reflect the underlying biology but differences in experimental and computational approaches may lead to slight differences.

In agreement with prior studies [225, 228, 235], we found that cell counts of all four subpopulations are expanded in response to inflammatory arthritis using two distinct and well-established mouse models: STIA and CIA [295, 308]. This likely reflects the infiltration of monocytes from circulation, which has been extensively documented in multiple tissues encountering immune challenges [79, 80, 88, 309, 310]. While we see a substantial rise in the expression of monocyte-associated genes in STIA, we do not observe a significant increase in cell cycle genes. In fact, cell cycle genes are actually downregulated in MB, indicating that recruitment is a more probably explanation for the expansion then proliferation. By flow cytometry, we observe the greatest increase in MB on day 7 post-STIA compared with other populations. Similarly, MB and MD exhibit a greater increase in relative abundance than MA and MC in the day 7 single-cell

data. This supports the idea that cells must transition through the MB and MD states before they are able to fill the tissue-resident niche of MA and MC [55, 56]. However, it is not entirely consistent with the niche model which proposes that there is a finite limit on the number of tissue-resident macrophages at any given time. Moreover, synovial lining macrophages However, this discrepancy may be related to the limitations of quantifying macrophage sub-populations by flow cytometry. Based on single-cell RNA-seq, we observe a relative decrease in MA and MC but this data does not provide an accurate method to achieve and absolute quantification.

Due to the functional heterogeneity of synovial macrophages at steady state, we might expect them to exhibit a variety of responses in the development of RA. However, our data demonstrates that in addition to the increase in monocyte gene expression mentioned above, all sub-populations generally exhibit a similar up-regulation of inflammatory and immune response pathways as well as a down-regulation of homeostasis. MB is distinguished by having the most extreme response, while MA exhibits the least. The exception is for genes associated with tissue-residence and the synovial lining phenotype. These genes are down-regulated in all sub-populations, but particularly in MA where they were highest at steady-state. These results suggest that there is a disruption of synovial lining macrophage function and phenotype during the development of arthritis and is consistent with the proposal that they form a physical barrier that shields the joint from external perturbation. The artificial depletion of CX3CR1+ synovial lining macrophages led to the onset of arthritis [220] and RA patients appear to be depleted of these cells [227, 228]. Nevertheless, the causal relationship in disease is unclear: does the loss of MA macrophage cause arthritis or is it simply a symptom of the development of inflammation.

Comparative analysis of acute (STIA) and chronic (CIA) arthritis mouse models highlights the differences over time in arthritis resolution vs. propagation. In STIA, the macrophage sub-populations largely return to their steady-state numbers and transcriptional profile by day 21. In contrast, the majority of CIA-associated transcriptional changes are maintained at day 62. Generally, the changes associated with the development of arthritis – increase of inflammatory genes and decrease of homeostasis – are shared in both models. However, the macrophage sub-populations in CIA exhibit additional long-term effects such as the upregulation of genes associated with connecting tissue replacement and neutrophil apoptosis in MA. These changes may be indicative of development of fibrosis [9] and a shift to a non-inflammatory equilibrium [10]. It is also noteworthy that the MB phenotype does as strongly drive CIA as it does in STIA, but this distinction may be more related to the change in timeframe of the experiments than a true difference in the pathogenesis.

In this study, we defined a classification system to describe murine synovial macrophage heterogeneity that we translated to characterize the macrophage niche in synovial tissue from RA patients. Our results demonstrate that the orthologous counterparts of murine macrophages exist in humans and are relevant to disease. Our findings may inform the development of novel therapeutic strategies for treating RA. Future studies can explore strategies to control the relative proportions of synovial macrophage subpopulations. Our results indicate the potential to alter the gene regulatory programs of individual macrophage subsets to induce and suppress pro-resolution and inflammatory phenotypes, respectively. Since the model of monocyte infiltration and macrophage expansion is generally relevant to inflammation, the findings from this study can also be applied to better understand the roles of different macrophages subpopulations in inflammatory

diseases of other tissues.

2.3 Critical role of synovial tissue–resident macrophage niche in joint homeostasis and suppression of chronic inflammation

2.3.1 Materials and Methods

Mice (Performed by Qi-Quan Huang)

CD11c-Flip-KO (HUPO) mice are generated by crossing $Flip^{flox/flox}$ mice with CD11c-Cre-GFP transgenic mice ($CD11c^{cre}$) [C57BL/6J-Tg (Itgax-cre,-EGFP) 4097Ach/J, Jackson stock 007567] [239]. Control mice were littermates or age/gender-matched mice that genotyped as $Flip^{flox/+}CD11c^{cre}$ or $Flip^{flox/flox}CD11c^{+}$. HUPO mice $22 \leq$ age ≤ 37 (median = 31) weeks old were used unless otherwise stated. For some experiments, HUPO mice (4 to 16 weeks) with no arthritis or within 1 to 2 or 4 to 6 weeks of arthritis onset were used.

HUPO or control mice on C57BL/6 CD45.1+ background were generated by crossing with CD45.1 congenic strain (B6.SJL-Ptprca Pepcb/BoyJ, Jackson stock 002014). All mice were breed on the C57BL/6 background. All genotyping was performed by polymerase chain reaction using genomic DNA extracted from tail biopsies. CD45.1 or CD45.2 background was determined by flow cytometry. The mice were bred and maintained in the Northwestern University barrier animal facility, and all procedures followed ethical guidelines and approved by Northwestern IACUC.

Clinical evaluation of arthritis (Performed by Qi-Quan Huang)

The incidence and severity of HUPO arthritis were assessed by clinical examination (8, 40).

The clinical score was quantitated from the sum of joint swelling/inflammation (graded 0 to 3 per each limb) and joint deformity (including toe flexion, contraction, and shortening; 0 to 3 per each limb) and the grip strength (0 to 4), the maximum score being 28 [240]. Incidence was defined as at least one swollen, inflamed, or deformed joint.

Flow cytometric immunophenotyping (Performed by Qi-Quan Huang)

Circulating monocytes and the BM precursors, as well as synovial macrophages, were characterized by flow cytometry using multicolor fluorochrome-conjugated antibodies to cell surface and intracellular markers. Blood was drawn by cardiac puncture immediately after euthanizing or submandibular puncture after isoflurane anesthesia. EDTA-anticoagulated whole blood was used for flow. BM cells were isolated from femurs after lysing the red blood cells. Synovial cells were dissected from ankle joints with the methods modified from our earlier publication [225]. Ankles were cut 3 mm above the heel, and skin was removed from the feet. The toes were disarticulated by pulling with blunt forceps, and tibiotalar joint was opened via posterior access to expose the synovial lining. The opened BM cavity in the tibia was thoroughly flushed with Hanks' balanced salt solution to remove BM cells. The dissected joint was incubated in collagenase D (1 mg/ml) for 60 min at 37°C. The released cells were filtered through a 40-mm nylon mesh, and the resulting single-cell suspensions were used.

BM monocyte progenitors and macrophage dendritic cell precursors and BMMs were defined using cocktails containing antibodies to CD117, CD115, CD135, CD11b, and Ly6C gating on the Lin− population (18). Circulating monocytes were defined using cocktails containing antibodies to CD45, CD11b, CD115, Ly6G, Ly6C, CD62L, F4/80, Cx3cr1, and MHCII. Synovial

macrophages were defined using cocktails containing CD45, CD11b, Ly6G, Ly6C, MHC class II (I-A/I-E, MHCII), F4/80, CD64, Siglec F, CD11c, and CX3CR1. Alveolar macrophages were defined as Siglec F+CD64+F4/80+CD11c+CD11b−. Synovial macrophage CD115, CD206, CD163, and TGFBR2 were identified by intracellular staining (41). The live/dead cell marker Aqua (Invitrogen) accompanied every run. Data were acquired on BD LSR II flow cytometer (BD Biosciences), and analysis was performed using FlowJo software (Tree Star Inc.).

For details on in vivo experiments including BrdU (Invitrogen) incorporation assays, monocyte and synovial depletion, and fluroscent Dil-lip uptake, please refer to the original manuscript for full details.

Bone marrow reconstitution and parabiosis (Performed by Qi-Quan Huang)

BM chimeras were established in a CD45.1/CD45.2 mismatched manner. BM recipients, control or HUPO on CD45.1 background were lethally irradiated (γ-radiation, 1100 rads). After 4 hours, 5 × 106 donor whole BM cells from CD45.2 HUPO or control mice were administered by retro-orbital injection. Recipients received sulfamethoxazole (50 mg/ml) and trimethoprim (8 mg/ml) in the drinking water for 8 weeks. Arthritis was evaluated by clinical scoring. Blood and ankles were harvested at 15 weeks and characterized by flow cytometry.

CD45.1/CD45.2 mismatched mice, control-control (three pairs), HUPO-control (six pairs), and HUPO-HUPO (three pairs) of the same gender and age and similar in body weight were paired by parabiosis (42) performed by the Microsurgery Core of Northwestern University. Food and water were supplied at the bottom of cages, and 1-ml saline per mouse was subcutaneously injected

as needed between 1 and 4 weeks after surgery, depending on the body weight. Two weeks after procedures, chimerism was determined by flow cytometry of peripheral blood cells. Arthritis was evaluated by clinical scoring starting 2 weeks after the procedure. Blood and ankles were harvested 5 or 10 weeks after parabiosis, and the cell populations were analyzed by flow cytometry.

Preparation of the RNA-seq library (Performed by Qi-Quan Huang)

HUPO mice with established arthritis and littermate or age/gender-matched control mice were euthanized to harvest blood and ankle cells for RNA-seq. CM and NCM and five subsets of STMs from ankles were stained by multicolor fluorochrome-conjugated antibodies as described and then sorted at the Northwestern University Flow Cytometry Core Facility using a FACSAria III instrument (BD Biosciences). Each HUPO sample was from an individual mouse; however, some ankle samples from control mice were combined because of the limited number of cells. RNA was extracted from each sorted cell population (>400 cell count) using the Arcturus PicoPure RNA Isolation Kit (Applied Biosciences) according to the manufacturer's instructions. Total RNA was used for library construction, which included 49 samples across six control and seven HUPO populations of cells. Full-length cDNA synthesis and amplification were carried out with the Clontech SMART-Seq v4 Ultra Low Input RNA Kit. Subsequently, Illumina sequencing libraries were prepared from the amplified full-length cDNA with the Nextera XT DNA Library Preparation Kit. Before sequencing, the prepared libraries were quantified with Qubit and validated on a Bioanalyzer with a high sensitivity DNA chip. The sequencing of the libraries used an Illumina NextSeq 500 NGS System. Single 75-bp reads were generated with dual indexing, and the libraries were sequenced to an average depth of 21.8 million reads. These procedures were performed at the NUSeq core Facility of Northwestern University.

RNA-seq analysis

The sequencing library was demultiplexed, and the quality of DNA reads was evaluated using FastQC. Adapters were trimmed, and reads of poor quality or aligning to ribosomal RNA sequences were filtered. The sequenced reads were aligned to the Mus musculus genome (mm10) using STAR [124]. Read counts for each gene were calculated using htseq-count [129] in conjunction with a gene annotation file for 23,337 genes obtained from UCSC (University of California Santa Cruz; http://genome.ucsc.edu). Raw gene expression counts were normalized to fragments per kilobase per million (FPKM) using cufflinks [131]. These procedures were performed at the Quantitative Data Science Core of Northwestern University.

Quality control for the 49 samples was performed excluding five low-quality samples (alignment < 82%, duplicates > 66%, and mapped reads < 2.3 × 106) from groups with >3 replicates. To maintain consistent numbers of replicates (n = 3) in each experimental group, we removed the least correlated sample from an additional three control mice under the assumption that the most similar samples better reflect steady-state conditions. We also removed two HUPO samples from F4/80^{hi} subsets that appeared to have been contaminated on the basis of the high expression of neutrophil-specific genes. Thus, 39 samples reflecting three replicates for each of the 13 groups remained. We defined 7778 expressed genes with log2 (FPKM + 1) ≥ 4 in at least 2 of the 39 individual samples. Mixed sexes were used in the experiments, since the HUPO phenotype was not significantly different between male and female mice. Further, removing genes on the sex chromosomes (X and Y) did not affect the results of the transcriptional analysis.

K-means clustering for control mice was performed for genes that fulfilled the following criteria: (1) mean expression across replicates in at least one of the subsets was [log2 (FPKM + 1)] ≥ 4; (2) there was a log2 fold change of ≥1 in one subset compared with any other; and (iii) $P < 0.05$ by analysis of variance (ANOVA) across subsets. DEGs for a given population of myeloid cells between HUPO and control mice were defined by the following criteria: (i) The mean expression for a given population in either group was ≥4; (ii) the magnitude of the log2 fold change between HUPO and control was ≥1; and (iii) $P < 0.05$ by t test between HUPO and control in any one of the five subsets. Because of the low cell numbers, there was no FI2 subset harvested from control mice for RNA-seq analysis. Therefore, the expression of the HUPO FI2 subset is given without a matching control.

The GENE-E software (https://software.broadinstitute.org/GENE-E/) was used for the pairwise Pearson's correlation and K-means clustering analyses, performed using the default settings. PCA was performed using the prcomp function in R with the FPKM matrix of the expressed genes. GO associations were determined by Gorilla [149].

To identify TF binding motifs, we used the findMotifs.pl function with default parameters in the HOMER software package [49]. To identify the potential contributions of TF regulation to HUPO macrophages, we compared the proportion of genes with selected TF binding motifs between the up- or down-regulated genes and those not significantly changed in HUPO macrophage subsets.

Human scRNA-seq analysis

A processed human scRNA-seq dataset was obtained from Zhang et al [227]. The top 20 up- and down-regulated genes ranked by the sum of fold changes in the HUPO FH1 and FH2 subset from Figure 2.3.6A, which have orthologs present in the human dataset, were used for module score calculation performed using FindModuleScore function in Seurat v3.1.0 package [174] with default parameters.

Statistical analysis

All quantitative data are presented as means ± SEM. Statistical analysis between two groups was performed with unpaired two-tailed Student's t test. For multiple comparisons, one-way ANOVA was used followed by Tukey's pairwise mean comparison. Correlations were determined by Pearson's linear correlation. The Bonferroni correction was performed when a single value was compared with multiple variables, and the corrected P value (pc) was presented. To investigate the enrichment of gene sets of interest, the significance of the observed gene numbers was determined by performing permutations (10,000×) to generate putative background distributions. Comparison of module scores and single-cell gene expressions between disease groups was performed with pairwise Wilcoxon rank-sum test. All significance levels were set at $P < 0.05$.

2.3.2 Results

STM subsets during homeostasis

To define STMs, we used an established gating strategy [311, 312], subsetting the CD11b+ population by expression of Ly6C, MHCII, and F4/80 (Figure 2.3.1A). ST F4/80+ macrophages were separated into those that were F4/80^{int} (FI) and F4/80^{hi} (FH). The FI population was then gated into three subsets based on Ly6C and MHCII expression (FI1, FI2, and FI3). We defined the Ly6C+ FI1 and FI2 subsets as STMs to distinguish them from true monocytes in the circulation, BM, and spleen [313]. Supporting this distinction, clear differences in the expression of CD64, F4/80, and MHCII were noted between classical monocytes (CM) and FI1 and FI2 macrophages, although the mean fluorescence intensity of Ly6C was similar (Figure 2.3.1B). The F4/80^{hi} population was further characterized as MHCII− (FH1) and MHCII+ (FH2). Peripheral blood classical and non–classical monocytes (NCM) were identified by subsetting CD11b+CD115+ cells by Ly6C and CD62L (Figure 2.3.1C).

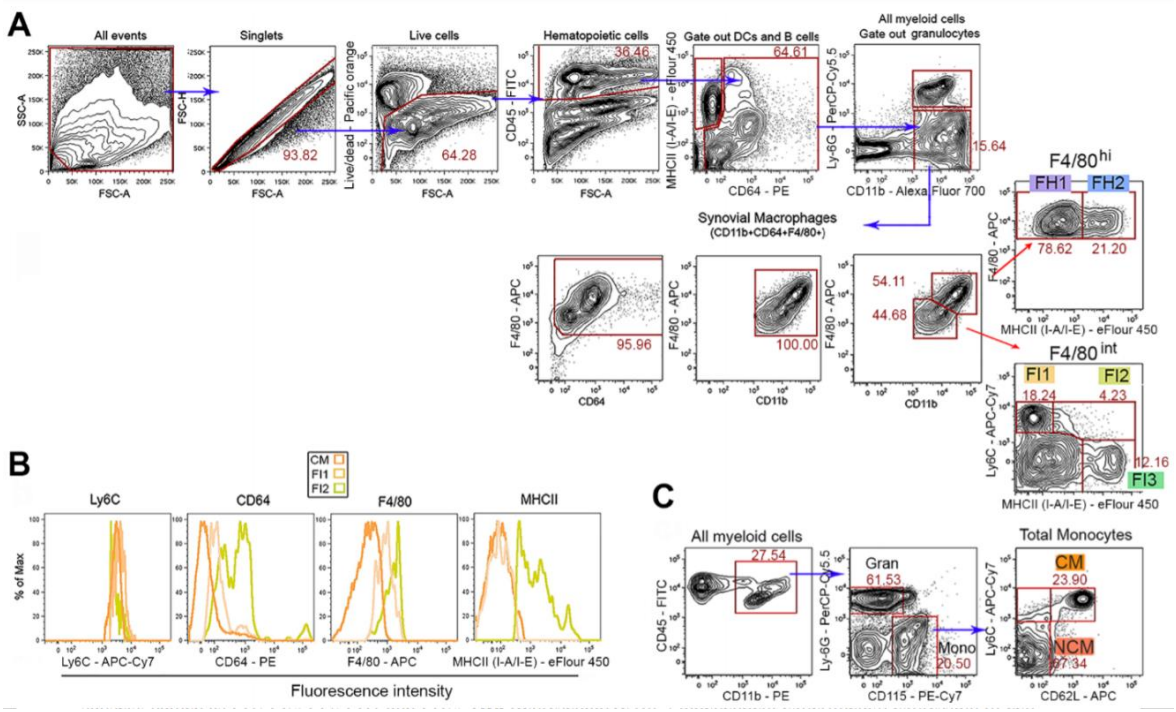


Figure 2.3.1 Gating strategy for synovial tissue macrophage subsets and monocytes during homeostasis defined by flow cytometry. (A) Gating Strategy for identifying

CD11b+F4/80+CD64+ total synovial tissue macrophages and the 5 distinct subsets, named as FI1, FI2, FI3, FH1 and FH2. (B) Representative (of 3 independent experiments) expression of cell surface markers on CM and FI1 and FI2 macrophages by fluorescent intensity. (C) Gating strategy for characterizing blood monocytes.

Distinct patterns of gene expression across subsets of STMs

We next examined transcriptional profiles by bulk RNA-seq. The control FI2 subset was omitted because of low cell numbers. We defined 7778 genes as expressed in this dataset. Principal components analysis (PCA) suggested that each subset represented a distinct population of myeloid cells with CM and NCM from the blood clearly segregating (Figure 2.3.2A). Pairwise correlation between subsets demonstrated a strong association between CM and NCM and a modest correlation of the FI1 synovial macrophage subset and CM (Figure 2.3.2B), consistent with the notion that this population differentiated from CM [26]. Notably, the FH1 subset demonstrates arguably the most distinct expression profile, in that it does not highly correlate to any other subset.

To identify specific patterns of gene expression shared across STM subsets and CM, we performed unsupervised k-means clustering of 1936 genes that were differential across these populations (Figure 2.3.2C). We included CM in the clustering to highlight gene sets that were specific to the STMs and because CM migrate into tissues under homeostatic conditions [311, 312]. We identified six distinct clusters of genes that were predominantly expressed in CM (cluster I), in FI1 (cluster II), or FI3 (cluster III) or shared to varying degrees across FI1, FI3, FH2, and FH1 macrophages (clusters IV to VI). The distinct set of genes associated with each cluster suggests differential functions, ontogeny, and degree of differentiation (Figure 2.3.2C-D). Genes in cluster

VI are expressed in all STM subsets, but not CM, and include generic macrophage functions such as inflammatory response and metabolic processes. In contrast, cluster IV peaks in the FH1 subset and contains genes associated with tissue-specific functions, such as Cfs1r and Il10rb, which are involved with the maintenance of mature macrophages [314]. This is reinforced by the expression of key macrophage maturity genes in cluster IV, such as Lamp1, Rab7, and Vamp3, important for lysosomal function and autophagy, and Msr1 and CD68, scavenger receptors important for maintaining homeostasis.

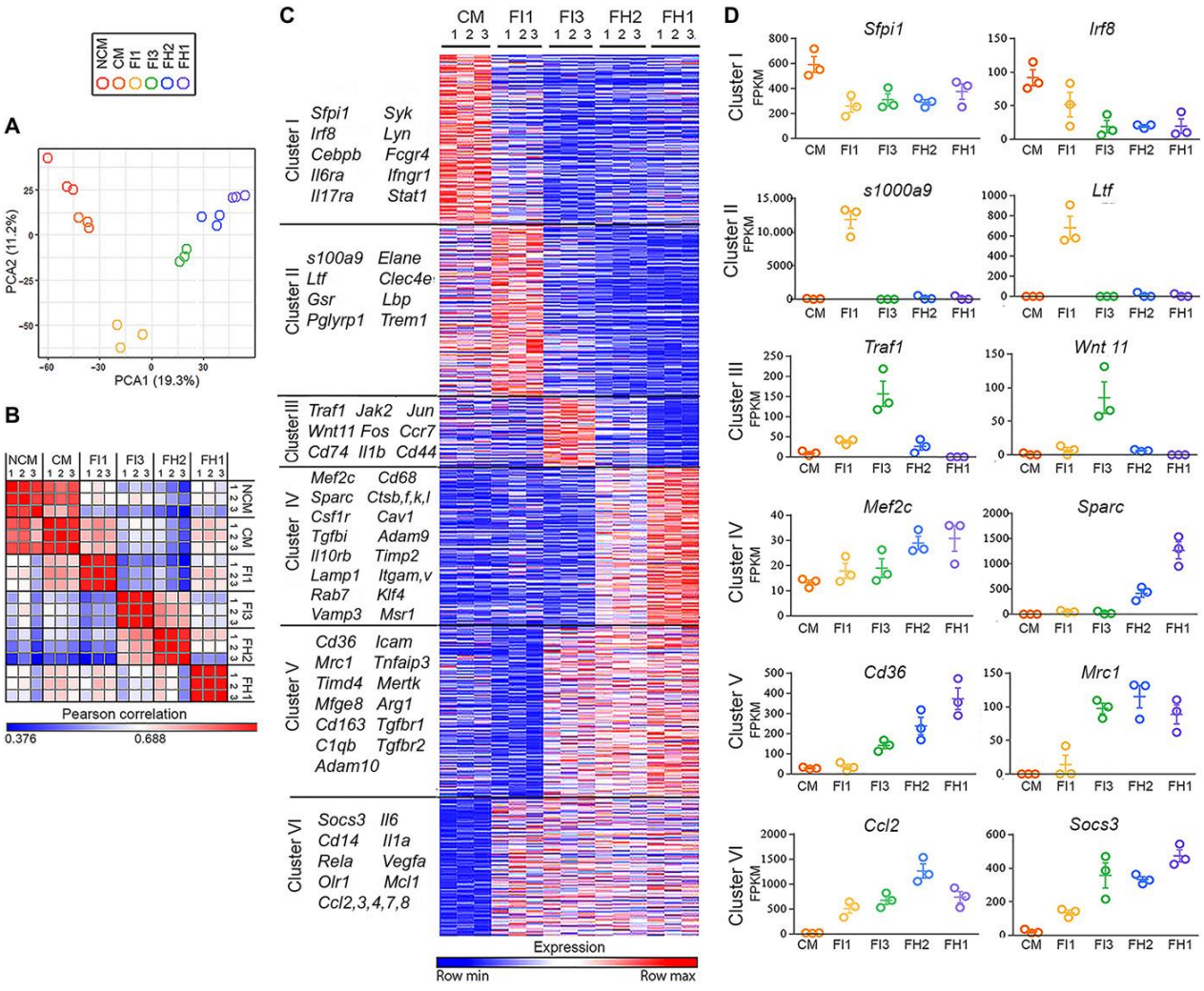


Figure 2.3.2 Distinct patterns of gene expression across subsets of STMs during homeostasis determined by RNA-seq. (A and B) PCA and pairwise Pearson's correlation coefficient of gene expression (total 7778 genes expressed) from wild-type mice, across individual samples from blood CM and NCM and STM subsets (FI1, FI3, FH1, and FH2). (C) K-means clustering of 1936 differentially expressed genes (DEGs) across CM and the STM subsets. (D) Representative examples of RNA expression from genes identified in each of the six clusters in (C), presented as the means ± 1 SE in fragments per kilobase per million (FPKM).

STMs are increased in HUPO mice, while the FH1 subset is reduced

Next, we examined STMs in HUPO mice and age-matched, littermate controls, at 22 to 37 weeks of age, when arthritis progression stabilized. The majority of F4/80+ macrophages of control mice were in the synovial lining, and they were increased in the both the lining and sublining of HUPO mice with arthritis (Figure 2.3.3A). The patterns of the populations of STMs defined by flow cytometry were distinct between HUPO mice with or without arthritis and age-matched littermate controls (Figure 2.3.3B). The total number of CD11b+CD64+F4/80+ macrophages was increased in HUPO mice with arthritis, and the percentages of the FI1 through FH2 subsets were variably increased in HUPO mice with arthritis (Figure 2.3.3C). In contrast, FH1 STMs were significantly reduced in HUPO mice with arthritis, compared to those without arthritis or controls, and this subset inversely correlated with arthritis score and duration, while granulocytes, but not B cells, were positively associated (Figure S2.3.3A). The expression of MHCII was higher and F4/80 was lower on HUPO FH1, compared with control macrophages (Figure S2.3.2B), suggesting reduced differentiation of the HUPO FH1 macrophages that were present. Although evaluation of every tissue was not performed, CD11c+ macrophages were also

reduced in the lungs of HUPO mice (Figure S2.3.3C). The FH1 subset was reduced in young (4 to 16 weeks) HUPO mice before the onset of arthritis on clinical exam and before a significant increase of total STMs or the FH2 subset (Figure 2.3.3D). Further, on histologic exam, low levels of inflammation and joint damage were observed before the onset of clinical arthritis, similar to preclinical or early RA (14), which increased over time, as arthritis severity increased (Figure 2.3.3E-F). These observations suggest that the reduction of the FH1 subset in HUPO mice may be associated with the development of arthritis. Of potential relevance to the increase of HUPO arthritis incidence with age (peak week, ~22), under homeostatic conditions, STMs decreased with age, and the FH1:FH2 ratio was ($P < 0.01$) reduced in older mice (Figure S2.3.3D).

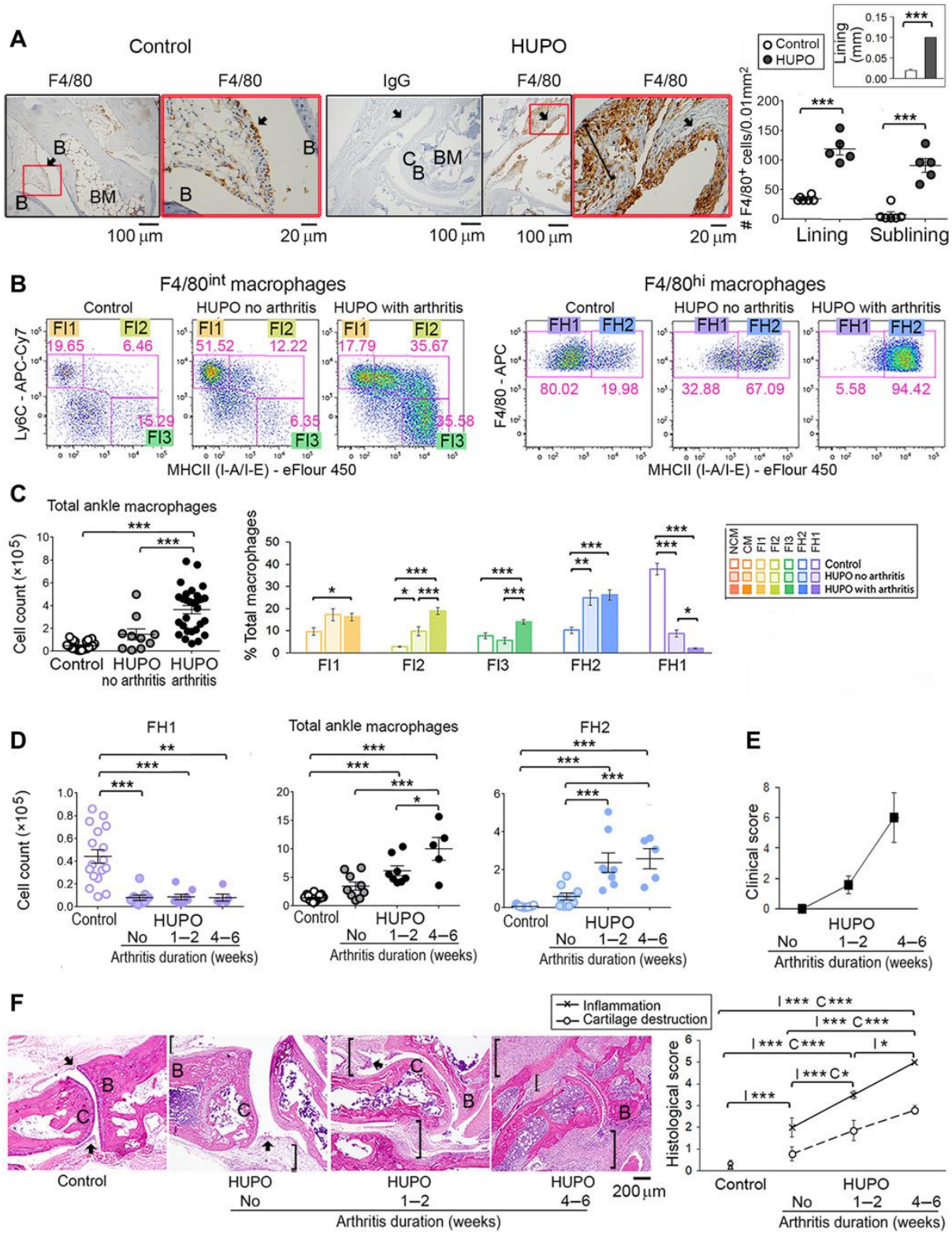

A
Control
HUPO
F4/80
F4/80
IgG
F4/80
F4/80
B
BM
100 μm
20 μm
Control
HUPO
Lining (mm)
F4/80+ cells/0.01mm2
Lining
Sublining
B
F4/80int macrophages
F4/80hi macrophages
Control
HUPO no arthritis
HUPO with arthritis
FI1
FI2
FI3
FH1
FH2
19.65
6.46
15.29
51.52
12.22
6.35
17.79
35.67
35.58
80.02
19.98
32.88
67.09
5.58
94.42
Ly6C - APC-Cy7
F4/80 - APC
MHCII (I-A/I-E) - eFlour 450
C
Total ankle macrophages
Cell count (×10^5)
Control
HUPO no arthritis
HUPO arthritis
% Total macrophages
NCM
CM
FI1
FI2
FI3
FH2
FH1
D
FH1
Total ankle macrophages
FH2
Cell count (×10^5)
Control
HUPO
No
1–2
4–6
Arthritis duration (weeks)
E
Clinical score
F
Inflammation
Cartilage destruction
Histological score
Control
HUPO No
HUPO 1–2
HUPO 4–6
200 μm
Arthritis duration (weeks)

Figure 2.3.3 Altered histology and STM subsets during chronic arthritis and before disease onset in HUPO mice. (A) Immunohistochemistry of ankle joints using anti-F4/80 or control IgG for control or HUPO mice with arthritis. Right panel presents number of F4/80+ cells/0.01 mm2 and average lining thickness. (B) Representative flow cytometry of STMs from control and HUPO mice. (C) Number of total CD11b+CD64+F4/80+ STMs (left) and frequency of each subset in control and HUPO with or without arthritis mice. (D) The number of FH1, total, and FH2 STMs for mice comparing control and HUPO mice with or without arthritis identified between 1 to 2 and 4 to 6 weeks. (E) Clinical scores and (F) histologic examination for young mice without or with arthritis. B: bone; C: cartilage; arrows identify synovial lining, and brackets identify the sublining in (A) and (F). Statistical analyses were performed by Student's two-tailed t test for (A) and one-way analysis of variance (ANOVA) plus Tukey. *P < 0.05, **P < 0.01, and ***P < 0.001 among indicated groups. I indicates inflammation and C indicates cartilage destruction in (F).

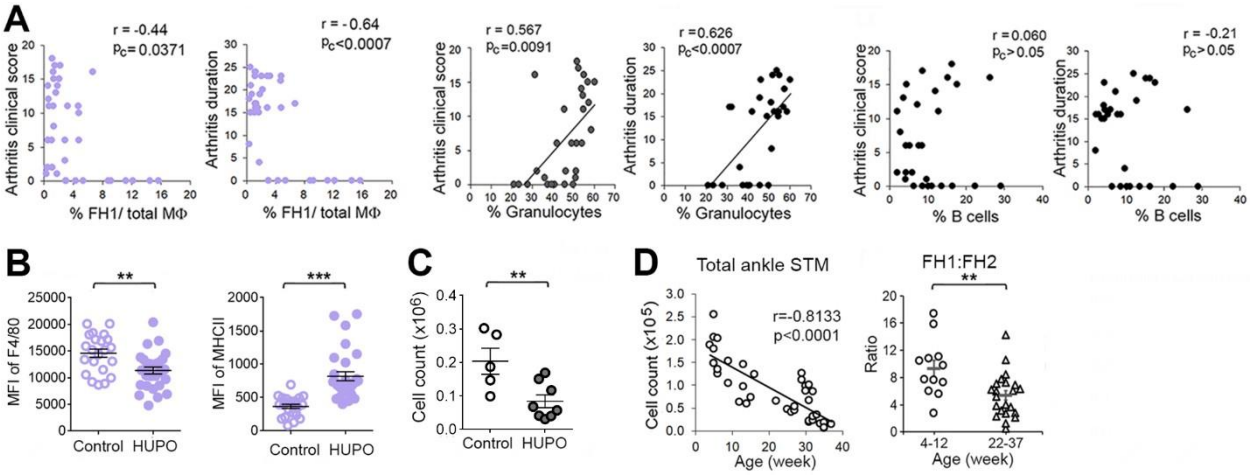


Figure S2.3.3 (A) Pearson's correlations of FH1 macrophages, granulocytes and B cells to arthritis clinical scores and arthritis duration in HUPO mice as indicated in each panel. (B) Differential

expression of MHCII and F4/80 on HUPO and control FH1 macrophages. (C) Alveolar macrophages, defined as SiglecF+CD64+CD11c+F4/80+CD11b-, from control and HUPO mice (B-C). (D) The number of total STM from control mice, and the FH1 to FH2 ratio for the indicated age groups. Statistical analyses were determined by Pearson's linear correlation for (left panel of D) or by 2-tailed Student's t- test performed for (B, C and right panel of D). ** $p< 0.01$ and *** $p< 0.001$ between indicated groups.

HUPO monocytes exhibit proinflammatory potential

Utilizing mainly 5-bromo-2'-deoxyuridine (BrdU) incorporation experiments, we found that migration of circulating monocytes, rather than local proliferation, is responsible for the increase of STMs in HUPO mice. The monocytes are able to readily differentiate into all subtypes of $F4/80^{hi}$ STMs by filling the vacant niches, although the FH1 subset remained reduced. We also confirmed that targeting Flip in STMs did not alter the apoptosis and proliferation of FH subsets during inflammation (See complete manuscript for full details). We therefore examined the circulating monocytes to better understand how they might contribute to HUPO arthritis. The total number of monocytes, specifically the CM, were significantly ($P < 0.001$) increased in HUPO mice with arthritis, and the expression of MHCII was increased ($P < 0.001$) on HUPO CM and NCM (Figure 2.3.4A). Of the 4480 genes expressed in either HUPO or control CM or NCM, 150 differentially expressed genes (DEGs) were up-regulated in both populations in HUPO mice, while 139 were down-regulated in both (Figure S2.3.4B). Of the DEGs in CM and NCM of HUPO mice, those up-regulated in both were enriched in the Gene Ontology (GO) pathway functions that may promote inflammation and adaptive immunity (Figure 2.3.4A), while no GO pathways exhibited a false discovery rate (FDR) P value < 1 for other combinations of genes up or down. In addition,

supporting the role of adaptive immunity in the progression of arthritis, although conventional dendritic cells (cDC) were reduced in the spleens of HUPO mice [240], dendritic cells (DC), identified as CD45+CD11b+MHCII+CD64−, were increased in the ST of HUPO mice with arthritis, compared to those without arthritis or controls (Figure 2.3.4D). Further, at 4 weeks before the onset of arthritis on clinical exam, cartilage proteoglycans were reduced in the cartilage of HUPO joints (Figure 2.3.4E), providing a potential antigenic source that may contribute to the anti-aggrecan antibodies previously observed in progressive HUPO arthritis [240].

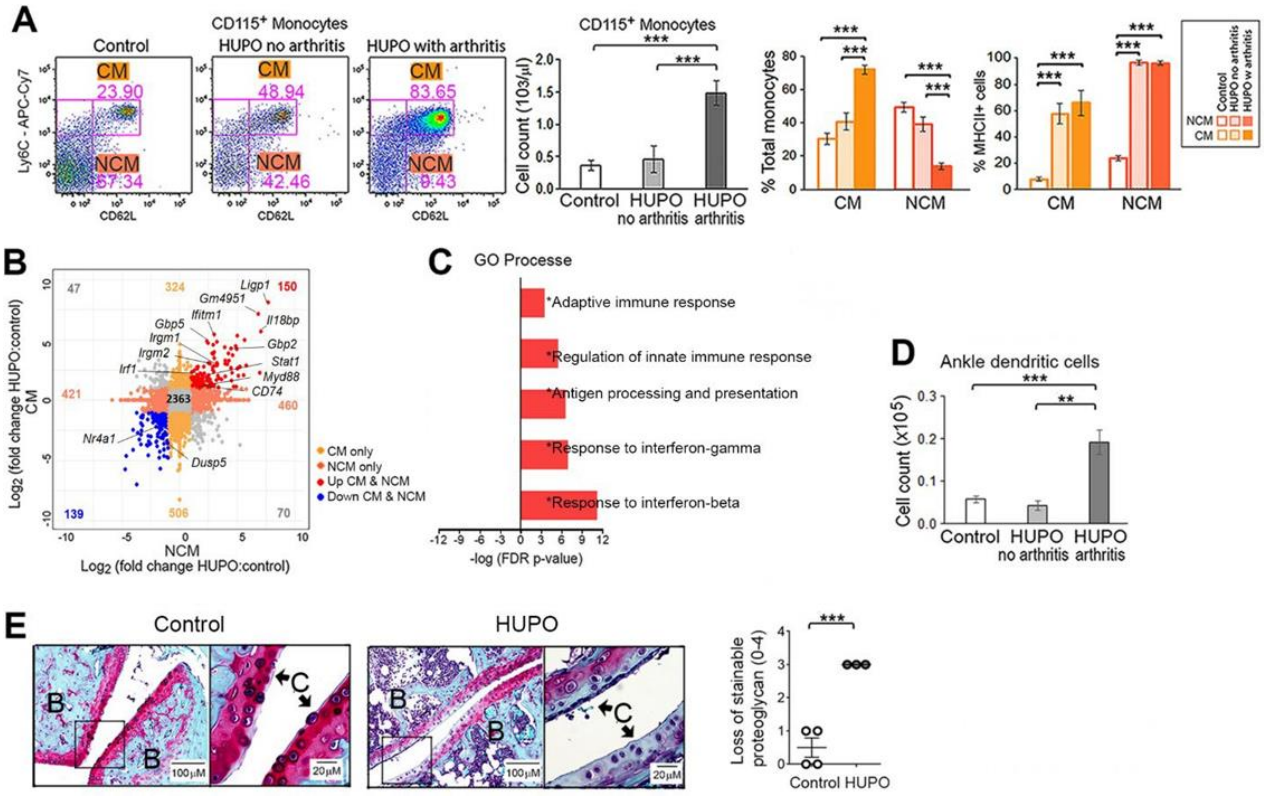


Figure 2.3.4 Increased circulating monocytes in HUPO mice exhibit arthritogenic potential. (A) Analysis of circulating monocytes. From left to right: the representative flow cytometry, the numbers of total monocytes, the frequency of CM and NCM and the MHCII+ CM and NCMs. (B)

Scatterplot showing fold change of genes expressed in HUPO or control, CM or NCM. DEGs are presented in colors, and the numbers of genes in each sector are identified. (C) The top GO terms for DEGs in red from panel B. All GO terms for the genes in other sectors were not significantly enriched. (D) DCs identified as CD45+CD64-MHCII+CD11b+, in HUPO compared with control joints. (E) Cartilage proteoglycans were identified by Safranin-O staining of ankle sections from 4-week-old control or HUPO mice. The areas in black boxes were enlarged in panels on the right. Data presented as loss of proteoglycan on a 0-4 scale. B= bone, C and arrows indicate cartilage. Values are mean ±1 SE. Statistical analyses by one-way ANOVA with Tukey adjustment (A, D) and by 2-tailed Student's t-test (E). ** $p< 0.01$ and *** $p< 0.001$ between the indicated groups.

Synovial macrophage subsets in HUPO mice are more similar to CM

Next, we isolated and performed RNA-seq on circulating monocytes and STMs from HUPO mice with arthritis. On the basis of global gene expression, we observed that the HUPO FI2, FI3, FH2, and FH1 subsets appear more closely related to each other (Figure 2.3.5A) than in controls (Figure 2.3.1A). In addition, when comparing each of the HUPO subsets to their control counterparts, we observed that FI1 subsets, which maintain their already high monocyte similarity, were only marginally different between HUPO and control mice (Figure 2.3.5B). In contrast, HUPO FH2 appears to be the most altered subset from its control counterpart, consistent with a high monocyte replacement in the HUPO FH2 population. In support of the monocyte origin of HUPO macrophages, the CM genes that were up-regulated in at least three of four HUPO macrophage subsets (n = 128) were enriched for interferon regulatory factor (IRF), PU.1-IRF, and CCAAT-enhancer-binding protein (C/EBP) binding motifs (Figure S2.3.6A).

To further determine how STMs differ between health and chronic inflammation, we defined 2021 DEGs, up- or down-regulated between HUPO and control mice in at least one of the myeloid subsets. We identified six HUPO-signature (HS) clusters demonstrating distinct expression patterns across STM subsets compared with control mice (Figure 2.3.5C). The HUPO FI2 subset, with no control counterpart, was included in the clustering. HS clusters I, IV, and V demonstrate expression patterns that were decreased in one or more HUPO subsets. Notably, cluster V genes, with decreased expression in HUPO FH2 and FH1 subsets, were consistent with functions associated with macrophage tissue residency during homeostasis (Figure 2.3.5C-D). In contrast, clusters II, III, and VI contain genes that generally were increased in expression in HUPO mice, which include GO pathways enriched for genes involved in inflammation, adenosine triphosphate metabolic processes, and glycolysis (Figure 2.3.5E).

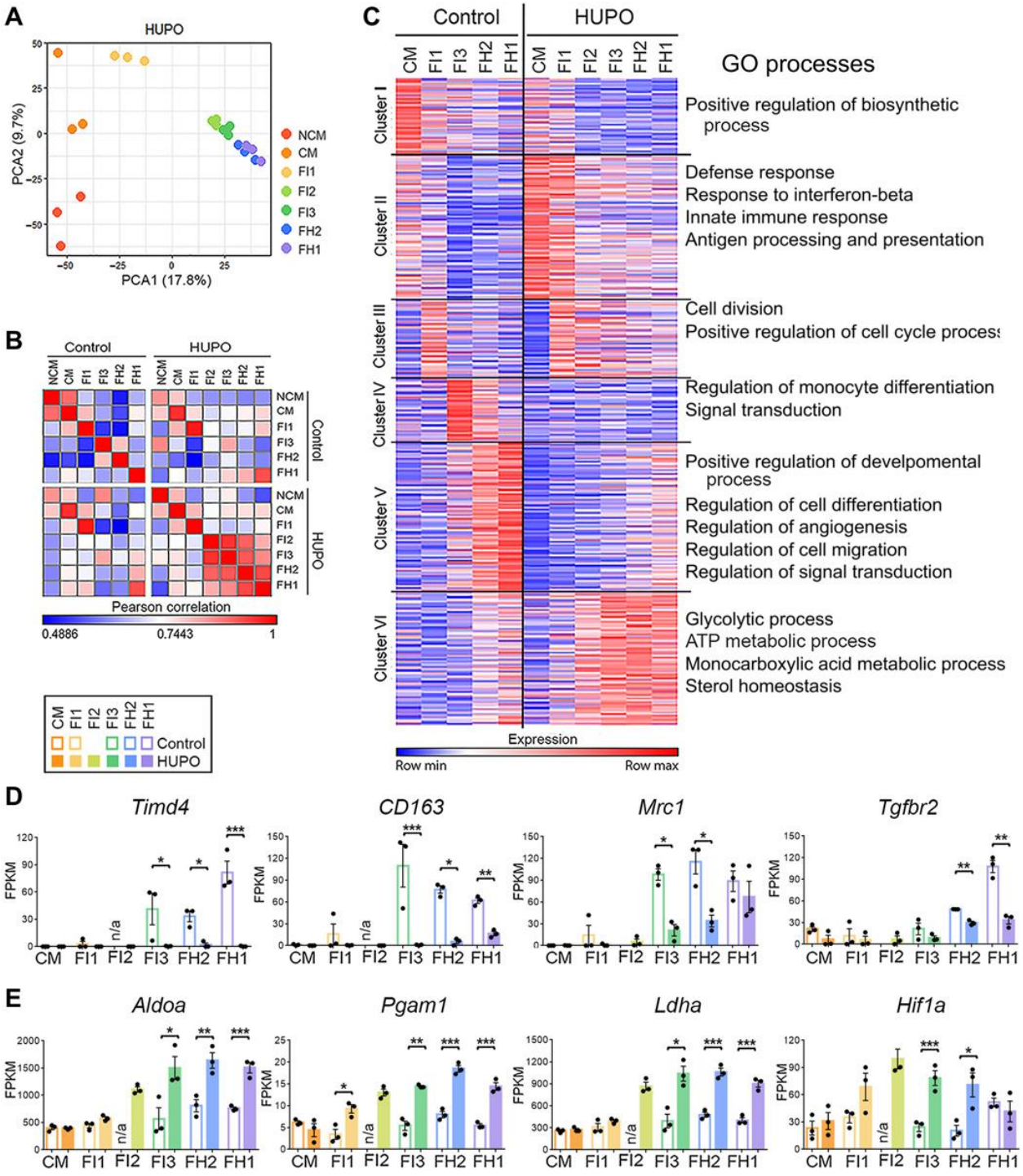


Figure 2.3.5 Transcriptional profiling supports the origin of HUPO STMs as circulating monocytes and identifies functional differences. (A) PCA of gene expression for individual

samples from HUPO mice. (B) Comparison of Pearson's correlation coefficient of mean gene expression between each cell population from control and HUPO mice (n = 3). Total genes expressed (n=7778) were used for (A) and (B). (C) K-means clustering of DEGs (n = 2021) between control and HUPO subsets, identifying HUPO-signature (HS) genes. The genes from HUPO FI2 subset, which do not have a control counterpart, were included in the clustering. (D) Examples of genes important in macrophage efferocytosis and tissue residency, identified in (C), cluster V. (E) Examples of genes involved in the glycolytic pathway, identified in (C), cluster VI. Statistical analysis was performed by two-tailed Student's t test (D and E). Individual values and the means ± 1 SE are presented. *P < 0.05, **P < 0.01, and ***P < 0.001 among indicated groups.

The tissue-resident macrophage phenotype is lost in HUPO and replaced by a proinflammatory profile

To further define the distinction between HUPO and control F4/80^{hi} macrophages, we compared the genes expressed (4500) in either FH1 or FH2 subset in HUPO mice or controls (Figure 2.3.6A). DEGs upregulated in the HUPO FH1 subset alone included IL1b and neutrophil chemokine genes (Cxcl1, Cxcl2, and Cxcl3), which may contribute to the neutrophil recruitment into the HUPO joints [240]. GO pathways increased in HUPO FH1 and FH2 macrophages were significantly enriched in genes involved in leukocyte migration and innate immune response (Figure 2.3.6B). A significant number (173, P < 0.0001) of genes were downregulated in both HUPO FH1 and FH2, including those important in maintaining macrophage tissue residency such as Csfr1, Cx3cr1, Timd4, and Vsig4. Further, proteins expressed by DEGs important in macrophage tissue residency were also reduced in the HUPO FH1 and/or FH2 subsets by flow cytometry (Figure 2.3.6C). In addition, Mef2c was significantly reduced in HUPO FH2

macrophages, and DEGs down-regulated in both the HUPO FH1 and FH2 populations were enriched for myocyte enhancer factor 2 (MEF2) transcription factor (TF) binding motifs compared with those genes not down-regulated in either (Figure S2.3.6B-C). These observations are in line with role of MEF2C in regulating the macrophage tissue–resident identity [47, 157]. Notably, of the 356 genes from Fig. 2.3.1C, cluster IV that we previously associated with macrophage tissue–resident identity, 57 ($P < 0.0001$) were reduced in both HUPO FH1 and FH2 subsets (Figure 2.3.6D), suggesting that the macrophage tissue residence phenotype was robustly diminished in HUPO mice.

To identify a mechanism for the reduction of the FH1 subset in HUPO mice, we examined the ability of phagocytosis of liposomes, as a surrogate for phagocytosis of apoptotic cells, to promote the differentiation of FH2 macrophages, determined by the reduction of MHCII expression. This approach was chosen since the expression of genes for receptors contributing to efferocytosis was reduced in HUPO FH2 macrophages (Figure 2.3.5D). Dil-lip injected into the ankles of control mice resulted in the reduction of MHCII expression on FH2 macrophages that phagocytosed Dil-lip compared to those that did not. No reduction of MHCII was observed on HUPO FH2 macrophages that phagocytosed Dil-lip, and this was significantly different ($P < 0.001$) compared with the controls (Figure 2.3.6E). These observations directly demonstrate that phagocytosis of liposomes was capable of promoting macrophage differentiation under homeostatic conditions, which was not observed with HUPO FH2 macrophages.

To determine whether the differences observed in FHI and FH2 macrophages from HUPO-associated arthritis were conserved in human disease, we compared gene expression in scRNA-

seq previously performed on RA synovium [227]. Overall, we found that of the 20 genes that were most up- or down-regulated in HUPO F4/80^{hi} macrophages, the module scores of those up-regulated were modestly increased, while those down-regulated were significantly reduced, in monocytes from leukocyte-rich RA ST (Figure 2.3.6F-G, S2.3.6D-E). Specifically, we noted that the expression of genes associated with increased HUPO expression, including S100A9 and VEGFA, was higher in ST monocytes from patients with leukocyte-rich RA ST (Figure 2.3.6F). On the other hand, genes associated with decreased HUPO expression, such as CX3CR1, TIMD4, VSIG4, and TGFBR2, were significantly decreased in expression in leukocyte-rich RA ST, suggesting that the loss of tissue residence phenotype is conserved in highly inflammatory human disease (Figure 2.3.6F). Together, these observations suggest that the reduction of the FH1 subset in HUPO mice emptied a niche, into which monocyte-derived FH2 macrophages entered but, in the absence of Flip and under chronic inflammatory conditions, demonstrated limited ability to differentiate into bona fide tissue-resident macrophages, capable of suppressing chronic inflammation, which may be relevant to the pathogenesis of RA.

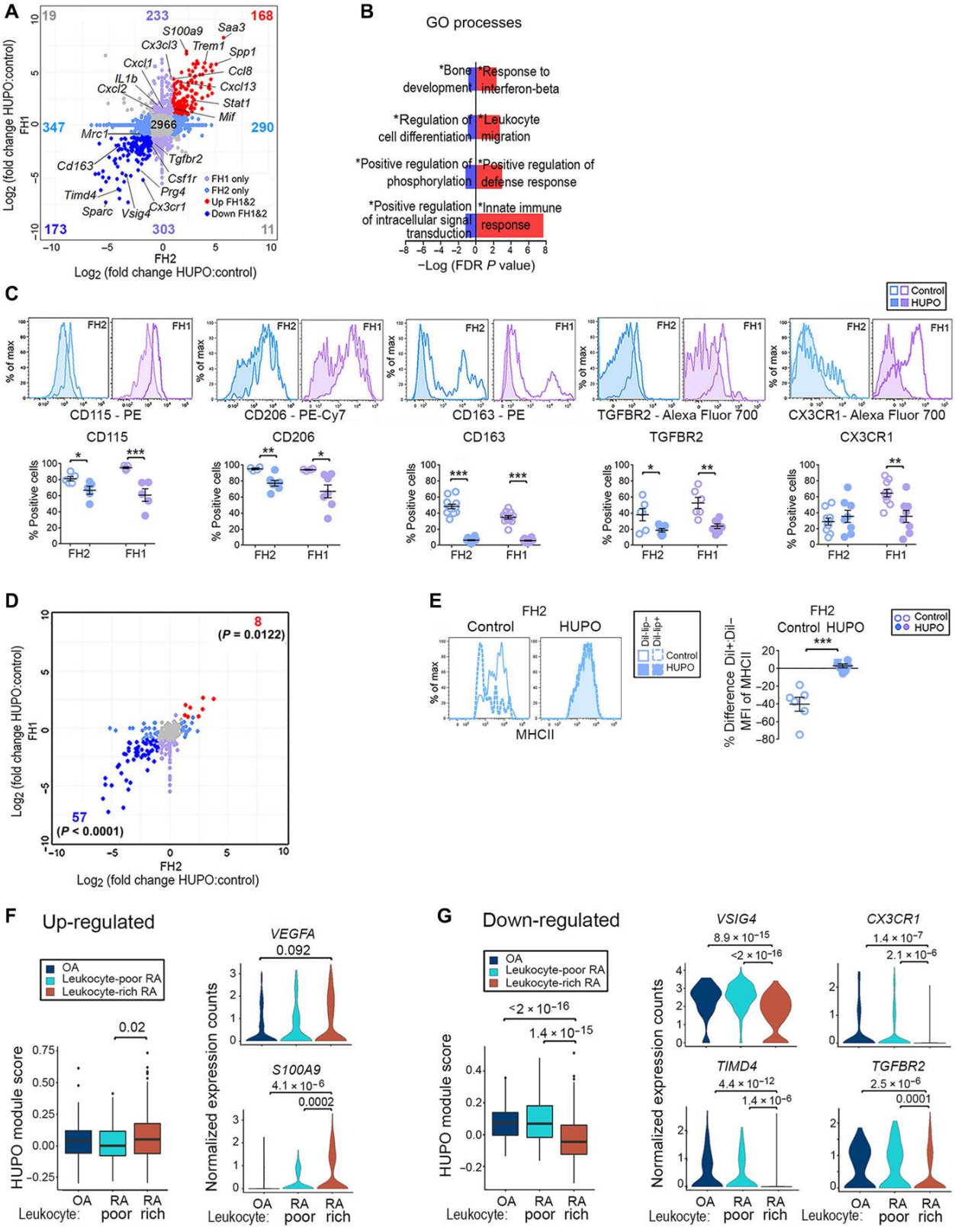

A
19
233
168
347
2966
290
173
303
11
S100a9
Saa3
Cx3cl3
Trem1
Spp1
Cxcl1
Ccl8
IL1b
Cxcl13
Cxcl2
Stat1
Mif
Mrc1
Tgfbr2
Cd163
Csf1r
Prg4
Timd4
Cx3cr1
Sparc
Vsig4
FH1 only
FH2 only
Up FH1&2
Down FH1&2
Log2 (fold change HUPO:control) FH1
Log2 (fold change HUPO:control) FH2
B
GO processes
*Bone development
*Response to interferon-beta
*Regulation of cell differentiation
*Leukocyte migration
*Positive regulation of phosphorylation
*Positive regulation of defense response
*Positive regulation of intracellular signal transduction
*Innate immune response
−Log (FDR P value)
C
Control
HUPO
FH2
FH1
% of max
CD115 - PE
CD206 - PE-Cy7
CD163 - PE
TGFBR2 - Alexa Fluor 700
CX3CR1- Alexa Fluor 700
CD115
CD206
CD163
TGFBR2
CX3CR1
% Positive cells
D
8
(P = 0.0122)
57
(P < 0.0001)
Log2 (fold change HUPO:control) FH1
Log2 (fold change HUPO:control) FH2
E
FH2
Control
HUPO
Dil-ip- Dil-ip+
MHCII
% Difference Dil+:Dil- MFI of MHCII
F
Up-regulated
OA
Leukocyte-poor RA
Leukocyte-rich RA
0.02
HUPO module score
Leukocyte:
VEGFA
0.092
S100A9
4.1 × 10−6
0.0002
Normalized expression counts
G
Down-regulated
<2 × 10−16
1.4 × 10−15
VSIG4
8.9 × 10−15
<2 × 10−16
CX3CR1
1.4 × 10−7
2.1 × 10−6
TIMD4
4.4 × 10−12
1.4 × 10−6
TGFBR2
2.5 × 10−6
0.0001

Figure 2.3.6 The tissue-resident macrophage phenotype is lost in F4/80hi HUPO STMs and in RA ST. (A) Scatterplot showing fold change of genes expressed in HUPO/control FH1/FH2 subsets. DEGs and their numbers are in colors, and select genes are labeled. (B) Top GO terms for genes in red and dark blue from (A). (C) Flow cytometry for protein expression of genes downregulated in HUPO mice from (A), as MFI and % positive. (D) Overlap of 356 genes from the control STM cluster IV in Figure 2.3.1C with genes in (A). (E) Following (16 to 40 hours) ankle injection of Dil-lip, control, and HUPO cells were harvested and MHCII on the FH2 subsets determined for Dil+ and Dil− cells. *P < 0.05, **P < 0.01, and ***P < 0.001 among indicated groups, for C and E. (F and G) Module scores (box plots) of top 20 up-regulated (F) or 20 down-regulated genes (G) in HUPO FH1 and FH2 (A) and the representative genes (violin plots) expressed in osteoarthritis (OA) and RA leukocyte–poor or RA leukocyte–rich synovial tissue [227]. Statistics: Two-tailed Student's t test (C and E); comparison with expected background distribution from 10,000 permutations (A and D) and pairwise Wilcoxon rank (F and G). Absolute P values between groups (F and G).

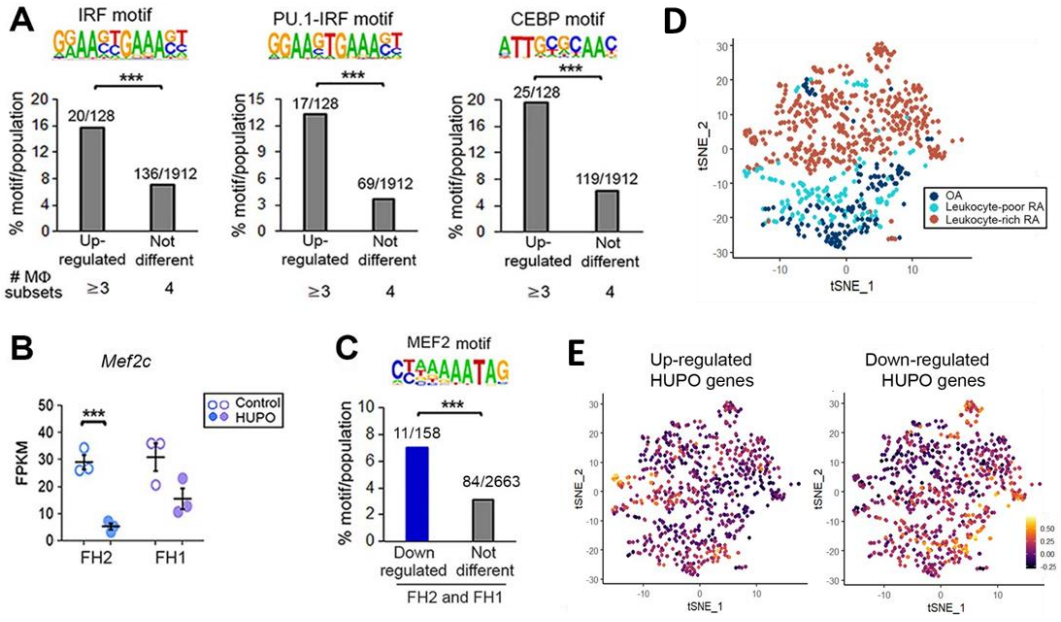


Figure S2.3.6 (A) The percent of monocyte genes that contain an instance of the given TF binding motif that were either up-regulated in at least 3 HUPO synovial tissue macrophage subsets (n=128) or not up-regulated in any subset (n= 1912). (B) Mef2c gene expression in HUPO or control FH1 or FH2 subsets. (C) The percent of genes containing an instance of the MEF2 binding motif in HOMER recognized genes that were either downregulated in both HUPO FH1 and FH2 (from dark blue in Figure 2.3.6A) or not differential in either (from grey in Figure 2.3.6A). (D) tSNE plot of human synovial monocytes categorized by disease types [227]. (E). Module scores of top 20 up- and down-regulated FH1/2 genes in HUPO, for the OA and RA synovial tissue monocytes. Statistical analysis was performed by 2-sided Student's t-test. *** $p < 0.001$ among indicated groups.

2.3.3 Discussion

Our observations suggest a novel mechanism for the transition from preclinical disease to active RA. During homeostasis, the FH1 subset was the dominant tissue-resident population, while the smaller FH2 subset was the F4/80^{hi} population that was slowly proliferating. In HUPO mice, macrophages are necessary for the initiation of disease, while T and/or B lymphocytes are required for progression [240]. Here, we demonstrate that the HUPO FH1 subset was reduced before and after disease onset, while the other subsets were variably increased shortly after arthritis onset and during chronic inflammation. BrdU labeling, transcriptional profiling, and parabiosis experiments established that circulating monocytes readily differentiated into F4/80^{hi} macrophages in HUPO mice, which was minimal during homeostasis, but with the reduction of Flip, in the environment of chronic inflammation, exhibited limited ability to differentiate into bona fide FH1 tissue–resident macrophages.

Recent studies demonstrate that, like most tissue-resident populations with the exception of microglia, which renew from fetal hematopoietic stem cells, CX3CR1+ synovial lining macrophages derive during early embryonic development [220, 311]. Our analysis demonstrates that the FH1 and FH2 populations represent long-lived tissue-resident macrophages, which maintain their population in steady state via local proliferation of the FH2 subset, with minimal replenishment from circulating monocytes. These findings support recent observations that identify CX3CR1+ lining and CX3CR1− interstitial STMs, consistent with our CX3CR1+ FH1 and CX3CR1^{lo} FH2 subsets, as the populations maintained with limited contribution from circulating monocytes [220]. Further, supporting the similarity of the interstitial CX3CR1− and FH2 populations, each was the primary subset that proliferated under homeostatic conditions. MHCII expression on most F4/80^{hi} populations of tissue-resident macrophages is relatively homogeneous, except for dermal macrophages, which demonstrate low and high MHCII subsets [311], similar to the STMs under steady-state conditions. Previous reports have shown that phagocytosis of apoptotic cells facilitates the differentiation of the MHCII+ subset to become MHCII− and is important in maintaining tissue residency and tissue homeostasis [225, 315, 316]. The differentiation of CX3CR1−MHCII+ macrophages into CX3CR1+ synovial lining macrophages [220] is consistent with the transition of FH2 into FH1 macrophages under homeostatic conditions. We speculate that the CD45+ cell apoptosis noted in the joints of the wild-type mice may contribute to this process through efferocytosis [316]. Our observations suggest the synovial tissue resident macrophages function to suppress chronic inflammation, not only serving as a barrier and dampening serum transfer–induced arthritis [220] but also capable of suppressing chronic inflammation following recruitment and differentiation of Flip-replete monocytes to the synovium.

In contrast to homeostatic conditions, the BrdU and parabiosis experiments demonstrated that HUPO circulating monocytes entered the joints readily differentiating into F4/80^{hi} macrophages; however, the transition from FH2 to fully functional FH1 cells was impaired. The lack of enhanced local proliferation of the F4/80^{hi} HUPO macrophages 30 min after BrdU administration is distinct from observations in acute serum transfer–induced arthritis [220] and the early inflammatory phase of atherosclerosis in which monocytes are recruited, differentiate into macrophages, and expand by proliferation [317]. Further, bacterial infection results in Kupffer cell necroptosis and recruited monocytes proliferate, differentiating into tissue-resident macrophages [66]. In contrast, in HUPO arthritis, which is chronic, the influx of monocytes primarily accounted for the expansion of the FH2 population. The reduction of FI3 and FH2 HUPO macrophages, concurrent with the reduction of NCM, but not CM, following the injection of clodronate liposomes, suggests that these macrophage subsets derive from NCM, although this was not directly documented. The HUPO monocytes entering the joints were highly enriched in pathways of innate immunity, antigen presentation, and interferon signaling, which, together with the increase of DCs in the HUPO synovial tissue and the systemic reduction of regulatory T cells present in HUPO mice, contribute to the development of autoantibodies to joint constituents such as aggrecan and the progression of arthritis [240]. This scenario is distinct from the setting where the niche is disrupted under noninflammatory conditions, into which homeostatic monocytes enter, acquire a tissue resident phenotype, and restore the niche [46, 69].

In HUPO mice with arthritis, the FH1 population was greatly reduced, while the FH2 population was expanded. Although Flip was reduced in both FH1 and FH2 macrophages under inflammatory conditions, neither population exhibited increased apoptosis compared with control

mice. Previously, we demonstrated that apoptosis of human monocytes was rescued by the increase of FLIP observed during in vitro differentiation into macrophages [318]. The reduction of FLIP in human in vitro–differentiated macrophages results in Fas-mediated apoptosis. In contrast, using lineage negative murine BM progenitors, we demonstrated that the reduction of Flip prevented macrophage differentiation in vitro, which was not associated with increased apoptosis and was not rescued by caspase inhibitors [239]. Together, these observations support the potential role of Flip in monocyte to tissue-resident macrophage differentiation, under chronic inflammatory conditions, which promoted macrophage survival despite the reduction of Flip.

In HUPO mice, the FH1 subset replenished to a limited degree by circulating monocytes, determined by BrdU and parabiosis, and demonstrated a restricted tissue-resident transcriptional profile. A potential mechanism preventing HUPO F4/80^{hi} macrophages from attaining tissue residency may be the reduction of the requisite genes. For example, CD115/CSFR1 was reduced in HUPO F4/80^{hi} macrophages, together with TFs downstream of CD115, such as Klf2 and Klf4, which are involved in the development of macrophage tissue residency and the ability to silently clear apoptotic cells [316]. Efferocytosis is known to drive the differentiation of monocytes to tissue-resident macrophages under homeostatic conditions, in an acute self-limited model of arthritis and following myocardial ischemia [225, 315, 316]. Our data suggest that decreased phagocytosis of apoptotic cells may also contribute to the decreased FH1 population in HUPO mice. Receptors responsible for efferocytosis including CD206 and CD163, which are highly expressed on phagocytic macrophages, were reduced in HUPO FH2 macrophages compared with controls, and apoptotic CD45+ cells were greatly reduced in the chronic inflammation of the HUPO joints. The forced reduction of CD206 reduces efferocytosis [315], and mice deficient in

CD206 develop more severe experimental arthritis [319]. CD163, a scavenger receptor, is important in suppressing inflammation [320]. Our observations, combined with the parabiosis data, demonstrate that, although the molecular mechanism remains to be defined, Flip was necessary, but not sufficient, for the differentiating monocytes to fully acquire a tissue-resident, homeostatic phenotype during chronic inflammation.

Our observations concerning the role of STMs and circulating monocytes in the initiation and progression of arthritis in HUPO mice appear relevant to the pathogenesis of RA, although FLIP is highly expressed in RA STMs [321]. The mechanisms targeting ST in the transition from preclinical to clinical disease are not known. Our data suggest a novel mechanism, which reduction of tissue-resident macrophages predisposes joints to the development of arthritis, by opening a niche permitting the influx of activated monocytes and other inflammatory cells. Consistent with this notion, the FH1 subset was reduced before the onset of clinical arthritis and restoration of this subset from wild-type mice suppressed inflammation in HUPO mice. HUPO monocytes expressed a transcriptional profile enriched in pathways that support inflammation and adaptive immunity, which likely contributed to the progression of disease. Similarly, monocytes from patients with RA are enriched in pathways involved in interferon signaling, inflammatory response, and anti-apoptosis [322], supporting a role for activated monocytes in the pathogenesis of RA. Before the onset of RA, patients also exhibit circulating autoantibodies and inflammatory mediators [323]. Further, the incidence of RA increases with age [324], while murine tissue–resident alveolar and human BM CD68+ macrophages [325, 326] and mouse STMs all decrease with age. In addition, MHCII− tissue–resident–like macrophages are reduced in patients with active RA compared with those with osteoarthritis [227], and genes relevant to macrophage tissue residency such as

CX3CR1, VSIG4, TIMD4, and TGFBR2 are reduced in leukocyte-rich RA ST. Also, circulating monocytes from patients with RA exhibit a defect in their ability to differentiate into M2-like macrophages, mediated by miR-155, resulting in reduced expression of CD206 and CD163, and increased proinflammatory mediators [327]. Consistent with our interpretation, recently published data demonstrate that when therapy is stopped or reduced in individuals with RA while their disease is in remission, those with increased MerTK+CD206+CD163+ tissue–resident macrophages, expressing transcriptomes enriched in anti-inflammatory signatures, on ST biopsy, are less likely to experience a recurrence of the joint inflammation [228]. Together, these observations identify a critical role for tissue-resident macrophages in the pathogenesis of chronic arthritis and raise the possibility for targeted treatments that prevent disease from developing in individuals with preclinical RA or maintaining remission by modulation of the STM populations.

CHAPTER 3

Uncovering the existence of extravascular synovial resident monocytes

Chapter 3 contains some results from an original research article entitled “Tissue-resident, extravascular Ly6c- monocytes are critical for inflammation in the synovium” (Montgomery et al., in revisions) and is reproduced here with the permission of the copyright holder.

3.1 Introduction

In recent years, our understanding of the mononuclear phagocyte system has expanded, highlighting previously unknown complexities in cell origin and function. However, to date,few studies have examined a role for monocytes in tissues, with the majority of studies centered on circulating monocytes, or monocyte-derived macrophages. Circulating monocytes exist in 3 main states, characterized by CCR2, CX3CR1, CD43 and Ly6c in mice: classical (CM) ($CCR2^{+}CX3CR1^{low}CD43^{-}Ly6c^{hi}$), intermediate (IM) ($CCR2+CX3CR1^{low}$ $CD43^{+}Ly6c^{int}$), and non-classical (NCM) ($CCR2^{-}CX3CR1^{hi}CD43^{+}Ly6c^{low}$) [328, 329]. Previous studies have shown that NCMs are derived from CM in bone marrow and in circulation. CM require CCR2 to exit the bone marrow, while NCM utilize sphingosine-1-phosphate receptor 5 (S1PR5) and/or CX3CR1. Consequently, CCR2-/- mice have reduced numbers of CM in circulation, while S1PR5-/- and CX3CR1-/- mice have reduced NCM [330]. NCMs also require CEBP/β for transcriptional activation of NR4A1 and CSF1R to maintain survival [331]. As such, NR4A1-/- and CEBP/β-/- mice also display markedly reduced numbers of circulating NCM. While transcriptional studies have exposed critical gene signatures for CM and NCM in the bone marrow and circulation, no such studies examined monocyte heterogeneity and their function at the tissue level.

In contrast to well-characterized inflammatory CM [330], the direct impact of NCM in steady state and inflammation is unclear. The current dogma for circulating NCM centers on barrier maintenance due to the ability of NCM to adhere and patrol the endothelium [332]. In this context, NCM maintain the endothelium, scavenge debris, and elicit removal of damaged endothelial cells by neutrophils [332]. To date, only one study has proposed the existence of an NCM population in tissue [333]. The investigators identified a $CD64^{+}$ $CD16.2^{+}$ subpopulation among extravascular

$CD45^{+}Ly6c^{lo}$ cells in the lung that are derived from circulating NCM and require NR4A1. These cells were considered monocytes but were putative precursors for interstitial macrophages. It is not clear whether similar populations are found in other tissues due to the dearth of studies that have identified NCM in tissue that are distinct from circulating NCM.

It is established that recruitment of monocytes to the inflamed synovium is a requisite for sustainment and progression of rheumatoid arthritis (RA) [235]. Support for a functional role for NCM in RA comes from murine models. While complete ablation of circulating monocytes using clodronate-laden liposomes (Clo-lip) prevents the effector phase of K/BxN serum transfer-induced arthritis (STIA) [225], pathology is restored exclusively with transfer of NCM, not CM [225, 334]. CX3CR1-/- mice also display a marked reduction in STIA [335]. In contrast, depletion of CM via anti-CCR2 antibody or deficiency in CCR2 has no effect on arthritis development in STIA [225], TNFα-Tg mice [336], or collagen-induced arthritis [337]. However, NR4A1-/- mice remain sensitive to STIA and CIA regardless of reduced numbers of circulating NCMs [338, 339]. Taken together, these studies present a quandary on the role that NCMs play in RA.

To distinguish differential and distinct functional roles of NCMs, we focused on identifying the heterogeneity of CD64-Ly6c- monocytes in tissue. We uncovered three subpopulations of synovial CD64-Ly6c- cells, which lake known markers of macrophages and can be separated by their expression of MHCII as well as their intra- and extra-vascular location in the mouse synovium. The CD64-Ly6c- cells which were MHCII+ were identified as dendritic cells and were localized to extra vascular compartment of the synovium. The remaining two populations, intra-vascular CD64-Ly6c- cells retain a similar phenotype to circulating NCM, independent of NR4A1 and

CCR2, and long-lived. Furthermore, extra-vascular CD64-Ly6C- are critical for pathology of inflammatory arthritis, via an LFA dependent mechanism. These data document an essential role for newly described tissue-resident CD64-Ly6c- cells in inflammatory arthritis.

3.2 Materials and Methods

Mice (Performed by Anna Montgomery)

Breeder pairs were purchased and experimental mice bred in house, and/or acclimated in barrier and specific pathogen-free animal facility at the Center for Comparative Medicine, Northwestern University. Female mice were used for all RA-like studies. All experimental procedures were carried out on mice aged 8-10 weeks (unless stated otherwise in aging studies). To induce serum transfer arthritis, 85µL/20g/mouse was given intravenously (I.V). All procedures were approved by the Institutional Animal Care and Use committee at Northwestern University.

Flow cytometry analysis (Performed by Anna Montgomery)

To prepare single-cell suspensions from joint, joints were removed from hind paws following euthanasia, perfused, and stored on ice in sterile HBSS. Skin and toes were removed from each paw and bone marrow flushed from exposed tibia with sterile HBSS through a 30G needle. Synovial tissue was then infused with 1.5mL/joint of ankle digestion buffer (2.4mg/mL dispase II, 2mg/mL collagenase D, 0.2mg/mL DNAse I in HBSS pH 7.2-7.6) before incubation at 37oC for 1h with shaking. Cells were then aggravated through a 40-µm mesh filter. Red blood cells were removed with lysis 250µL/sample (1x PharmLyse in sterile water) at room temperature for 1 minute. Dead cells were stained with eFluor 506 viability dye (1:1000 dilution). Cells were incubated with FcBlock (BD Bioscience) and stained with selected antibodie. Cells were fixed

with 10% PFA at 4C for 20 minutes. To prepare single cell suspensions from blood, 90µL blood collected by cardiac puncture was incubated with FcBlock and selected antibodies. Red blood cells were lysed with FACS lyse at rt for 10 minutes (1x in sterile water) and single-cell suspensions were acquired on BD LSR II or BD Symphony. For all FACSorting studies, cells were acquired on a BD FACSAria. Count eBeads were used in joint preparations to calculate cell numbers. Fluorescence minus one sample were used to set gates. Compensation and analysis of flow-cytometry data was carried out in FlowJo V10.

Bulk RNA sequencing (Performed by Gaurav Gadhvi and Shang-Yang Chen)

RNA from FACSorted synovial cells was extracted using PicoPure RNA Isolation kit as per manufacturer's instructions. Bulk RNA-seq shown in Figure 3.1 was carried out using QuantSeq 3' mRNA sequencing kit, while bulk RNA-seq shown in Figure 3.4 utilized full-length SMART-seq v4 Ultra Low Input Kit for Sequencing.

Following the sequencing, libraries in the form of BCL files were obtained from Illumina's BaseSpace platform and demultiplexed (using bcl2fastq v2.17.1.14) to convert them into FastQ read format for further processing. The QuantSeq reads were then processed further by trimming the adapters, low quality bases and short reads (using BBDuk version 37.22 with the following parameters: k=13 ktrim=r useshortkmers=t mink=5 qtrim=r trimq=10 minlength=20). After trimming, remaining reads were aligned to the mouse genome reference mm10 (Mus Musculus / UCSC assembly GRCm38) using STAR [124]. Aligned reads in BAM format were mapped to the reference transcriptome (Mus Musculus GRCm38.87) to obtain exon counts and generate gene expression tables using the tool HTSeq [129]. The SMART- seq reads were trimmed using

Trimmomatic (version 0.36) [118] to remove adapter sequences, low quality bases and short reads (minimum length = 20bp). After trimming, remaining reads were aligned to the mm10 genome reference (Mus Musculus / UCSC assembly GRCm38) with Tophat aligner (tophat 2.1.0) [126]. The aligned reads in BAM format were mapped to gene exons by HTSeq as above using the reference transcriptome GTF file (Mus Musculus GRCm38.87).

All gene expression counts were scaled to read depth using counts per million reads mapped (CPM). To filter out lowly expressed genes, genes with no group mean above 7 CPM in the relevant cell types were excluded from the analysis. Differentially expressed genes (DEG) across multiple cell types were defined as genes with a difference of 2-fold between any two groups. K-means clustering of DEG was carried out in Gene-E. GO enrichment was calculated using GOrilla [149] on each cluster with all DEGs as background. Expressed genes, K-means clustering, and GO processes for all datasets are provided in supplemental files for each figure. Volcano plots were generated in R (version 3.3.1). Principal Component Analysis (PCA) and Pearson's correlation were performed on expressed genes and visualized with R (version 3.3.1). In Figure 3.1, monocyte populations were compared with CD64+ macrophages that were isolated from the same mice and sorted into 4 subpopulations based on the cell surface expression of CX3CR1 and MHCII.

Single-cell RNA sequencing

RNA libraries for single-cell analysis were prepared using 10x Chromium Single Cell 3' Solution v3. Reads were processed and aligned to mm10 mouse reference genome using mkfastq and count commands of cellranger 3.1.0 pipeline [163]. Subsequent analyses, including quality control, unsupervised clustering, identification of cluster markers, and visualization of gene

expression were carried out using Seurat v3.1 package in R. Samples were individually assessed and filtered based on the number of UMI counts and % mitochondrial reads per cell. To account for technical variability, sample-specific thresholds were used as indicated below:

Table 3.1 Quality control metrics and cutoffs for scRNA-seq samples

Genotype	Total Cells Detected	Median Genes/Cell	Median UMIs/Cell	Min UMI/cell	Max UMI/cell	Max % MT reads/Cell	# cells after filtering
C57Bl/6	9447	2681	9323	4000	35000	7	7160
$CCR2^{-/-}$	4867	3184	12927	5500	40000	7	3621
$NR4A1^{-/-}$	13375	818	1494	4000	38000	7	2867
$LFA1^{-/-}$	7908	2622	8880	3500	31000	5	5444
Ly6C-MHCII-CITE-seq	3529	3236	11887	3500	45000	5	2967
Ly6c-MHCII+ CITE-seq	8753	3196	11547	4500	40000	5	6061
CD64+ CITE-seq	9914	3001	9238	5000	40000	10	6051
RA patient CITE-seq	2754	4270	20758	8454	63404	10	1614

For the initial analysis on the C57BL/6 sample, selection of variable genes was performed using the default vst method with nfeatures set to 2000. UMAP dimensionality reduction and unsupervised graph-based clustering were performed with top 16 principal components (PC) and resolution parameter of 0.2. SingleR package v1.0.5 [187] was used to annotate cells with Immgen reference cell types [155]. Differentially expressed genes for each subpopulation were defined log(fold-change)>|0.25| and adjusted p-value<0.05 by Wilcoxon test with Benjamini-Hochberg procedure for False Discovery Rate. Module scores of cell type signatures based on manually

selected genes were computed using AddModuleScore function with default parameters (Table 3.2). Pearson's correlations were calculated between the averaged expression profiles of single-cell subpopulations and bulk RNA-seq on monocyte populations (Figure 3.1). Cell cycle scoring was performed using G2/M and S phase gene sets provided in Seurat, converted to orthologous mouse genes using BioMart R package. Cells in subpopulations 0, 1, and 3 were further classified as either MHCII+ or MHCII– by a threshold of 2 on normalized expression of H2-Eb1. Co-expression of H2-Eb1 with MHCII+/- compartment genes was visualized using DimPlot function with blend=TRUE parameter.

Table 3.2 Genes used in calculation of cell-type-specific module scores

PB CM	PB NCM	Syn Ly6c-	cDC
Fcgr1	Cebpb	C1qa	Adam19
Ifnar2	Nr4a1	C5ar1	Ap1s3
Irf5	Cx3cr1	Ccl2	Ass1
Irf7	Il17ra	Cd74	Bcl11a
Myd88	Il10ra	Gas6	Btla
Stat1	Csf1r	H2-Aa	Ccr7
	Irf1	Lyve1	Flt3
		Mafb	Gpr132
		Retnla	Gpr68
		Cd14	Gpr82
		Cd209d	H2-Eb2
		Trem2	Hmgn3
		Rab31	Kit
			Klri1
			Kmo
			P2ry10
			Rab30
			Septin6
			Slamf7
			Traf1
			Zbtb46

Similarity of our annotation with published clusters in Culemann et al [220] was determined by calculating the fraction of their top 20 reported markers that overlapped with the list of differentially expressed genes for each of our subpopulations: Cycling (2), Monocyte (4) and cDC (5) as well as MHCII+ and MHCII-. Significance of overlap was determined by hypergeometric distribution with the expressed genes as background, defined as those that were present in more than 25% of cells for at least one subpopulation (n = 6344). A FWER cutoff of 0.05 was applied based on Bonferroni correction for 35 comparisons (Supp Table 2E). To assign C57BL/6 myeloid cells to either i.v. or e.v. Syn Ly6c-, or mono-DC singleR was run with the bulk RNA-seq on these populations as reference.

For the processing of CITE-seq samples, dimensionality reduction was performed with top 10 PCs. Antibody intensities were normalized using CLR method in NormalizeData function. Identities of individual cells were annotated via the FindTransferAnchors and TransferData functions using the top 30 PCs and C57BL/6 sample as reference data. The analysis of scATAC-seq experiment was carried out using Signac 1.3.0 package. Quality control was performed with the following cutoffs: fragments in peak region > 4000 and < 40000, percentage of reads in peaks > 40%, ratio reads in ENCODE blacklist regions < 0.02, nucleosome binding signal < 1.2, and transcriptional start site (TSS) enrichment score > 2. Latent semantic indexing (LSI) was performed through RunTFIDF and RunSVD functions, followed by UMAP dimensionality reduction using 2:30 LSI components. Annotation of cell identities was performed by integrating and transferring labels from C57BL/6 CITE-seq data. Per-cell activity scores of JASPAR2020 transcription factor (TF) motifs was inferred through chromVAR wrapper function implemented within Signac package.

For annotation of human myeloid cells, we first merged the Ly6c-MHCII- and the CD64+ CITE-seq data carrying over the annotations from the former. Then, we defined the markers genes for MHCII- cells (e.v. Syn Ly6c-) using FindAllMarkers and chose the top 10 by fold-change with human orthologs. Next, we calculated the module score in cells from either the AMP CD14+ scRNA-seq dataset [227] or our own RA Patient CITE-seq data using AddModuleScore.. We also performed de novo clustering (10 PCs and 0.6 resolution) on the latter data to determine which cluster exhibited the highest expression of the MHCII- module and link this with RNA expression and ADT intensity of other genes/surface markers.

Integration of C57BL/6 cells with other samples was executed using Seurat anchoring method with 30 CCA dimensions and visualized by UMAP using top 13 (B6 with CCR2-/- and NR4A1-/-) and 15 (B6 with LFA1-/-) PCs. Subpopulation labels were determined by the majority label of C57BL/6 annotated cells in each cluster. Relative contributions of the integrated samples were calculated by down-sampling to 2000 (B6 with CCR2-/- and NR4A1-/-) and 3000 (B6 with LFA1-/-) cells for each sample after clustering, with significance determined through chi-square test. GO processes were obatined using GOrilla on genes increased or decreased in expression log2(fold-change)>|0.25|, adjusted p-value<0.05 by Wilcoxon test with Benjamini-Hochberg) in LFA1-/-MHCII- compartment compared to C57Bl/6 with the 14144 expressed genes as background.

Intra and extra vascular labeling of immune cells (Performed by Anna Montgomery)

To label intra-vascular immune cells, anti-CD45 BUV661 antibody was administered I.V. at 6µg/mouse in 200µl sterile PBS. Mice were then returned to housing environment for 5 minutes

before euthanasia or for 1 hour before administration of second I.V. anti-CD43 BUV395 antibody. In studies using STIA, K/BxN serum was administered 5 minutes after anti-CD45 BUV661 antibody, 60 minutes prior to anti-CD43 BUV395 antibody as previously described.

Monocyte depletion (Performed by Anna Montgomery)

For depletion studies, 200µl/mouse clodronate-laden liposomes were given I.V. 24 hours prior to euthanasia. All mice were perfused with 20mL with ice-cold HBSS following euthanasia to remove circulating cells and retain adherent intra-vascular cells.

Bone marrow chimeras (Performed by Anna Montgomery)

See original manuscript for details

Statistical analysis

All statistical analysis was carried out in GraphPad Prism V8. P-values less than 0.05 were considered statistically significant using two-tailed unpaired t-test with equal variance.

3.3 Results

Non-classical monocytes in the synovium are distinct from those in the circulation

We sought to determine the contribution of circulating NCM to inflammatory arthritis by inducing STIA in NR4A1-/- mice, which are depleted of NCM in peripheral blood (PB) (Figure 3.1A, S3.1A-B). NR4A1-/- mice developed STIA of comparable severity and onset to C57Bl/6 controls (Figure 3.2B), in agreement with a previous report [338]. Flow cytometry was then performed to identify monocyte populations that are preserved in the synovium of NR4A1-/- mice, which may explain the sensitivity of these mice to inflammatory arthritis. We identified a novel synovial myeloid niche defined as CD45+CD11b+Ly6G-SigF-CD64- of which the majority were Ly6C-. Based on this gating strategy synovial macrophages (CD64+) were the most abundant myeloid population in the synovium (Figure S3.1C), while Ly6C- represented 10% of the synovial mononuclear phagocyte compartment, and Ly6C+ and Ly6C^{int} cells composed less than 1%. Synovial Ly6C- (Syn Ly6C-) cell numbers remained unchanged in NR4A1-/- mice (Figure 3.1C) even while NCM were markedly reduced in PB. In order to confirm that Syn Ly6C- cells were not dependent on CCR2, STIA was induced in CCR2-/- mice lacking CM in PB (Figure 3.1D, S3.1B). CCR2-/- mice also showed comparable clinical scores in STIA to C57Bl/6 controls (Figure 3.1E) as reported in previous studies [225, 335] and their numbers of synovial Ly6c- cells were unchanged compared to controls (Figure 3.1F). Taken together, these data confirm that neither subtype of circulating monocyte is required for inflammatory arthritis, while a newly identified population of Syn Ly6C- cells is independent of NR4A1 and CCR2, lacks expression of macrophage associated markers and may play an essential role in STIA.

Cell numbers were measured throughout disease to determine the response of synovial tissue

myeloid populations during STIA. Synovial Ly6C+ and Ly6C- numbers significantly expanded on D3 (p=0.003, p=0.006) and D7 (p=0.003, p=0.03) post-STIA compared to D0, while Ly6Cint cells were not significantly different at any timepoint (Figure 3.1G). By D14, synovial Ly6C+ cells returned to baseline whereas Ly6C- cells plateaued on D14 and D21 compared to D0 (p= 0.05, p=0.02). Similarly, PB CM reach a peak prior to 21 days post-STIA while PB NCM continue to increase (Figure 3.1H).

To determine whether synovial NCM exhibit a distinct transcriptional state from circulating monocytes, we isolated CM and NCM from PB (PB CM, PB NCM) and Syn Ly6c- from hind joints of mice for bulk RNA sequencing (RNA-seq). Given PB IM are likely an intermediate cell state, these cells were excluded from our studies. PB CM, PB NCM and Syn. NCM exhibited distinct transcriptional profiles from each other (Figure 3.1I, S3.1D). We then compared expression of genes preferentially associated with PB CM vs. PB NCM as described in Mildner et al. [331]. Expression of monocyte genes in PB CM and PB NCM largely aligned with expectations, but Syn Ly6C- cells did not uniformly express PB CM genes – such as Irf7, Ccr2, Ifi30, Mmp8, and Cebpd – or those associated with PB NCM – such as Apoe, Csf1r, Fcgr4, Pparg, Nr4a1, and Cx3cr1 (Figure 3.1J). Furthermore, loss of NR4A1 had a minimal effect on the transcriptional profile between synovial Ly6C- cells compared to WT (Figure 3.1K, S3.1E). These data show that Syn Ly6C- cells are transcriptionally distinct from circulating cells and are not NR4A1 dependent.

K-means clustering of 5116 differentially expressed genes identified 3 gene clusters preferentially expressed by PB CM, PB NCM, or Syn Ly6C- cells (Figure 3.1L). Compared to other clusters, the Syn. NCM cluster (cluster 3) was enriched for genes associated with

extracellular matrix organization, hormone secretion, cell division, cell adhesion, and regulation of biological processes (Figure S3.1F). Additionally, increased expression of genes involved in antigen presentation (H2-Aa and Cd74), immune activation (Cd9, Pf4, and Cd36), complement (C1qa, C1qb, and C1qc) and pro-fibrotic/repair (Gas6 and Cd163) were detected in NCMs compared to PB CM and PB NCM (Figure 3.1L). We found that on the individual gene and global level, Syn Ly6C- cells exhibited a distinct transcriptional profile from each of the four CD64+ synovial macrophage populations (Figure 3.1M, Figure S3.1G-H). Taken together, these data uncover a novel CD64-Syn Ly6c- population present in the synovium.

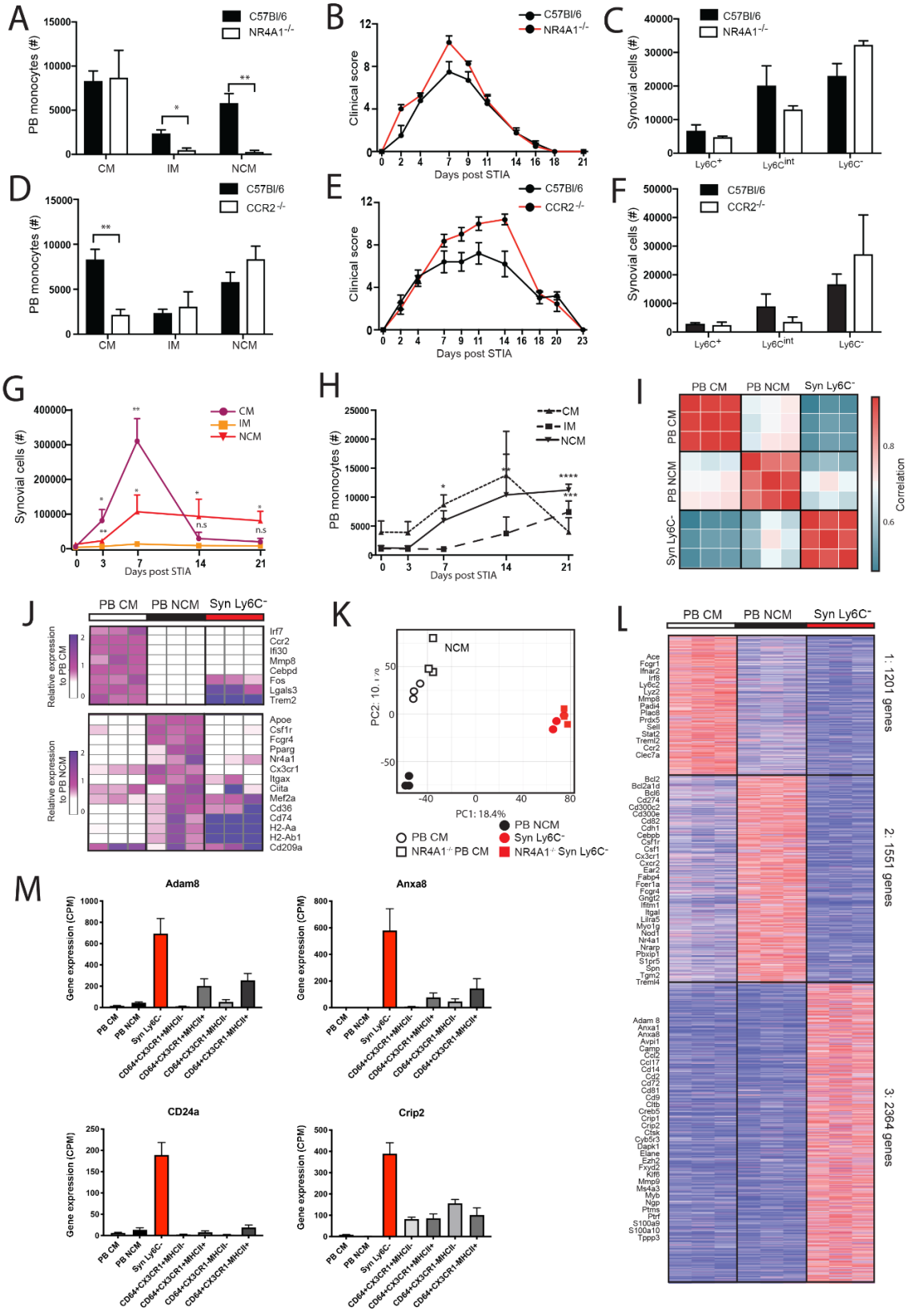

A
PB monocytes (#)
C57Bl/6
NR4A1-/-
CM
IM
NCM
B
Clinical score
Days post STIA
C
Synovial cells (#)
Ly6C+
Ly6Cint
Ly6C-
D
CCR2-/-
E
F
G
Synovial cells (#)
H
PB monocytes (#)
I
PB CM
PB NCM
Syn Ly6C-
Correlation
J
Relative expression to PB CM
Relative expression to PB NCM
Irf7
Ccr2
Ifi30
Mmp8
Cebpd
Fos
Lgals3
Trem2
Apoe
Csf1r
Fcgr4
Pparg
Nr4a1
Cx3cr1
Itgax
Ciita
Mef2a
Cd36
Cd74
H2-Aa
H2-Ab1
Cd209a
K
PC2: 10.1%
PC1: 18.4%
PB CM
NR4A1-/- PB CM
PB NCM
Syn Ly6C-
NR4A1-/- Syn Ly6C-
L
1: 1201 genes
2: 1551 genes
3: 2364 genes
M
Adam8
Anxa8
CD24a
Crip2
Gene expression (CPM)
CD64+CX3CR1+MHCII-
CD64+CX3CR1+MHCII+
CD64+CX3CR1-MHCII-
CD64+CX3CR1-MHCII+

Figure 3.1 Synovial NCM are phenotypically distinct from circulating NCM. (A) Numbers of classical (CM), intermediate (IM), and non-classical (NCM) monocytes in the peripheral blood (PB), (B) STIA severity and (C) numbers of CM, IM, and NCM in synovium of C57Bl/6 compared to NR4A1-/-mice, and in C57Bl/6 compared to CCR2-/-mice (D-F). (G) Changes in numbers of synovial CM, IM, and NCM, and (H) PB CM, IM, and NCM during STIA. Data shown are n>=4 ±S.E.M, *= P<0.05, **= P<0.01, ****=P<0.001. (I) Pairwise Pearson's correlation of global gene expression between replicates of Syn. NCM, PB CM and PB NCM. (J) Fold-change expression of monocyte associated genes from Mildner et al. [331] (K) PCA of 10206 genes expressed by PB CM, PB NCM, Syn. NCM from C57Bl/6 and NR4A1-/-mice. (L) k-means clustering of 5115 differential genes (LogFC >1) across PB CM, PB NCM, and Syn. NCM, (M) Mean expression of representative genes from PB CM, PB NCM, Syn Ly6C- and Syn Mac populations (RNA-seq data: n=3, error bars indicate SEM).

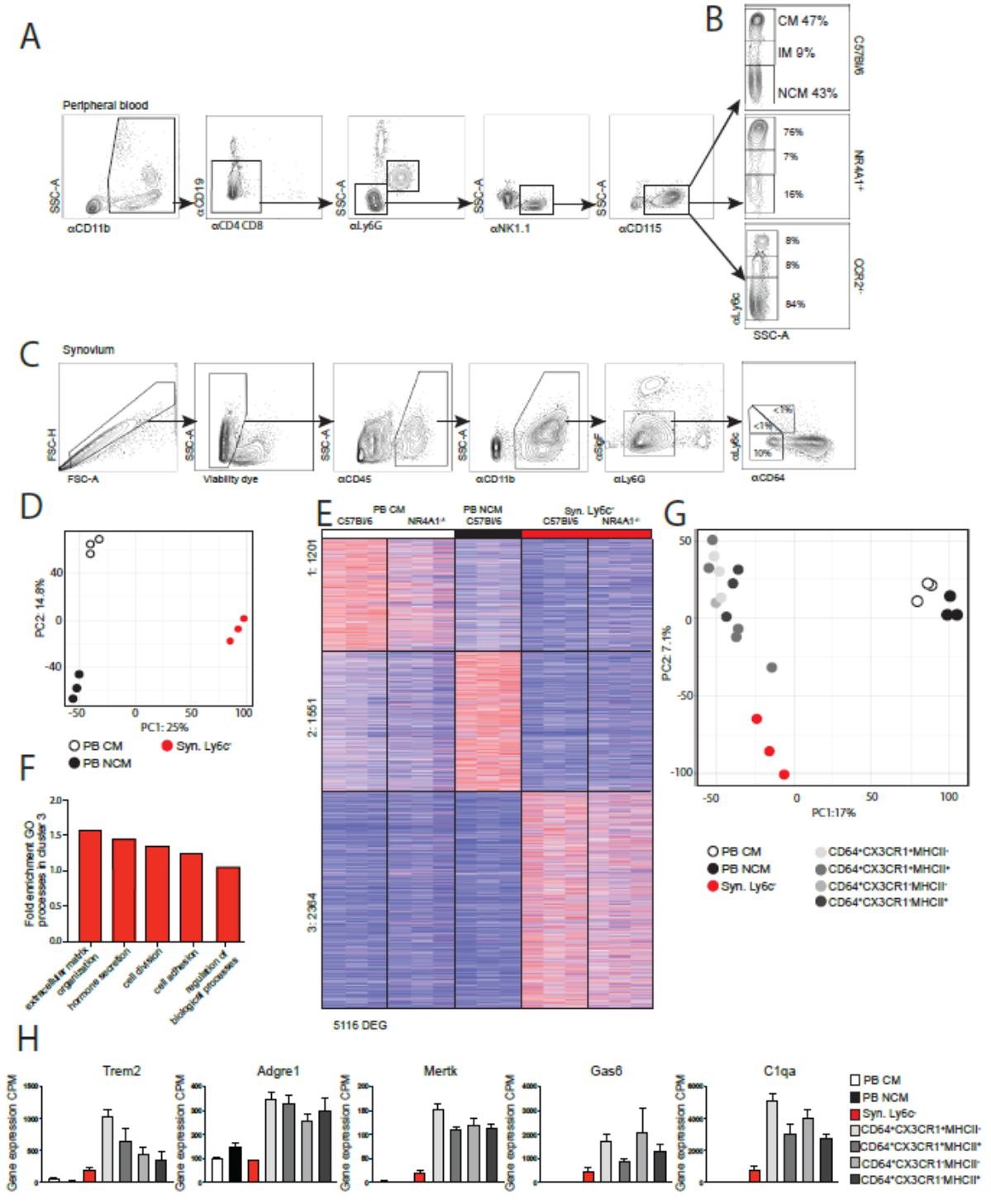


Figure S3.1 (A) Flow gating strategy for identification of PB monocyte sub-populations in C57Bl/6, (B) NR4A1-/- and CCR2-/- mice, and (C) Flow gating strategy for identification of Syn.

monocytes (CD64-) and macrophages (CD64+). (D) PCA of 10206 genes expressed by PB CM, PB NCM, and Syn. NCM. (E) Visualization of clusters from Figure 1L with expression in PB CM and Syn. NCM from NR4A1-/- mice. F) Significantly enriched GO processes in cluster 3 from Figure 1L with preferential expression in Syn. NCM ($P<0.05$). (G) PCA of 9661 genes expressed in PB CM, PB NCM, Syn. NCM and Syn. macrophages. (H) Volcano plot of differentially expressed genes (LogFC > 1) between Syn. NCM and Syn. macrophages.

Single-cell RNA-sequencing identifies novel synovial NCM population

We utilized single-cell RNA-seq (scRNA-seq) to investigate heterogeneity of Syn Ly6C- (CD45+CD11b+Ly6G-SigF-Tim4-CD64-). We defined 6 clusters (0-5) of mononuclear subpopulations using unsupervised graph-based clustering of 7160 cells sequenced from C56BL/6 mice (Figures 3.2A, S3.2A-B). The expression of Cd14 and Itgam (CD11b) confirmed that the sorted cells were of myeloid lineage, while low expression of Fcgr1 (CD64), Timd4 (Tim4) and Mertk confirms non-macrophage classification (Figures 3.2B, S3.2C). Using SingleR and reference ImmGen dataset we found that subpopulations 0-3 exhibited a mixture of monocyte, macrophage, and DC annotations expected of tissue-resident myeloid cells, while subpopulations 4 and 5 were primarily assigned to monocytes and DC respectively (Figure 3.2C). Subpopulation 4 also displayed similarity to PB CM and PB NCM, suggesting these may be monocytes retained in tissue vasculature or cells that were not removed by perfusion (Figures 3.2D, S3.2B, S3.2D-F). Subpopulation 5 had the highest expression of genes associated with conventional DCs [340] (Figures 3.2E, S3.2G). Subpopulation 2 expressed high levels of cell cycle genes (Figure S3.2H-J), suggesting theses cells were actively dividing. Therefore, the formerly labelled Syn Ly6c- compartment contains a mixture of cells with different identities from monocytes to DC.

We then performed scRNA-seq on CD45+CD11b+Ly6G-SiglecF-Tim4-CD64- synovial cells from NR4A1-/- and CCR2-/- mice at steady state to determine whether their compositions were 179 affected by CCR2 or NR4A1 deficiency (Figure 2F, Supplemental Table 2C). These datasets were integrated with control cells (C57Bl/6) and the subpopulation annotation was superimposed (Figure S2K). The distribution of CCR2-/- and NR4A1-/- cells across the 6 subpopulations was significantly different from controls ($p<2.2e-16$, $p=4.89e-10$) (Figure 2G), suggesting that at least some of these cells are derived from circulating monocytes. In particular, the depletion of subpopulation 4 in CCR2-/- supports the contribution of classical monocytes.

Although subpopulations 0, 1 and 3 all had high similarity to Syn Ly6C- cells from Figure 3.1 (Figures 3.2D, S3.2E), the variability in gene expression profiles suggested they may be distinct subtypes. In particular, subpopulation 1 exhibited higher expression of certain genes associated with the DC lineage (Figures 3.2E, S3.2B). Since MHCII genes (H2-Eb1, Ab1, Aa, DMb1, DMa) are associated with DCs, we partitioned subpopulations 0, 1 and 3 based on their expression, specifically on H2-Eb1 (Figure S3.2L-M). As expected, the MHCII+ compartment exhibited elevated expression of genes associated with monocyte derived DCs (mono-DCs), including Cd209a, Cd74, and Nr4a3 [341-343] (Figure 3.2I). The remaining non-DC MHCII- fraction was enriched for genes that are known to regulatory function in inflammation, lipid metabolism and angiogenesis including Crip2, Fxyd2, and Rnase4 [344-346] (Figure 3.2G). Furthermore, there was no significant difference in the proportion of cells annotated as MHCII- vs MHCII+ in CCR2-/- mice and MHCII- cells were increased in NR4A1-/- mice (Figure S3.2N). Collectively, these data demonstrate that MHCII- cells represent a distinct Ly6c- population residing in the synovium of mice, which is independent of CCR2 and NR4A1.

We next compared our single-cell data to those recently published using on CD11b+ myeloid cells from murine synovium. The differentially expressed genes associated with our subpopulations were compared to the top 20 marker genes for the single-cell myeloid subpopulations (excluding neutrophils) identified by Culemann et al. at day 5 post-STIA [220] (Figure S2O). The MHCII- cells displayed a high unique overlap in genes with the CX3CR1+, RELMa+, MHCII+, and CCR2+ARG1+ populations, while MHCII+ cells overlapped genes from MHCII+DC as well as CX3CR1+ and MHCII+ populations from the Culemann dataset (Figure S2O). The cycling cells and monocyte populations overlapped significantly with STMN1+ proliferating cells and CCR2+IL1B+ mononuclear cells respectively (Figure S2O). These results support the identification of these individual populations, but do not exactly match the Culemann clusters, which were sorted on CD45+CD11B+Ly6G- and so contain large numbers of macrophages which greatly outnumber the tissue-resident synovial Ly6c- cells [220].

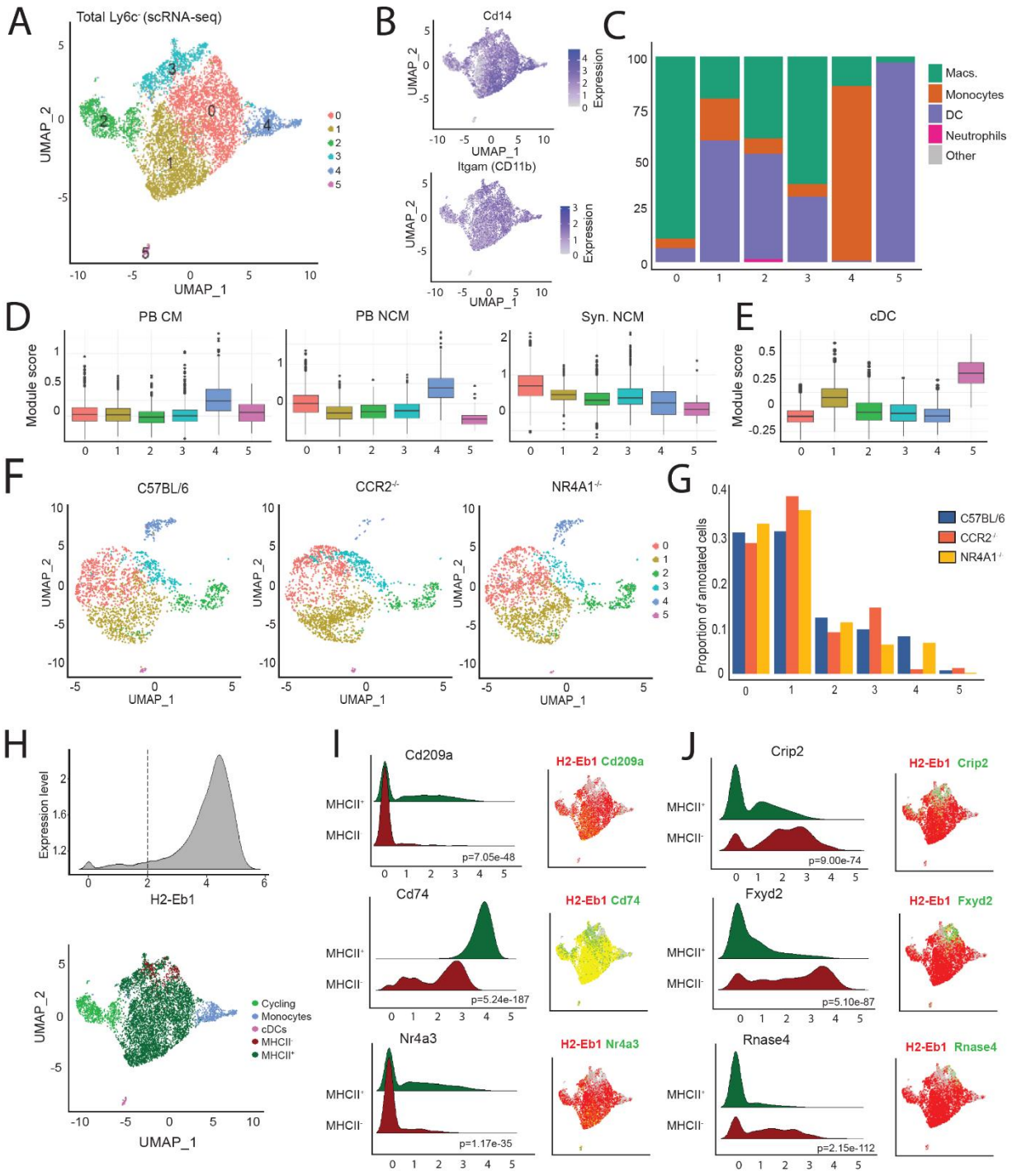

A
Total Ly6c- (scRNA-seq)
UMAP_2
UMAP_1
0
1
2
3
4
5
B
Cd14
Itgam (CD11b)
Expression
C
Macs.
Monocytes
DC
Neutrophils
Other
D
PB CM
PB NCM
Syn. NCM
Module score
E
cDC
F
C57BL/6
CCR2-/-
NR4A1-/-
G
Proportion of annotated cells
H
Expression level
H2-Eb1
Cycling
Monocytes
cDCs
MHCII-
MHCII+
I
Cd209a
H2-Eb1 Cd209a
p=7.05e-48
Cd74
H2-Eb1 Cd74
p=5.24e-187
Nr4a3
H2-Eb1 Nr4a3
p=1.17e-35
J
Crip2
H2-Eb1 Crip2
p=9.00e-74
Fxyd2
H2-Eb1 Fxyd2
p=5.10e-87
Rnase4
H2-Eb1 Rnase4
p=2.15e-112

Figure 3.2: Single-cell RNA-sequencing analysis of joint myeloid niche identifies tissue Syn Ly6c$^-$ cells. (A) UMAP depicting 6 subpopulations of total Ly6c$^-$ (CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$ Tim4$^-$CD64$^-$Ly6c$^-$) cells from scRNA-seq data. (B) Expression of myeloid markers Cd14 and Itgam. (C) Percent of cells in each subpopulation assigned to ImmGen cell types by singleR. (D) Module score for each scRNA-seq subpopulation representing expression of key genes in PB CM, PB NCM, Syn Ly6c$^-$ and E) cDC. (F) Integration of scRNA-seq data on total Ly6c$^-$ (CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$Tim4$^-$CD64$^-$Ly6c$^-$) cells from CCR2$^{-/-}$ and NR4A1$^{-/-}$ mice with C57Bl/6 (sub-sampled to 2000 cells). (G) Proportion of cells annotated as each subpopulation in C57Bl/6, CCR2$^{-/-}$ and NR4A1$^{-/-}$ mice. (H) Re-classification of cluster 2 as Cycling cells, cluster 4 as Monocytes, cluster 5 as cDCs and clusters 0, 1, and 3 as MHCII$^+$ or MHCII$^-$ based on expression of H2-eb1. (I-J) Ridge plots and UMAP visualization of gene expression by MHCII. compartment. P-value by Wilcoxon test is indicated.

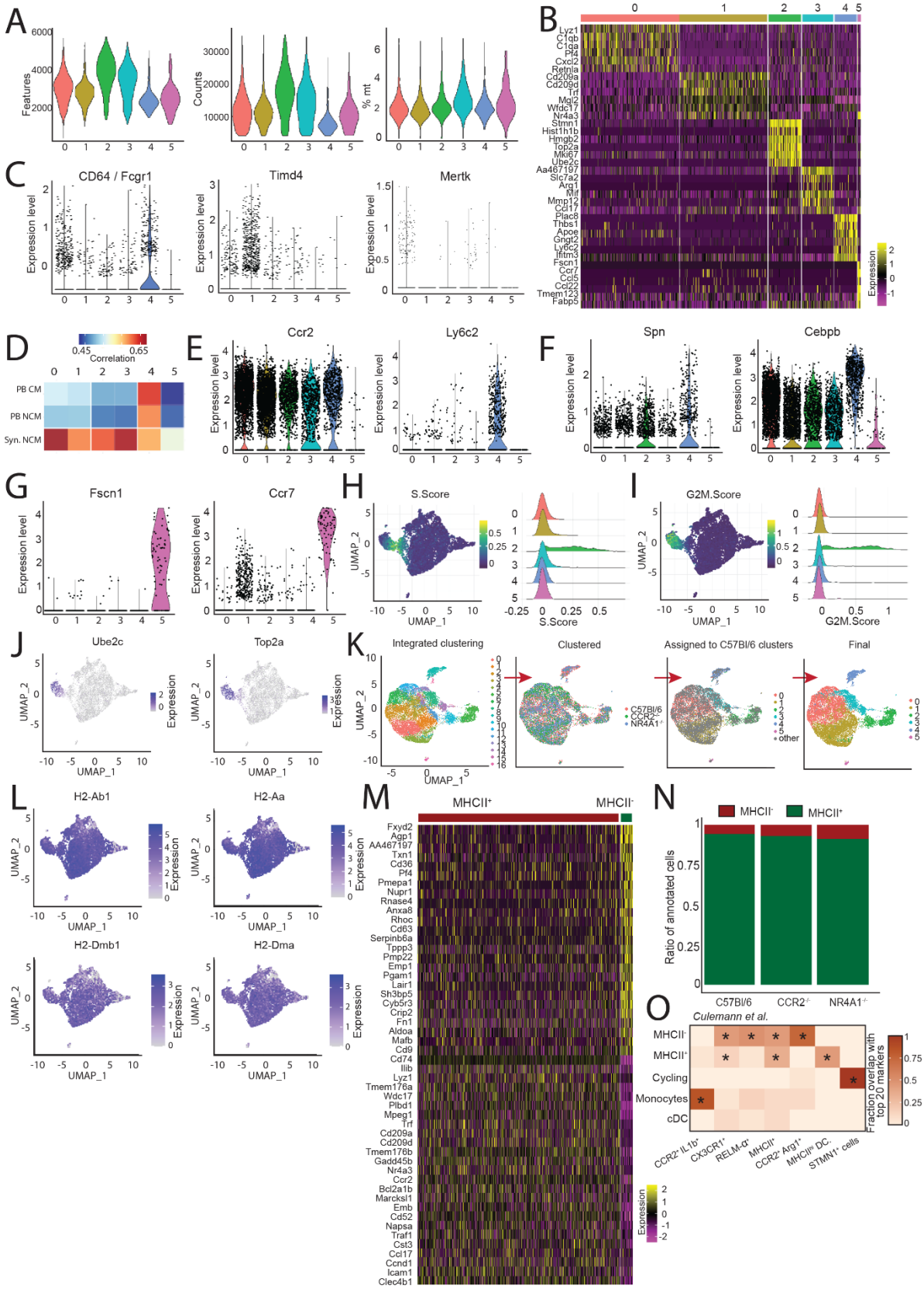

A
Features
Counts
% mt
B
Lyz1
C1qb
C1qa
Pf4
Cxcl2
Retnla
Cd209a
Cd209d
Trf
Mgl2
Wfdc17
Nr4a3
Stmn1
Hist1h1b
Hmgb2
Top2a
Mki67
Ube2c
Aa467197
Slc7a2
Arg1
Mif
Mmp12
Ccl17
Plac8
Thbs1
Apoe
Gngt2
Ly6c2
Ifitm3
Fscn1
Ccr7
Ccl5
Ccl22
Tmem123
Fabp5
Expression
C
CD64 / Fcgr1
Timd4
Mertk
Expression level
D
Correlation
PB CM
PB NCM
Syn. NCM
E
Ccr2
Ly6c2
F
Spn
Cebpb
G
Fscn1
Ccr7
H
S.Score
I
G2M.Score
J
Ube2c
Top2a
K
Integrated clustering
Clustered
Assigned to C57Bl/6 clusters
Final
C57Bl/6
CCR2-
NR4A1-
other
L
H2-Ab1
H2-Aa
H2-Dmb1
H2-Dma
M
MHCII+
MHCII-
N
Ratio of annotated cells
C57Bl/6
CCR2-/-
NR4A1-/-
O
Culemann et al.
MHCII-
MHCII+
Cycling
Monocytes
cDC
Fraction overlap with top 20 markers
CCR2+ IL1b+
CX3CR1+
RELM-α+
MHCII+
CCR2+ Arg1+
MHCIIlo DC
STMN1+ cells

Figure S3.2: (A) Quality control of CD45+CD11b+Tim4-CD64- cells by subpopulation from C57Bl/6 mice showing number of genes, number of UMIs, and percent mitochondrial reads per cell. (B) Relative expression of top 5 marker genes (by fold change) across subpopulations 0-5. (C) Normalized expression of macrophage-associated genes across clusters. (D) Pearson's correlation between gene expression in scRNA-seq subpopulations and bulk RNA-seq of PB CM, PB NCM, and Syn Ly6c-. (E) Normalized expression of CM, (F) NCM and (G) DC associated genes. (H) Normalized expression of S-phase and (I) G2-phase module genes. (J) Visualization of expression of cell cycle genes. (K) Integration of scRNA-seq data from CCR2-/- and NR4A1-/- mice with superimposed C57Bl/6 annotations. (L) Expression of MHCII genes H2-Ab1, H2-Aa, H2-DMb1, H2-DMa. (M) Relative expression of top 20 differentially expressed genes between the $MHCII^{+}$ and $MHCII^{-}$ cells. (N) Ratio of cells annotated as either $MHCII^{+}$ or $MHCII^{-}$ in C57Bl/6, $CCR2^{-/-}$ and $NR4A1^{-/-}$ mice. (O) Fraction overlap of differentially expressed genes from C57Bl/6 scRNA-seq subpopulations with top 20 markers of myeloid populations from murine synovium at day 5 post-STIA identified in [220]. * indicates significant p-value by hypergeometric test after FWER correction.

Synovial $Ly6c^{-}$ cells exist as three distinct populations

To validate our partitioning of cells based on MHCII expression, we separately sorted synovial $CD45^{+}CD11b^{+}Ly6G^{-}SiglecF^{-}CD64^{-}$ cells into $Ly6c^{-}MHCII^{+}$ and $Ly6c^{-}MHCII^{-}$ compartments from the same mice for Cellular Indexing of Transcriptomes and Epitopes by sequencing (CITE-seq) providing both transcriptional (RNA) and surface marker (Antibody-Derived Tags – ADT) data (Figure S3.33A). By comparison of the transcriptional profiles, we

annotated these cells as belonging to the 5 populations from Figure 3.2. As expected, the Ly6c$^-$ MHCII$^-$ cells contained an expanded proportion of MHCII$^-$ and monocyte populations with a very small MHCII$^+$ population and no cDC (Figure 3.3A-B, S3.3B). In contrast, the Ly6c$^-$MHCII$^+$ cells were almost entirely annotated as the MHCII$^+$ population with a very minor MHCII$^-$ population (Figure S3.3B-C). In addition, a substantial portion of the Ly6c$^-$MHCII$^-$ cells were annotated as cycling cells and these appeared to include neutrophils (Figure S3.3D). Further, the surface marker intensity of CD163, FrB, C5aR and Vista discriminated the MHCII$^-$ populations from the MHCII$^+$ and monocytes (Figure 3.3C). Next, we examined the epigenomic landscape of synovial CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$CD64$^-$Ly6c$^-$MHCII$^-$ cells using single-cell assay for transposase-accessible chromatin by sequencing (scATAC-seq) assay. We transferred the transcriptional annotations based on similarity with chromatin accessibility levels at the corresponding genes and obtained similar proportions of the populations (Figure 3.3D). Using this data to assess transcription factor (TF) activity, we found that MAFB and MYC were specific to the MHCII$^-$ population while FLI1 and IRF8 were affiliated with the cycling cell and the monocyte populations, respectively (Figures 3.3E-F). Thus, sorting on cell surface levels of MHCII effectively enriched for non-DC Syn Ly6c$^-$ cells. However, the Ly6c$^-$MHCII$^-$ compartment still contains both standard monocytes and the novel MHCII$^-$ population.

To further distinguish these 2 populations of Syn Ly6c$^-$cells, we investigated whether they differed in anatomical location. We used an established *in vivo* intra-vascular labeling system followed by flow cytometry [347]. Only intra-vascular cells were labelled by administering intravenously (I.V.) a fluorescently conjugated anti-CD45 antibody (αCD45-BUV661-(I.V.) prior to euthanasia (Figures 3.3G-H). Over 90% of the circulating leukocytes were labeled 5 min post

I.V. (Figure S3.3E). Thus, immune cells which co-label with αCD45-BUV661-(I.V.) and the *ex vivo* (E.V.) Anti-CD45 antibody (αCD45-AF700-(E.V.) are considered intra-vascular, while single positive αCD45-AF700-(E.V.) cells are extra-vascular. Synovial macrophages were distinguished using $CD64^{+}$ and $Tim4^{+}$ (Figure 3.3G). Intra-and extra-vascular cells were then gated based on $CD45^{+}CD11b^{+}Ly6G^{-}SiglecF^{-}Tim4^{-}CD64^{-}Ly6c^{-}$ expression to isolate the synovial $Ly6c^{-}$ cells. The intravascular population (i.v. Syn $Ly6c^{-}$) displayed higher expression of CX3CR1 and CD43 (Figure 3.3H), reminiscent of circulating NCMs. Meanwhile extra-vascular (e.v.) cells were divided based on the expression of MHCII to obtain e.v. Syn $Ly6c^{-}$ and mono-DC. We confirmed the identity of the $MHCII^{+}$ compartment as mono-DC due to increased expression of DC transcription factor ZBT46 using zDC-cre mice crossed with zsGFP reporter mice (zDC-GFP), consistent with the scRNA-seq data (Figure S3.3F-H). Numerically, the mono-DC had the highest number, while the e.v. Syn Ly6c- cells were ~3X less and the i.v. Syn. Ly6c- cells were less than 500 per synovium (Figure 3.3I). These numbers are consistent with the scRNA-seq data in Figures 1 and 2. Using flow cytometry to compare cell surface levels, we found that FcγRIV and Treml4 discriminated i.v. Syn $Ly6c^{-}$ cells, while Folrb, VISTA and Lyve1were higher in some but not all e.v. Syn $Ly6c^{-}$ cells (Figure S3.3I).

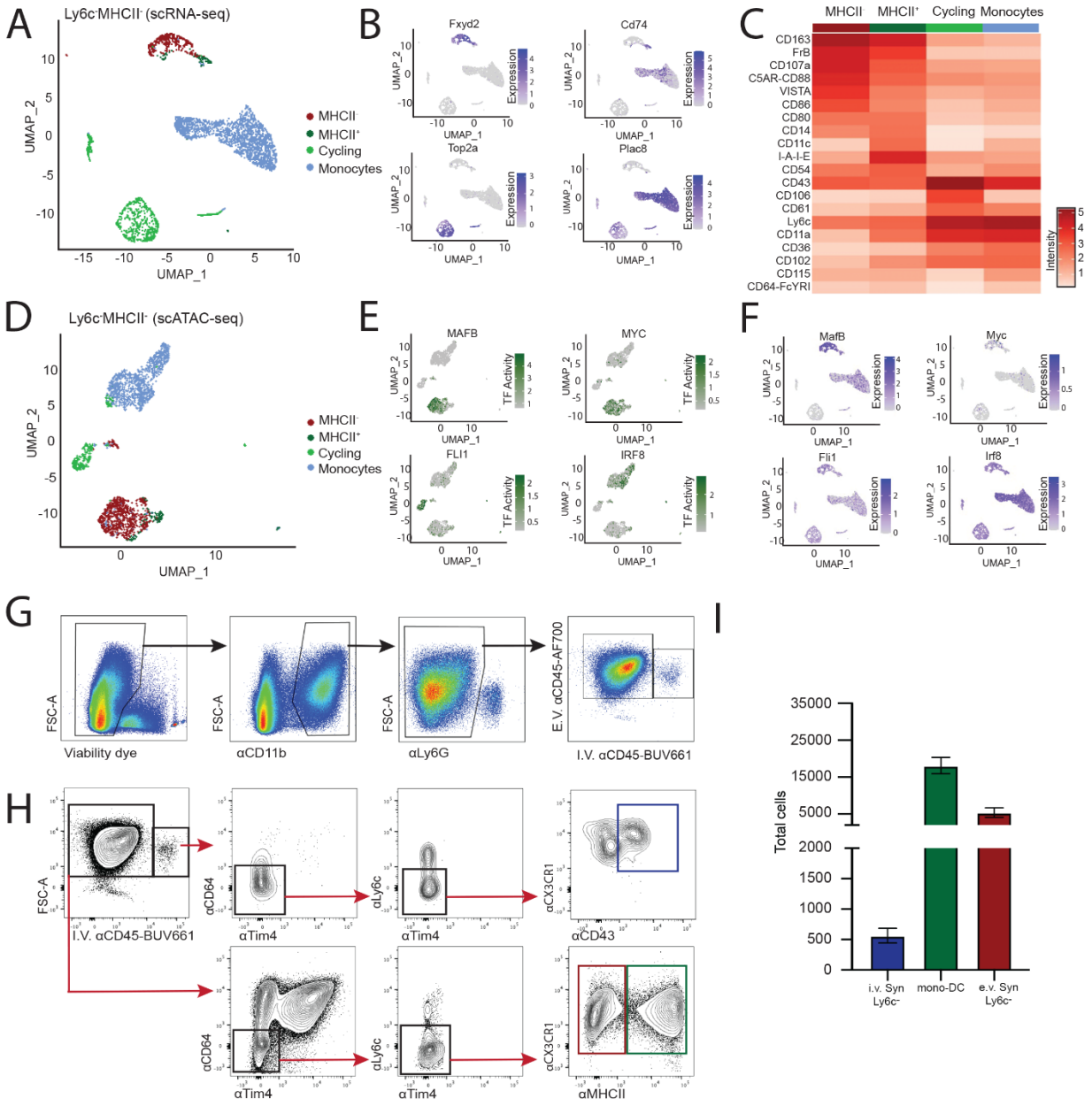


Figure 3.3 Identification of intra- and extra- vascular Syn Ly6c$^-$ cells by flow cytometry. (A) Annotation of single-cell RNA-seq data on CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$Tim4$^-$CD64$^-$Ly6c$^-$ cells from C57Bl/6 mice based on subpopulations defined in total Syn Ly6c$^-$ cells (CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$Tim4$^-$CD64$^-$Ly6c$^-$) (Figure 3.2H). (B) Expression of genes associated with MHCII$^-$

(Fxyd2), MHCII^{+} (Cd74), cycling (Top2a), and monocyte (Plac8) subpopulations. (C) Intensity by ADT count of surface markers using CITE-seq. (D) Annotation of scATAC-seq data on CD45^{+}CD11b^{+}Ly6G^{-}SiglecF^{-}Tim4^{-}CD64^{-}Ly6c^{-}MHCII^{-} cells from C57Bl/6 mice, based on subpopulations defined in total Syn Ly6c^{-} cells (Figure 3.2H). (E) Transcription Factor (TF) activity and (F) expression of corresponding genes associated with MHCII^{-} (MafB & MYC) Cycling (Fli1), and monocyte (Irf8) subpopulations. (G-H) Gating strategy to distinguish intra-vascular and extra-vascular Syn Ly6c- cells and mono-DC. (I) Numbers of Syn Ly6c- and mono-DC in hindjoints of C57Bl/6 mice in steady state.

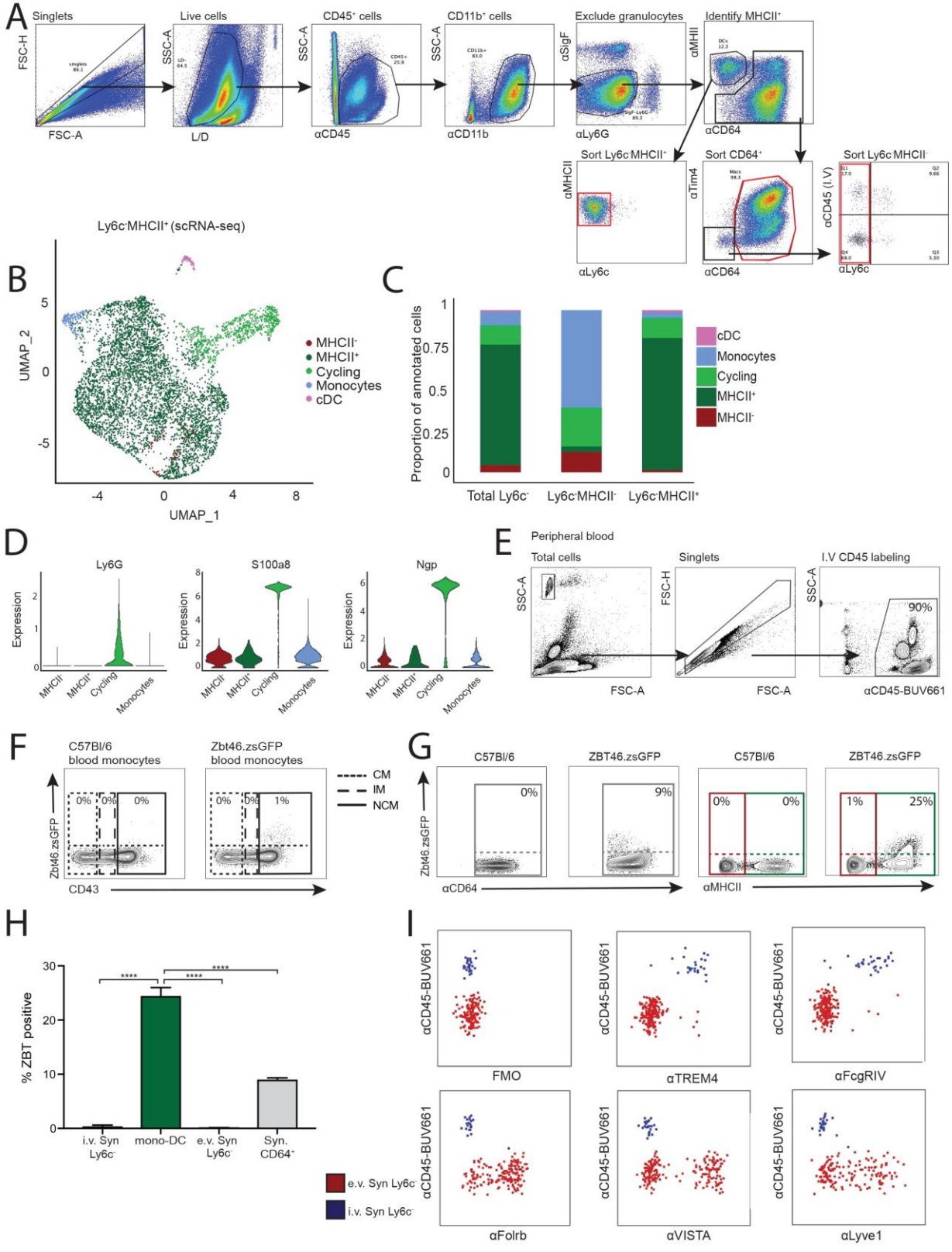

A
Singlets
Live cells
CD45+ cells
CD11b+ cells
Exclude granulocytes
Identify MHCII+
FSC-H
FSC-A
SSC-A
L/D
αCD45
αCD11b
αSigF
αLy6G
αMHII
αCD64
Sort Ly6c-MHCII+
Sort CD64+
Sort Ly6c-MHCII-
αMHCII
αTim4
αCD45 (I.V)
αLy6c
B
Ly6c-MHCII+ (scRNA-seq)
UMAP_2
UMAP_1
MHCII-
MHCII+
Cycling
Monocytes
cDC
C
Proportion of annotated cells
Total Ly6c-
Ly6c-MHCII-
Ly6c-MHCII+
cDC
Monocytes
Cycling
MHCII+
MHCII-
D
Ly6G
S100a8
Ngp
Expression
MHCII-
MHCII+
Cycling
Monocytes
E
Peripheral blood
Total cells
Singlets
I.V CD45 labeling
SSC-A
FSC-H
FSC-A
90%
αCD45-BUV661
F
C57Bl/6 blood monocytes
Zbt46.zsGFP blood monocytes
CM
IM
NCM
Zbt46.zsGFP
CD43
0%
1%
G
C57Bl/6
ZBT46.zsGFP
Zbt46.zsGFP
αCD64
αMHCII
9%
25%
H
% ZBT positive
i.v. Syn Ly6c-
mono-DC
e.v. Syn Ly6c-
Syn. CD64+
e.v. Syn Ly6c-
i.v. Syn Ly6c-
I
αCD45-BUV661
FMO
αTREM4
αFcgRIV
αFolrb
αVISTA
αLyve1

Figure S3.3. (A) Sorting strategy for $CD45^+CD11b^+Ly6G^-SiglecF^-Tim4^-CD64^-$ into $Ly6c^-MHCII^-$ (lower left pane), $CD64^+$ (middle pane), and $Ly6c^-MHCII^-$ (lower right pane). (B) Annotation of single-cell RNA-seq data from $CD45^+CD11b^+Ly6G^-SiglecF^-Tim4^-CD64^-Ly6c^-MHCII^+$ cells from C57Bl/6 mice, with data from total Syn $Ly6c^-$ cells clusters from Figure 3.2H. (C) Frequency of cells annotated as $MHCII^-$, $MHCII^+$, cDC, monocytes or cycling cells in total Syn $Ly6c^-$ cells, $Ly6c^-MHCII^-$ and $Ly6c^-MHCII^+$ ($CD45^+CD11b^+Ly6G^-SiglecF^-Tim4^-CD64^-Ly6c^-MHCII^+$) cells. (D) Expression of neutrophil-associated genes in subpopulations from the $Ly6c^-MHCII^-$ sorted cells. (E) I.V. labeling of cells in peripheral blood using i.v. anti-CD45 antibody. (F) Expression of GFP in peripheral blood monocytes and (G) synovial cells from representative zbt46.zsGFP and control mice. (H) Quantification of the percent of cells that are ZBT positive (based on GFP expression) in i.v. Syn $Ly6c^-$, mono-DC, e.v. Syn $Ly6c^-$ and Syn $CD64^+$ cells. Values are mean of N>4 ±SEM. P-value was calculated using unpaired t-test. ***=$p<0.005$, ****=$p<0.001$. (I) Surface expression of FMO, TREM4, FcgRIV, Folrb, VISTA, and Lyve1 in i.v. Syn $Ly6c^-$ (blue) and e.v. Syn $Ly6c^-$ cells (red) measured by flow cytometry.

Intra-and extra-vascular synovial $Ly6c^-$ exhibit different functionality

To investigate the properties of the two Syn $Ly6c^-$ populations (i.v. and e.v.) and contrast them with mono-DCs (Syn $Ly6c^-MHCII^+$), we compared the transcriptional profiles of these cells using

bulk RNA-seq. Each population was highly reproducible across replicates and characterized by unique transcriptional profiles (Figures 3.4A, S3.4A). We further confirmed the relationship of the bulk-sorted cells to the subpopulations identified by scRNA-seq using singleR to assign labels based on the bulk data. We merged the $CD45^{+}CD11b^{+}Ly6G^{-}SiglecF^{-}CD64^{-}$ cells from the $Ly6c^{-}MHCII^{+}$ and $Ly6c^{-}MHCII^{-}$ compartments (Figure3. 3) into one dataset and confirmed that the $MHCII^{-}$ cells were assigned to e.v. Syn $Ly6c^{-}$ cells monocytes to i.v. Syn $Ly6c^{-}$ cells and the $MHCII^{+}$ population to mono-DCs (Figure 3.4B). Comparable results were obtained when assessing the level of expression of markers for the single-cell annotations in the bulk dataset (Figure S3.4B). Cycling cells from $Ly6c^{-}MHCII^{+}$ and $Ly6c^{-}MHCII^{-}$ were assigned to mono-DC and e.v. Syn $Ly6c^{-}$, respectively and associated genes were expressed in both populations as well. Taken together, these data demonstrate that there are 3 populations of Syn $Ly6c^{-}$ ($CD45^{+}CD11b^{+}Ly6G^{-}SiglecF^{-}Tim4^{-}CD64^{-}$), i.v. Syn $Ly6c^{-}$, e.v. Syn $Ly6c^{-}$, and mono-DC based on 3 separate scRNA-seq studies as well as bulk RNA-seq.

We then performed k-means clustering on 5127 differentially expressed genes across these populations to define 4 clusters: one each with expression specific to i.v. Syn $Ly6c^{-}$ cells (1), mono-DC (2) and e.v. Syn $Ly6c^{-}$ cells (3) and a final cluster (4) preferentially expressed in both mono-DC and e.v. Syn $Ly6c^{-}$ cells (Figure S4C, Supplemental Table 4A). The i.v. Syn $Ly6c^{-}$ cells expressed the highest levels of Cx3cr, Spn (CD43), Cebpb and Nr4a1, genes associated with PB NCM (Figure3.4C). Further, the bulk sorted mono-DC population preferentially expressed DC-associated genes including Cd74, Cd209a, Zbtb46 and Flt3 (Figure3.4D). Meanwhile, e.v. Syn $Ly6c^{-}$ cells expressed genes observed above in $MHCII^{-}$ cells, such as Alox5, Pf4, Fxyd2 and Rnase4 (Figure 3.4E). The cluster specific to i.v. Syn $Ly6c^{-}$ cells was enriched for genes associated with

collagen fibril organization (Col6a1/2, Col4a1/2, Adamts2), blood vessel morphogenesis (Smad4, Pdgfra, Syk). and blood vessel development (Akt1, Notch2, Foxo1) (Figure 3.4F). As expected, the mono-DC-specific cluster was enriched for genes involved in MHCII antigen presentation molecules (H2-Aa, Cd74), as well as genes involved in myeloid DC differentiation (Irf4, Dcstamp, Btf3), and genes that play a role in T-cell differentiation (Ccr7, Stat5a, Tnfsf9) (Figure 3.4F). For the e.v. Syn Ly6c$^-$ cell-specific cluster, enriched GO processes included complement receptor mediated signaling (C1qa/c, C5ar1, Fcna/b), and chemotaxis and cell migration pathways including Pf4, Pmp22, P2ry12 (Figure 3.4F). Additionally, the leukotriene metabolic pathway was enriched in this cluster with genes including Alox5, Ltc4s, and Ncf1. The cluster shared between mono-DC and e.v. Syn Ly6c$^-$ included cell cycle (Ccna2, Tubb6, Cdc23), DNA repair (Lig1, Hdac9, Cdca5) and cellular component organization or biogenesis (Rpf2, Ipo4, Lmna) (Figures 3.4F, S3.4D).

Next, we sought to identify the human equivalent of the e.v. Syn Ly6c$^-$ population utilizing single-cell data sets on myeloid cells isolated from the human synovium. To help distinguish these cells from that macrophage majority, we first merged the Ly6c$^-$MHCII$^-$ CITE-seq data with CD64$^+$ macrophages from the same mice (Figure S3.4E). We were then able to define an e.v. Syn Ly6c$^-$ module based on the unique gene markers for MHCII$^-$ cells. Previous studies from the AMP consortium presented scRNA-seq data on CD14$^+$ mononuclear cells sorted from human synovium of RA, split into Leukocyte-rich vs. Leukocyte-poor, and OA patients (Figure 3.4G) [227]. We were able to identify a small subset of these cells that resembled e.v. Syn Ly6c$^-$ (Figure 3.4H); OA and Leukocyte-poor RA patients tended to exhibit higher levels of the MHCII module (Figure 3.4I). Similarly, these genes were highest in the NUPR1$^+$ and IFN-activated clusters defined by AMP to

be affiliated with leukocyte-poor RA and OA (Figures S3.4F-G). Next, we performed CITE-seq on sorted $CD45^{+}CD11b^{+}$ synovial cells from an RA patient, which was collected through the RhEumatoid Arthritis SynOvial tissue Network (REASON) [244]. We performed de novo clustering to divide these 1614 cells into 11 clusters (Figure 3.4J). Then, using the same module score approach, we found that cluster 7 expressed the higher levels of e.v. Syn $Ly6c^{-}$ genes (Figure 3.4K, S3.4H). Using the ADT data, we found that these cells exhibit myeloid cell surface markers (Figure S4I), including CD14, and are TIM4+ (Figure 3.4L). These cells also express TimD4 and Trem2 at the RNA level, suggesting they may be related to the $TREM2^{high}$ cells reported by Alivernini et al [228] to be over-represented in Healthy synovium (Figure S3.4J). Overall, these findings demonstrate that Syn $CD14^{+}Tim4^{+}$ cells represent the human ortholog of e.v. Syn $Ly6c^{-}$ cells.

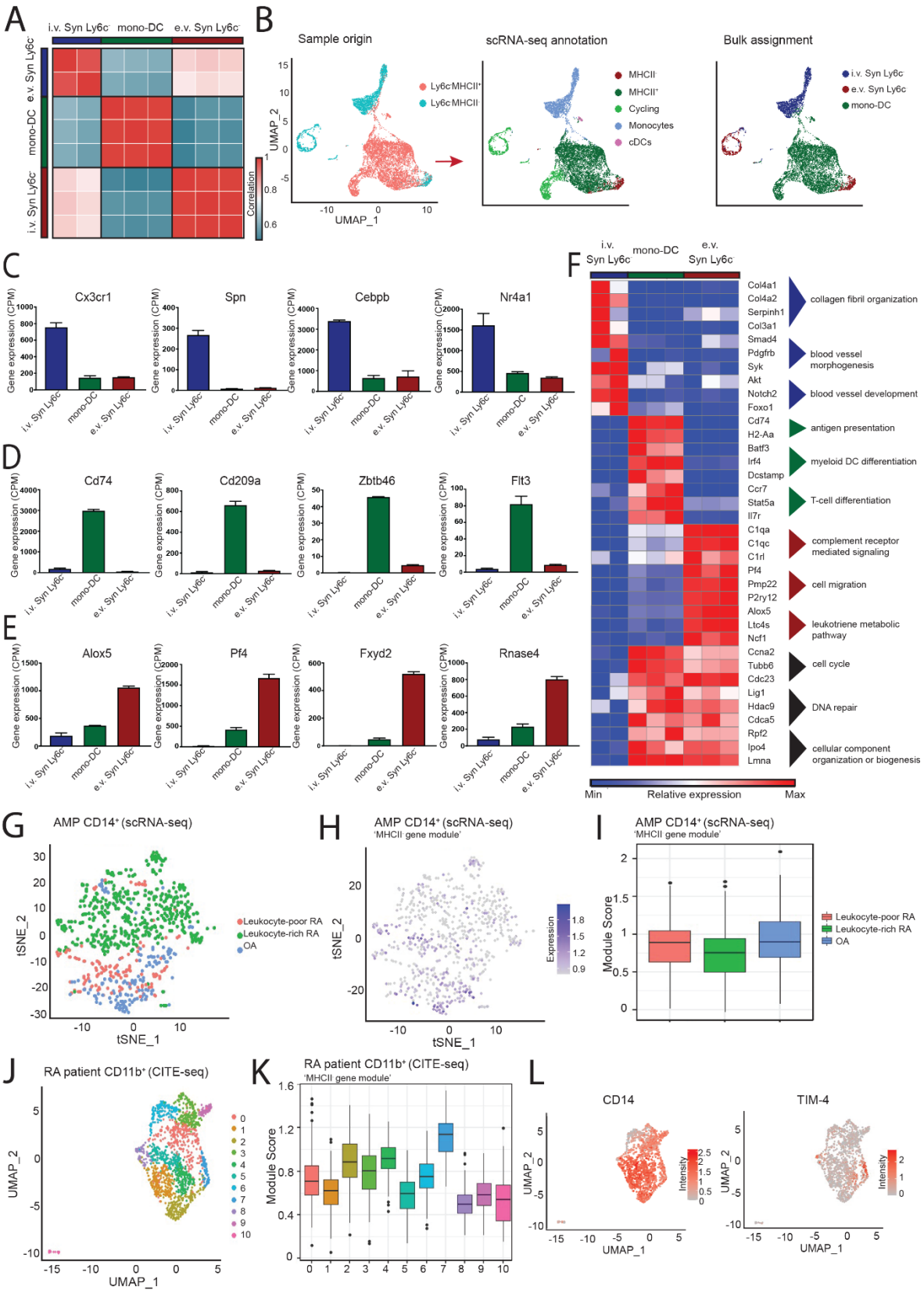

A
i.v. Syn Ly6c⁻
mono-DC
e.v. Syn Ly6c⁻
Correlation
B
Sample origin
scRNA-seq annotation
Bulk assignment
Ly6c⁻MHCII⁺
Ly6c⁻MHCII⁻
MHCII⁻
MHCII⁺
Cycling
Monocytes
cDCs
i.v. Syn Ly6c⁻
e.v. Syn Ly6c⁻
mono-DC
UMAP_1
UMAP_2
C
Cx3cr1
Spn
Cebpb
Nr4a1
Gene expression (CPM)
D
Cd74
Cd209a
Zbtb46
Flt3
E
Alox5
Pf4
Fxyd2
Rnase4
F
i.v. Syn Ly6c⁻
mono-DC
e.v. Syn Ly6c⁻
Col4a1
Col4a2
Serpinh1
Col3a1
Smad4
Pdgfrb
Syk
Akt
Notch2
Foxo1
Cd74
H2-Aa
Batf3
Irf4
Dcstamp
Ccr7
Stat5a
Il7r
C1qa
C1qc
C1rl
Pf4
Pmp22
P2ry12
Alox5
Ltc4s
Ncf1
Ccna2
Tubb6
Cdc23
Lig1
Hdac9
Cdca5
Rpf2
Ipo4
Lmna
collagen fibril organization
blood vessel morphogenesis
blood vessel development
antigen presentation
myeloid DC differentiation
T-cell differentiation
complement receptor mediated signaling
cell migration
leukotriene metabolic pathway
cell cycle
DNA repair
cellular component organization or biogenesis
Min
Relative expression
Max
G
AMP CD14⁺ (scRNA-seq)
Leukocyte-poor RA
Leukocyte-rich RA
OA
tSNE_1
tSNE_2
H
AMP CD14⁺ (scRNA-seq)
'MHCII gene module'
Expression
I
AMP CD14⁺ (scRNA-seq)
'MHCII gene module'
Module Score
J
RA patient CD11b⁺ (CITE-seq)
UMAP_1
UMAP_2
K
RA patient CD11b⁺ (CITE-seq)
'MHCII gene module'
Module Score
L
CD14
TIM-4
Intensity

Figure 3.4 Extra-vascular tissue location confers phenotype of Syn Ly6C- cells. (A) Pairwise Pearson's correlation of global gene expression between replicates of i.v. Syn Ly6c$^-$, mono-DC and e.v. Syn Ly6c$^-$ cells. (B) Merging of scRNA-seq cells from CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$Tim4$^-$CD64$^-$Ly6c$^-$MHCII$^-$ (Figure 3.3A) and CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$Tim4$^-$CD64$^-$Ly6c$^-$MHCII$^+$ cells (Figure S3.3B), annotation by subpopulation, and assignment to i.v. Syn Ly6c$^-$, mono-DC or e.v. Syn Ly6c$^-$ populations using the bulk transcriptional data as reference in singleR. (C) Genes with preferential expression in i.v.NCM, (D) mono-DC and (E) e.v. Syn Ly6c$^-$. (F) Relative expression of representative genes from GO processes associated with i.v. Syn Ly6c$^-$ cells, mono-DC, e.v. Syn Ly6c$^-$ cells. (G) CD14$^+$ scRNA-seq cells originating from Leukocyte-poor RA, Leukocyte-rich RA, and OA patients in published AMP data [227]. (H) Expression of MHCII$^-$ gene module in CD14+ cells from AMP data. (I) MHCII$^-$ module score by cell origin in AMP data. (J) Clustering of CD45$^+$CD11B$^+$ synovial cells based on gene expression from CITE-seq on an ultrasound guided synovial biopsy from a RA patient collected through the RhEumatoid Arthritis SynOvial tissue Network (REASON). (K) MHCII$^-$ module score by cluster in RA patient data. (L) Intensity by ADT count of surface markers CD14 and TIM-4 on RA patient synovial cells.

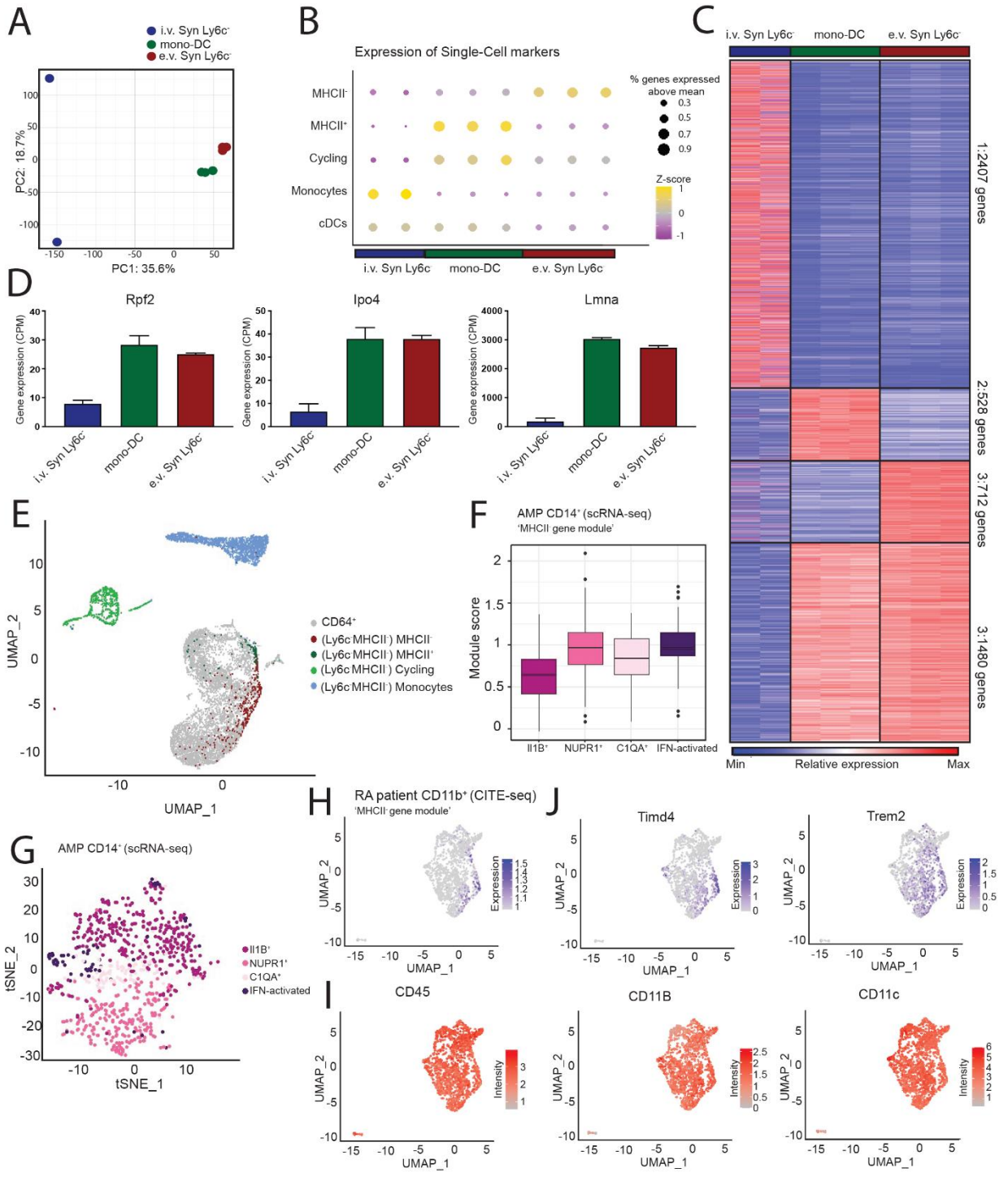


Figure S3.4 (A) PCA of 10270 expressed genes from i.v. Syn Ly6c-, mono-DC and e.v. Syn Ly6c- cells. (B) Expression of marker gene sets from single-cell RNA-seq total Ly6c- subpopulations

(Figure 3.2H) in bulk populations. Color of circle indicates z-score normalized expression while size indicates % of genes in the set expressed in the given sample above their mean value. (C) K-means clustering of 5127 differentially expressed genes across i.v. Syn Ly6c$^-$ cells, mono-DC and e.v. Syn Ly6c$^-$ cells. (D) Example genes with preferential expression in both e.v. Syn Ly6c$^-$ cells and mono-DC. (E) Merging of scRNA-seq cells from CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$Tim4$^-$CD64$^-$ Ly6c$^-$MHCII$^-$ and CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$ CD64$^+$ sorted cells (Figures 3.3A, S3.3B), with Ly6c$^-$ subpopulation annotations. (F-G) CD14$^+$ scRNA-seq cells annotated as IL1B$^+$, NUPR1$^+$, C1QA$^+$, or IFN-activated clusters in published AMP data [227], and their MHCII- module score. (H) Expression of MHCII$^-$ gene module, (I) Intensity by ADT count of surface markers CD45, CD11B, and CD11C and (J) gene expression of Timd4 and Trem2 in CD45$^+$CD11b$^+$ synovial cells from an ultrasound guided synovial biopsy of a RA patient collected through the RhEumatoid Arthritis SynOvial tissue Network (REASON).

LFA1 is required for STIA induced reverse extravasation of e.v. Syn Ly6c$^-$ cells and its deletion reduces pro-inflammatory transcriptional profile of e.v. Syn Ly6C- cells

Using lineage tracing techniques, we confirmed that e.v. Syn Ly6C- cells are derived embryonically, radioresistant, and capable of self-renewal but can be replenished from circulating monocytes when the synovial niche is disrupted. We further observed that the rapid expansion of e.v. Syn Ly6C- cells are critical to the development of inflammatory arthritis. Interestingly, using a modified in vivo intra-vascular labeling system, we uncovered evidence that e.v. Syn Ly6C- cells are capable of reverse transmigration, giving them continuous access to the vasculature. (See original manuscript for full details)

To investigate if contact with vasculature by e.v. Syn Ly6c^{-} mice is required for development of STIA, we used LFA1$^{-/-}$ mice. Previously studies have shown that LFA is required for leukocyte extravasation into tissue and for STIA [279]. LFA1$^{-/-}$ mice failed to develop inflammatory arthritis (Figure 3.5A) consistent with other's work [279, 348] There were no significant differences in the numbers of PB CM, PB NCM, i.v. and e.v. Syn Ly6c^{-} cells between LFA1$^{-/-}$ mice and control mice at steady state (Figure 3.5B-D). However, neither LFA$^{-/-}$ Syn Ly6c^{-} population exhibited the expansion in response to arthritogenic serum (Figure 3.5E) observed in C57Bl/6 mice. In addition, LFA1$^{-/-}$ e.v. Syn Ly6c^{-} cells did not undergo increased labeling with αCD43-BUV395-(IV)) 1hr following STIA as compared to C57Bl/6 mice, suggesting access to vasculature does not increase in response to inflammatory stimulus in LFA1$^{-/-}$ mice (Figure 3.5F).

To determine how LFA1 deletion affected Syn Ly6c^{-} cells on the transcriptional level, we performed scRNA-seq on isolated CD45^{+}CD11b^{+}Ly6G^{-}SiglecF^{-}Tim4^{-}CD64^{-}Ly6c^{-} synovial cells from LFA1$^{-/-}$ mice at steady state. As before, we integrated and superimposed the 6 subpopulations (0-5) defined from C57Bl/6 mice on 3500 LFA1$^{-/-}$ cells (Figures 3.5G, S3.5A). LFA1$^{-/-}$ mice displayed an altered distribution of cells across the 6 subpopulations ($p<2.2x10^{-16}$) (Figure S3.5B-C) but the ratio of MHCII^{+} to MHCII^{-} cells was comparable to C57Bl/6 mice (Figure 3.5H). LFA$^{-/-}$ e.v. Syn Ly6c^{-} cells exhibited decreased expression of genes associated with chemotaxis (Ccl17, Itgb2, Ccl9), defense response (Arg1, Mif, Cfp, Itgax), regulation of cell adhesion (Ccr5, Lgals3, Adam8) and stress response (Pycard, Flt1, Prdx5, Vegfa) (Figure 3.5I-J). In contrast, genes associated with regulation of cell differentiation (Fos, Cd36, Mef2c, Mafb, Csf1r), regulation of angiogenesis (Tcf4, Pf4, Tgfbr2), and response to wounding (Macf1, Aqp1, Cfh) were increased in expression in LFA$^{-/-}$ e.v. Syn Ly6c^{-} cells compared to C57Bl/6 (Figure 3.5I-J). Based on these

data, we demonstrate that e.v. Syn Ly6c^{-} are associated with chemotaxis of leukocytes such as granulocytes to the synovium during inflammation.

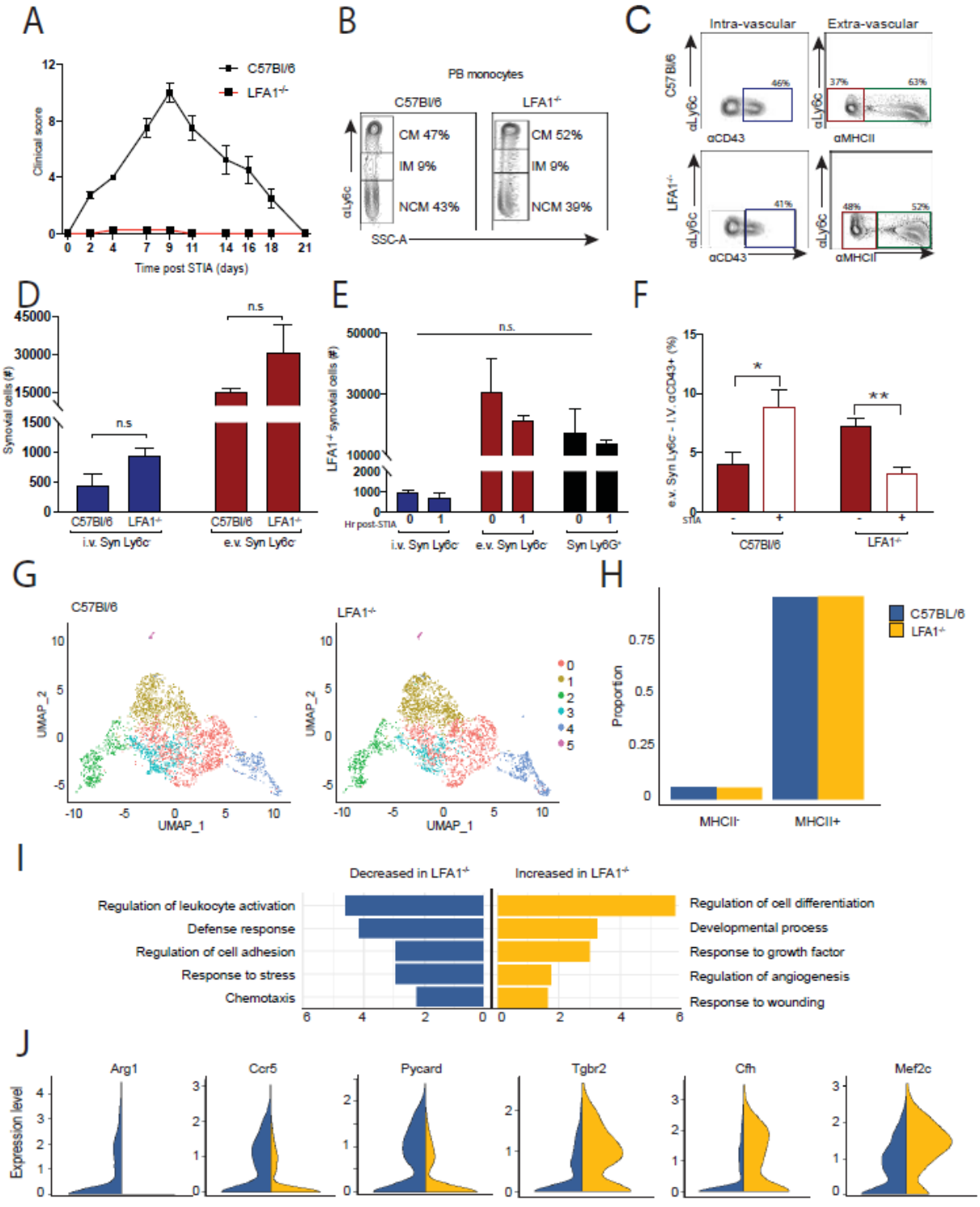


Figure 3.5 Deletion of LFA1 reduced pro-inflammatory phenotype of e.v. Syn Ly6C- cells. (A) STIA clinical score in C57Bl/6 and LFA1-/- mice. (B) PB monocytes and (C-D) i.v. Syn Ly6C-, and e.v. Syn Ly6C- cells in C57Bl/6 and LFA1-/- in steady-state. (E) Number of i.v. Syn Ly6C-, e.v. Syn Ly6C- cells and Syn Ly6G+ cells 1hr post-STIA in C57Bl/6 and LFA1-/- mice. (F) e.v. Syn Ly6C- cells labeled with I.V αCD43-BUV395 in steady state and 1hr post-STIA. (G) Integration of scRNA-seq data on CD45+CD11b+Tim4-CD64- cells from LFA-/- mice with C57Bl/6. To obtain comparable numbers, both datasets were sub-sampled to 3000 cells. (H) Ratio of cells annotated as either MHCII+ or MHCII- in C57Bl/6 and LFA1-/- mice. (I) Selected GO processes associated with differentially expressed genes in the MHCII- compartment (representing e.v. Syn Ly6C- cells) between LFA1-/- and C57Bl/6 mice. (J) Ridge plots of representative genes that are increased or decreased in expression in MHCII- cells from LFA1-/- compared with C57Bl/6 mice. Graphs are mean N>4 +SEM P-value was calculated with unpaired t-test. * = $p<0.05$, ** = $p<0.01$.

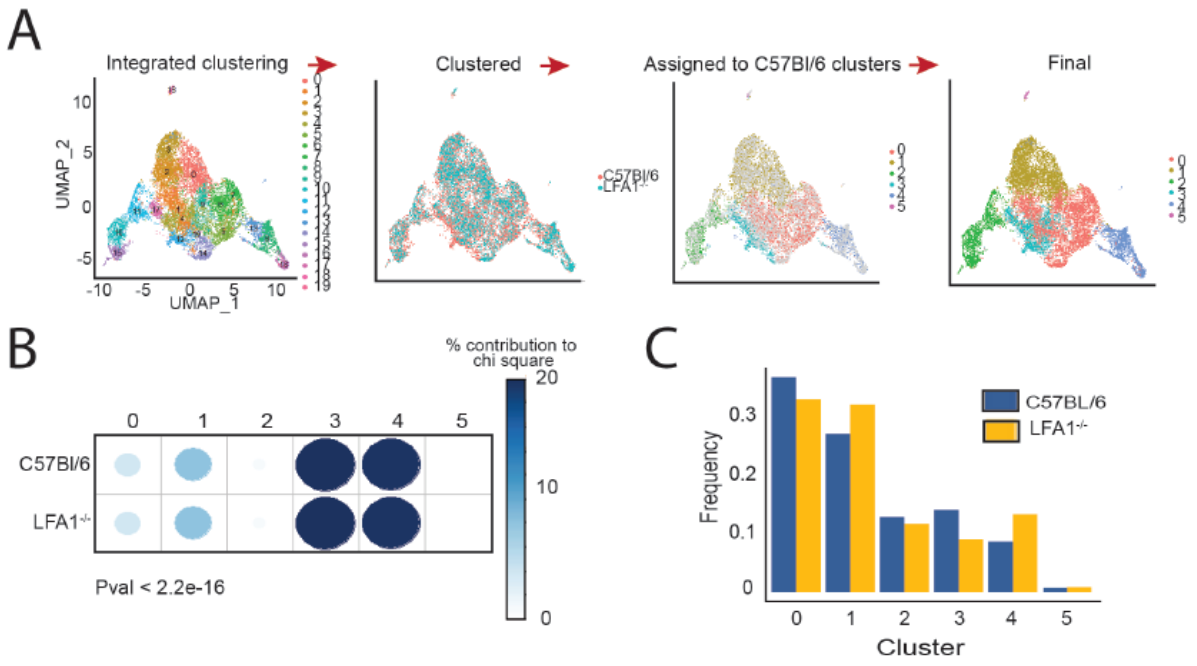


Figure S3.5 (A) Integration of scRNA-seq data from LFA1-/- mice with superimposed C57Bl/6 annotations. (B-C) Proportion of cells annotated as each subpopulation in LFA1-/- and C57Bl/6 mice and chi-squared residuals.

3.4 Discussion

Over the past several years, numerous studies have characterized CM and NCM in circulation and contrasted these with differentiated macrophages in the tissue. Our study is the first to identify two synovial populations that exist in the tissue but do not exhibit canonical macrophage markers. First, we described a distinct population of i.v. Ly6c- cells that are transcriptionally similar to PB NCM and require NR4A1 but remain attached to the vessel wall even after perfusion. Then, through injection of intravenous antibody we confirm the extravascular localization of an e.v. Syn Ly6c$^-$ population that is negative for surface expression of CD64 and Tim4 and is transcriptionally distinct from both circulating monocytes, synovial macrophages and DCs. These Syn Ly6c$^-$ cells are long-lived, terminal cells that do not require CCR2, NR4A1 or LFA for development. In fact, e.v. Syn Ly6c$^-$ cells development may be NOD2 dependent, as this has previously been linked to restoring circulating NCM lacking NR4A1 [349]. Finally, we show a population of cells corresponding to e.v. Syn Ly6c$^-$ cells are present in synovial biopsies from RA patients and their depletion prevents experimental RA in mice. Taken together, our data suggests that e.v. Syn Ly6c$^-$ cells represent a novel population of synovial mononuclear cells involved in pathology of RA.

In order to explore the heterogeneity of tissue Syn Ly6c$^-$ myeloid cells but exclude macrophages, we performed single-cell RNA-seq on CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$CD64$^-$MHCII$^-$ and CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$CD64$^-$MHCII$^+$ mononuclear phagocytes from the synovium

during steady state. Analysis of these data enabled us to distinguish a subpopulation of monocyte-derived dendritic cells from i.v. and e.v. Syn Ly6c$^-$ cells via MHCII expression. We also compared our total Ly6c$^-$ fraction (CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$CD64$^-$Ly6c$^-$) subpopulations with 6 myeloid cell (CD45$^+$CD11b$^+$Ly6G$^-$) populations harvested from STIA mouse joints on disease day 5 [220]. Although subpopulations defined by Culemann were annotated as macrophages, their FACS strategy did not exclude monocytes, thus explaining the high level of overlap between their populations with our CD45$^+$CD11b$^+$Ly6G$^-$SiglecF$^-$Tim4$^-$CD64$^-$ MHCII$^+$ and MHCII$^-$ cells.

Further, two recent studies utilized scRNA seq to characterize human myeloid cells from the joints of RA patients [227, 228]. By using a module score based on marker genes to annotate cells, we identified MHCII$^-$ (e.v. Syn Ly6c$^-$) cells among the 4 monocyte subpopulations defined by Zhang and colleagues [227]. The data support a higher presence of a corresponding e.v. Syn Ly6c$^-$ population in human synovium from leukocyte poor RA and OA tissue, although the study is limited due to the small number of CD14$^+$ cells (750) in the AMP study. The Alivernini group performed scRNA-seq on CD64$^+$CD11b$^+$CD3$^-$CD19$^-$CD20$^-$CD56$^-$CD49$^-$CD117$^-$CD15$^-$ synovial cells from healthy controls as well as from RA patients who are treatment-naïve/resistant or in clinical remission [228]. Based on the transcriptional profile of the cells we identified as e.v. Syn Ly6c$^-$, we propose there is overlap with their definition of Trem2$^+$ cells. However, these prior studies did not enable us to determine whether e.v. Syn Ly6c$^-$ cells could be identified by flow cytometry. Thus, we utilized our CITE-seq data from the REASON studies and determined that the e.v. Syn Ly6c$^-$ cells corresponded with the TIM4$^+$ cells. Our data as well as those from Alvernini et al demonstrate that Timd4 is only expressed in one synovial monocyte/macrophage population (TREM2^{hi}), which are the largest population in the healthy and UPA patients. Moreover, we

observed that the human ortholog of e.v. Syn Ly6c$^-$ cells were present in normal and OA patients at higher proportion than RA patients using the AMP dataset. Since in mice, TIM4$^+$ cells are typically considered macrophages and Timd4 was not uniquely expressed on murine e.v. Syn Ly6c$^-$ cells, these data suggest a lack of conservation of Tim4 between mice and human and that TIM4 may have a different function in mice vs humans. Thus, these provide the foundation for future study of our novel e.v. Syn Ly6c$^-$ cells in the context of human disease.

Prior studies on the role of monocytes in RA have presented conflicting results. As has been observed previously and confirmed in this study, clo-lip prevent the development of RA by ablating all circulating monocytes [225, 334]. Our data expand upon these findings by establishing the redundant role of PB monocytes in inflammatory arthritis and identifying the critical e.v. Syn Ly6c$^-$ population distinct from PB monocytes but susceptible to clo-lip killing. Our previous finding that restoring NCM following monocyte depletion enables the progression of STIA appears to conflict with sensitivity of NCM-deficient NR4A1$^{-/-}$ mice to arthritis [338, 339]. The data presented here resolve this conflict, by confirming the preservation of e.v. Syn Ly6c$^-$ cells in NR4A1$^{-/-}$ mice and by extension of this PB NCM may replenish the e.v. Syn Ly6c$^-$ niche when required. This process has been well established in embryonic macrophages, which can be replenished from bone marrow derived cells following injury [350, 351]

e.v. Syn Ly6C- cells may represent a terminal monocyte population. Although they lack canonical macrophage surface markers CD64, MERTK, and TIM4, e.v. Syn Ly6c$^-$ express the gene for MafB and our results indicate that MAFB is highly active as a TF in these cells. MafB has been previously associated with tissue macrophages over monocytes [47] and is central for suppressing

macrophage proliferation [352]. Our experiments determined that e.v. Syn Ly6C- cells maintain their ability to proliferate and do not differentiate into macrophages. The e.v. Syn Ly6C- cells are also negative for the DC master regulator Zbt46 [353, 354] as compared to the mono-DCs. Collectively, these data support e.v. Syn Ly6c$^-$ cells as a self-renewing terminal monocyte population that is distinct from DCs and macrophages in the tissue.

Increased vascularity and enhanced permeability of the synovium has been associated with RA and experimental models of arthritis [355-357]. In steady-state, access to the synovium is restricted by the size of the particle [358], consistent with failure of the Clo-lip to eliminate synovial macrophages. However, since e.v. Syn Ly6C- cells in the synovium are susceptible to depletion by Clo-lip, we proposed that these cells have access to the vasculature. By using a second i.v. labeling system, we demonstrate labeling of e.v. Syn Ly6c$^-$ cells over time indicating access to the vasculature by antigen sampling across the endothelium as exhibited by CX3CR1$^+$ macrophages [359] or through full reverse transmigration from synovium to the intra-vascular space. One group demonstrated bidirectional transmigration of monocytes across hepatic sinusoidal endothelium [360], and another showed that reverse transmigration contributes to the development of pathogenic foam cells in atherosclerosis [361]. In support of the former, we show that e.v. Syn Ly6c$^-$ cells from LFA deficient mice do not respond to STIA with proliferation or increased vascular labeling, suggesting an e.v. Syn Ly6c$^-$ cell restricted mechanism. This suggests a role for e.v. Syn Ly6c$^-$ cells in surveying the vascular endothelium, maintaining barrier integrity, and responding to potential pathogens. These roles have been established in PB NCM within the vasculature [332] and therefore e.v. Syn Ly6c$^-$ may play a complementary role in the extra-vascular niche.

LFA1 is required for monocyte crawling [362] and has been implicated in the pathogenesis of arthritis [279, 363]. Neutrophils also express LFA1 and are a critical cell for inflammation in STIA, where their influx into joints is LFA1 dependent [348]. Therefore, involvement of neutrophils in e.v. Syn Ly6c$^-$ cell response to STIA cannot be excluded. Nonetheless, these data support that e.v. Syn Ly6c$^-$ cells are involved in inflammatory responses during STIA via LFA1 dependent mechanisms.

Future experiments may be required to further investigate the phenotype of e.v. Syn Ly6c$^-$ cells. Given the highly plastic nature of mononuclear phagocytes, culturing cells *in vitro* results in populations skewed artificially based on culture conditions, and cell differentiation. For these reasons, e.v. Syn Ly6c$^-$ cells were not cultured *in vitro* to assess functionality or morphology. Although beyond the scope of this study, a future approach using *in vivo* imaging may be useful to further characterize the location and morphology of e.v. Syn Ly6c$^-$ monocytes.

To the best of our knowledge, we are the first to identify a population of tissue Ly6c$^-$ cells distinct from DCs and macrophages in mice, with a corresponding population identified in human RA synovium. These e.v. Syn Ly6c$^-$ cells respond rapidly to inflammatory signals, drastically expand in numbers, and traverse perivascular space via an LFA-dependent mechanism in arthritis. Our data support a role for e.v. Syn Ly6c$^-$ cells as instigators of synovial inflammation leading to the pathogenic cascade in inflammatory arthritis.

CHAPTER 4

Defining macrophage subsets in human pediatric cholestatic livers

Chapter 4 contains analytical findings from an original research article entitled “Transcriptional Profiling of Pediatric Cholestatic Livers Identifies Three Distinct Macrophage Populations” (Taylor et al. 2021, PLoS One) and is reproduced here with the permission of the copyright holder.

4.1 Introduction

Macrophages are a heterogeneous and plastic cell population that respond to environmental signals in various cholestatic liver diseases [364-366]. Tissue-resident macrophages of the liver, also termed Kupffer cells, are self-renewing cells that are present in the liver at birth and promote tolerance in homeostasis [367]. In the setting of liver injury, tissue-resident macrophages can adopt a pro-inflammatory state and additional monocyte-derived macrophages may be recruited from the peripheral circulation to the liver [69, 368, 369]. This leads to a heterogeneous population of macrophages that may have distinct functions in disease.

Prior studies have presented conflicting evidence for a role of macrophages in obstructive cholestasis. Recruited monocytes have been shown to have a protective role against infection in the setting of murine bile duct ligation [370]. In contrast, CCR2-mediated recruitment of monocyte-derived macrophages in a murine model of primary sclerosing cholangitis has been implicated in the mechanism of liver injury and fibrosis [371]. Similarly, macrophages have been associated with the pathogenesis of murine parenteral nutrition-associated cholestasis via toll-like receptor 4 (TLR4)-mediated activation [372] and production of interleukin-1 beta (IL-1β) [373]. Furthermore, reduced farnesoid x receptor (FXR) signaling is thought to induce activation of the macrophage inflammasome in cholestasis and endotoxemia, thereby promoting IL-1β release and increasing immune susceptibility in cholestasis [374]. However, the precise subsets of macrophages responsible for cholestatic liver injury and repair have not been fully characterized.

Macrophages have also been more specifically implicated in biliary atresia (BA), an

obstructive cholangiopathy of infants thought to arise from an aberrant immune response to a self-antigen. While there are two major forms of BA, isolated BA (iBA) and syndromic BA (BASM) with associated malformations, evidence supports a similar antigen-driven immune response in both subtypes [375]. Evidence supporting a role for macrophages in this mal-adaptive immune response include the observation that increased numbers of macrophages correlate with poor prognosis in human BA [376-378]. Hepatic macrophages are also increased in the rotavirus-induced murine model of BA [379]. In addition, macrophage depletion in a murine model of BA improved bile duct obstruction [380]. These studies demonstrate a central role for macrophages in promoting liver injury in BA but fail to identify the specific pathogenic versus pro-restorative macrophage subsets.

In this chapter, we defined human liver macrophage heterogeneity in cholestasis by comparing scRNA-seq data between patients with cholestasis from BA or Alagille Syndrome (ALGS, a non-immune etiology of obstructive cholestasis), non-diseased pediatric liver, and previously published normal hepatic macrophages [381]. We identify novel hepatic macrophage subsets in obstructive cholestasis that are distinct from non-diseased macrophages by leveraging the ability of scRNA-seq to define cell sub-populations. We further demonstrate reduced expression of regulatory genes across all cholestatic macrophage subsets that may contribute to loss of immune tolerance in cholestasis. Taken together, our results lay the foundation for future mechanistic studies and development of macrophage-specific immune modulatory therapies.

4.2 Materials and Methods

Human tissue samples (Performed by Sarah Taylor)

Formalin-fixed, paraffin-embedded liver tissue sections from non-diseased donor liver (n = 5), and BA (n = 6), and ALGS (n = 6) patients at the time of liver transplantation were obtained from the pathology archives of Ann & Robert H. Lurie Children's Hospital of Chicago. Fresh liver tissue was obtained from the explanted liver of 3 patients with cholestatic liver disease (2 with BA and 1 with ALGS) and 1 patient with a hepatic tumor at the time of liver transplantation. Laboratory data was collected retrospectively from the hospital admission for liver transplantation. Written informed consent was obtained from each patient's legal guardians including in the study. The study protocol conforms to the ethical guidelines of the Declaration of Helsinki as reflected in a prior approval by the Institutional Review Board of Lurie Children's Hospital of Chicago. All methods were conducted in accordance with the Institutional Review Board's guidelines and regulations.

Flow cytometry and scRNA-seq library construction (Performed by Sarah Taylor)

A total of 1.9 x 107 cells were obtained from digestion of ALGS liver, 2.2 x 107 from BASM, 4.92 x 107 from iBA, and 1.1 x 108 from CL. We stained single cell suspensions from each sample with antibodies to detect cell viability and expression of the CD45 common leukocyte antigen. 90–100,000 live CD45+ cells were collected by fluorescence activated cell sorting with a viability of 94% for ALGS, 84% for BASM, 76% for iBA, and 87% for NC (Figure S4.1). scRNA-seq libraries were prepared using the Single Cell 3' v2 Reagent Kit for BASM and ALGS and the v3 Reagent Kit for iBA and NC (Figure S4.2). Gel Beads in Emulsion containing single cells were generated

by the 10x Genomics Chromium Controller in the Northwestern Next Generation Sequencing Facility. Barcoded libraries were sequenced on the Illumina HiSeq 4000 platform. Raw sequence data was processed using the 10X Genomics Cell Ranger 3.1.0 pipeline for de-multiplexing, trimming, aligning, and mapping to genes. After filtering of the scRNA-seq data 5,027 immune cells in ALGS, 2,633 immune cells in BASM, 5,927 immune cells in iBA, and 4,691 immune cells in NC were detected (Figure S4.2).

Single-cell RNA-seq analysis

To define the hepatic immune cell heterogeneity, we analyzed each single cell library using the Seurat v3 package [174, 199]. Filtering parameters for each sample were set to include genes expressed in > 3 cells. Cells were included with gene counts >200 and < 5000, and with < 20% mitochondrial genome. We next ran the functions LogNormalize (scale factor 10,000), ScaleData, and RunPCA on each dataset. Variability in each principal component was visualized by the ElbowPlot function (Figure S4.2C). Based on this analysis we clustered the cells by the FindNeighbors function (15 dimensions for ALGS, 17 for BASM, 12 for iBA, and 10 for NC) and FindClusters (resolution of 0.5 for each cholestatic sample and 0.2 for NC). Cell clusters were visualized by Uniform Manifold Approximation and Projection (UMAP) using the function RunUMAP. Using lineage-specific marker genes, we annotated each cluster as myeloid (CD68, CEBPB, CLEC9A), T and natural killer (NK) cells (CD3D, CD8A, NKG7), and B cells (CD79A without MZB1) plasma cells (CD79A co-expressed with MZB1), and dividing cells (TOP2A). To confirm our cell assignments we used SingleR [187] to compare all clusters from each patient to the reference bulk transcriptome data from Immgen [155]. We also separately compared our disease-specific myeloid clusters to the Immgen database to further refine our myeloid subset

annotations prior to integrated analysis.

We next performed integrated clustering on the mononuclear phagocyte cells from each cholestatic patient and ran FindIntegrationAnchors and IntegrateData on ALGS clusters 5, 8, 10, BASM clusters 0, 6, 7, and iBA 3, 6, 9, 11, and 12. We determined the conserved genes within each integrated myeloid cluster by the function FindConservedMarkers. To compare our diseased macrophages to normal macrophages, we imported previously published single-cell data on non-diseased adult human liver from 5 donors with a median age of 41.0 years (interquartile range 23.5 to 54.5 years) [381]. We used the same cell-specific annotations and assigned this normal dataset as the reference in further SingleR analysis of our cholestatic macrophages. The degree of similarity between groups was further assessed visually by UMAP and by correlation analysis of shared genes (Morpheus, https://software.broadinstitute.org/morpheus).

We used Monocle 3 [203] for trajectory analysis of non-diseased macrophages, where a principal graph was fitted on to a dimensional-reduced UMAP space and then each cell assigned a “pseudotime” value depending on its relative location on the inferred graph. By grouping cells into 5 clusters based on their pseudotime values, we applied the function FindGeneModules to identify 2 gene modules upregulated at the beginning (pseudotime 0-5) and end (pseudotime (20-25) of the trajectory to best represent the non-inflammatory and inflammatory macrophage profiles, respectively. Finally, to ascertain if differences in the transcriptional signatures may be secondary to patient age, we compared macrophages from the NC liver sample to the adult normal macrophages by correlation and pseudotime analyses as described above.

Immunofluorescence and quantification of macrophage subsets (Performed by Sarah Taylor)

We next characterized protein expression for genes that differentiated the 3 cholestatic macrophage subsets by immunofluorescence using the Vectra Multispectral Imager in the Northwestern Immunotherapy Assessment Core. Baking and dewaxing were performed on formalin-fixed, paraffin-embedded tissue sections. Using the Opal 7-color automation kit (Akoya Biosciences, Marlborough, MA, USA) slides were stained for CD68 (Abcam ab955), CD69 (Abcam ab233396), C1Q (Abcam ab268120), and S100A8/9 (Abcam ab22506). Whole slide fluorescent imaging was performed followed by multispectral imaging of three 2.01 mm x 1.5 mm areas per slide (Phenochart and Vectra software). We next used inForm software to phenotype the cells and analyzed the cell data with phenoptrReports 0.2.9 package in R. Based on gene expression data we defined LAM on histology as CD68+C1Q+S100A8/9-CD69-, MLM as CD68+C1Q-S100A8/9+CD69-, and AM as CD68+C1Q+/-S100A8/9-CD69+. Using these definitions, we compared abundance on histology by disease group.

4.3 Results

Increased macrophage numbers in obstructive cholestasis as compared to healthy liver controls

We performed immunohistochemistry on histology samples from donor livers, and BA and ALGS patient livers at the time of liver transplantation to determine whether the hepatic macrophage population is expanded in cholestatic liver disease (Figure 4.1A). No histologic abnormality was present among donors with the exception of one individual liver which exhibited hepatocyte swelling. Mean age for donor patients rounded down to the nearest month was 68 months (SD 113, n = 5). No laboratory data was available for donor controls. All BA and ALGS

liver samples had prominent fibrosis or cirrhosis at the time of tissue collection. Mean age rounded down to the nearest month for BA patients was 7 months (SD 1, n = 6) and 105 months (SD 78, n = 6) or 8 years and 9 months for ALGS cases. Difference in age between the 3 groups was not statistically significant by ANOVA (p = 0.12). Mean direct bilirubin within 24 hours prior to transplant was not significantly different between disease groups at 9.6 mg/dL (SD 8.0, n = 6) for BA and 12 mg/dL (SD 7.7, n = 6) for ALGS (p = 0.65 by paired t-test). We found increased number of CD68+ macrophages in BA liver as compared to control with a mean of 1332 cells/mm^2 in BA versus 601 cells/mm2 in non-diseased pediatric liver tissue (p = 0.04) (Figure 4.1B). While ALGS samples also exhibited greater numbers (1040 cells/mm2) of CD68+ macrophages, it did not reach significance compared to control (Figure 4.1B). The pronounced influx of macrophages in cholestatic liver disease suggests they may play a pathogenic role in cholestatic-induced liver injury.

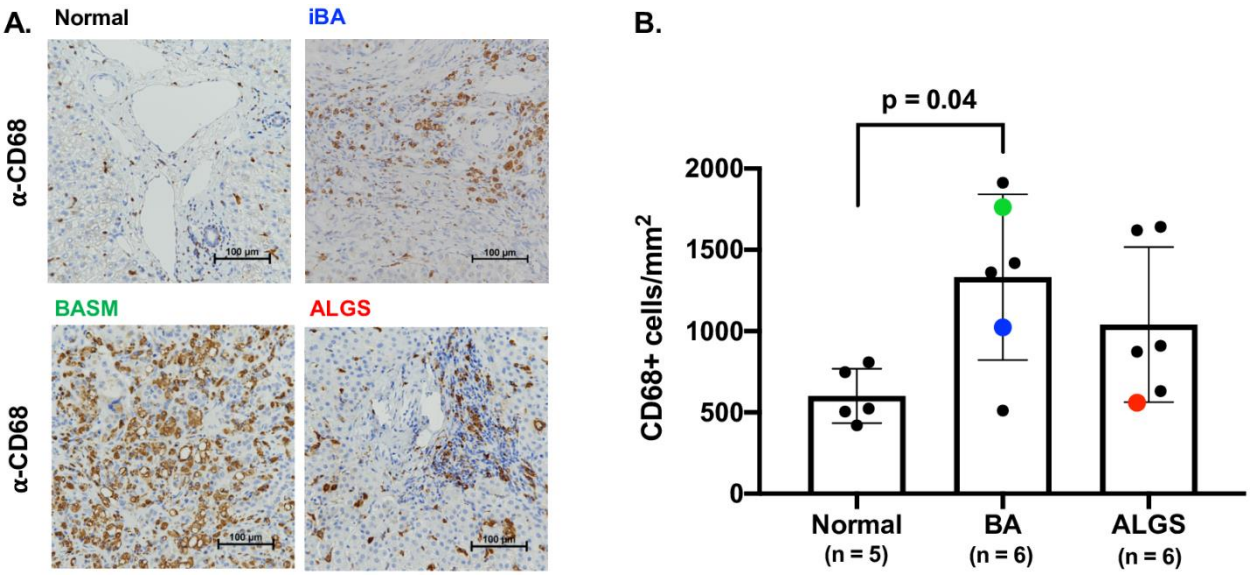


Figure 4.1 Increased hepatic macrophages in cholestatic liver disease. (A) Representative immunohistochemistry staining with the macrophage marker anti-CD68 in samples taken at the time of liver transplantation from the iBA, BASM, and ALGS patients also used for scRNA-seq are shown compared to a normal donor liver sample. (B) Quantitative analysis of entire sections from wedge biopsies showed a significantly increased number of CD68+ macrophages in BA patients, with individual samples processed for scRNA-seq shown in blue (iBA), green (BASM) and red (ALGS).

Variable immune cell composition between BASM and ALGS

We next performed scRNA-seq on CD45+ live cells isolated from each liver sample to better evaluate immune cell infiltration in obstructive cholestasis (Figure 4.2A). We classified single-cell clusters into 5 immune cell types and a population of dividing cells in the cholestatic liver samples using lineage-specific marker genes (Figure 4.2B-C). Different clusters of the same cell type were highly correlated within each sample and between the samples thereby supporting the lineage annotation (Figure S4.2A-B). Further, the lineage annotations were confirmed by SingleR [187], which compares each cell against a reference dataset of population-level transcriptional profiles (in this case, the Immgen database [155]) (Figure S4.2C). Lastly, one cluster in each patient expressed high levels of cell cycle genes [382], which would indicate dividing cells (Figure S4.2D). T and NK cells were the most abundant immune cell population in both samples, comprising 73%, 48%, and 54% of total immune cells in ALGS, BASM, and iBA respectively (Figure 4.2D). Mononuclear phagocytes were the next largest population in BASM and iBA, but not in ALGS. This discrepancy may reflect the difference in disease etiology.

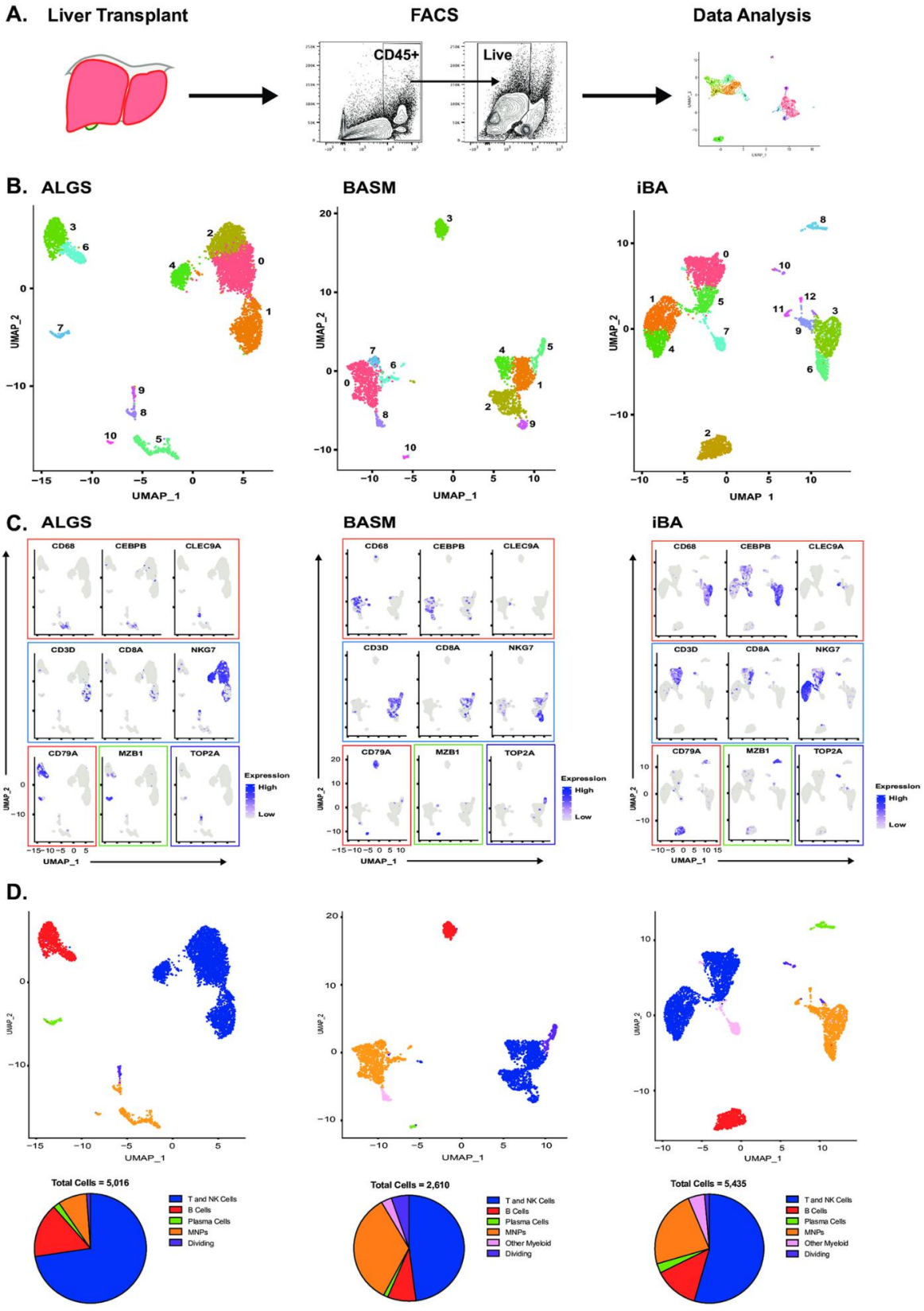

A.
Liver Transplant
FACS
CD45+
Live
Data Analysis
B.
ALGS
BASM
iBA
UMAP_1
UMAP_2
C.
ALGS
BASM
iBA
CD68
CEBPB
CLEC9A
CD3D
CD8A
NKG7
CD79A
MZB1
TOP2A
Expression
High
Low
D.
Total Cells = 5,016
Total Cells = 2,610
Total Cells = 5,435
T and NK Cells
B Cells
Plasma Cells
MNPs
Other Myeloid
Dividing

Figure 4.2 Single-cell RNA-seq enables immune cell characterization in cholestatic liver disease. (A) Hepatic CD45+ cells were isolated from liver tissue at the time of liver transplantation by FACS for scRNA-seq analysis. (B) UMAP of scRNA-seq data showing 11 clusters in ALGS (left) and BASM (middle) and 13 clusters in iBA (right) patient samples. (C) Clusters were assigned to cell types based on the expression of lineage-specific genes (blue = T/NK cells; red = B cells; green = plasma cells; orange = MNP; pink = other myeloid cells; purple = dividing cells) in ALGS, BASM, and iBA (left to right). (D) UMAPs were re-colored by cell type and proportion of immune cells demonstrates greater numbers of MNP cells in BASM (middle) as compared to ALGS (right) and iBA (right).

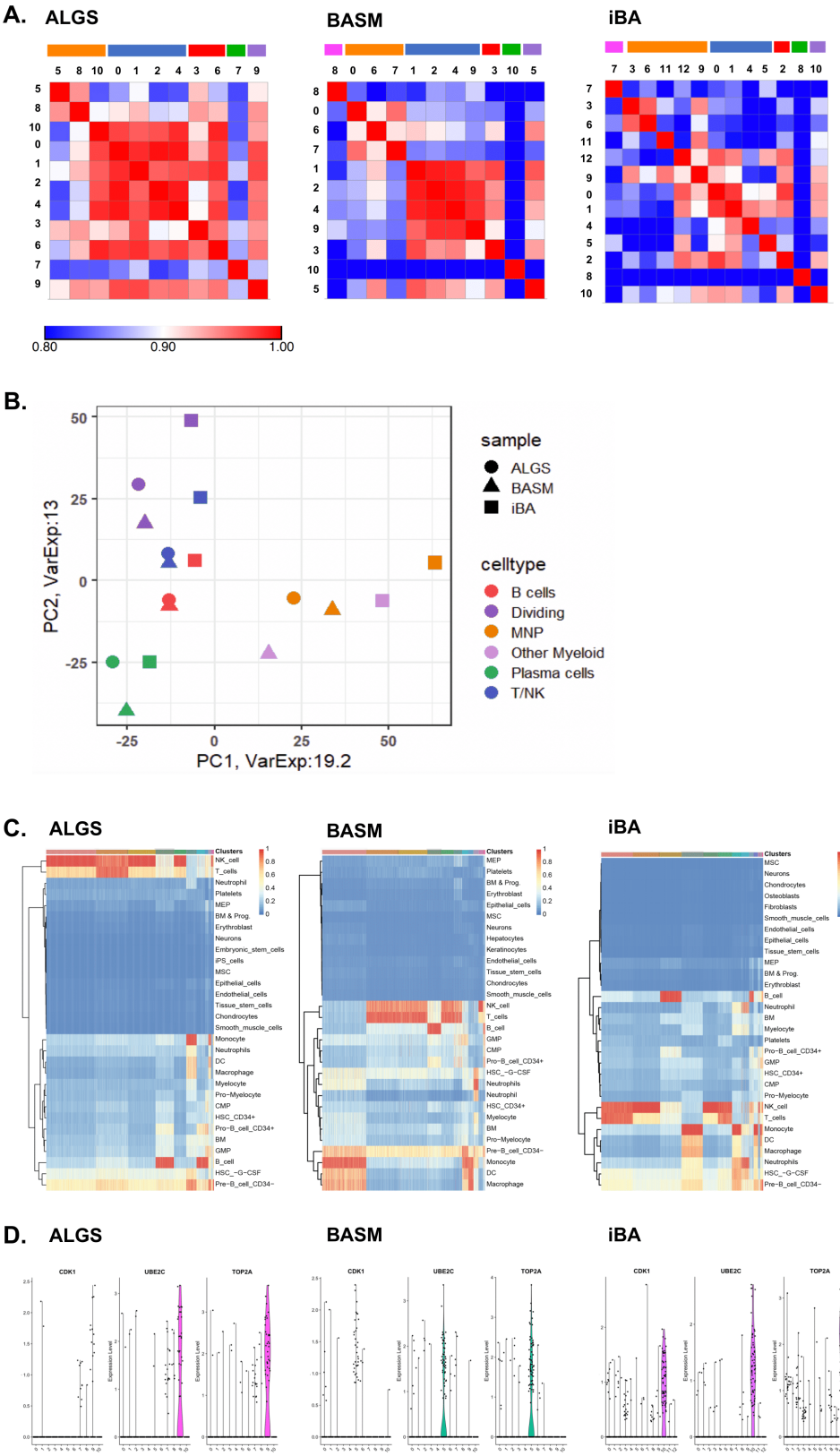

A.
ALGS
BASM
iBA
0.80
0.90
1.00
B.
sample
ALGS
BASM
iBA
celltype
B cells
Dividing
MNP
Other Myeloid
Plasma cells
T/NK
PC2, VarExp:13
PC1, VarExp:19.2
C.
ALGS
BASM
iBA
D.
ALGS
BASM
iBA
CDK1
UBE2C
TOP2A

Figure S4.2: (A) Pairwise Pearson's correlation of average gene expression between each cluster in ALGS, (left), BASM (middle), and iBA (right) organized by cell type annotation (pink = other myeloid; orange = MNP; blue = T/NK cells; red = B cells; green = plasma cells; purple = dividing cells). (B) Clustering of cell types between patients by principal component analysis. (C) SingleR analysis of clusters from ALGS (left), BASM (middle), and iBA (right) compared to Immgen database reference dataset confirmed our cell cluster assignments. (D) Dividing cells were identified in each patient sample by expression of the cell cycle genes CDK1, UBE2C, and TOP2A.

Three distinct macrophage populations in obstructive cholestasis

To better understand macrophage heterogeneity in obstructive cholestasis, we focused our analysis on the clusters annotated as MNP and other myeloid cells. Our SingleR results suggested these clusters contained a mixture of macrophages, dendritic cells (DCs), and neutrophils (Figure 4.3A). For further analysis, we excluded neutrophils, which were found in BASM (cluster 8) and iBA (cluster 7) and defined by distinct expression of neutrophil genes, such as FCGR3B and S100P [383], and lack expression of macrophage genes, such as CD68 and CTSB (Figure 4.3B). We then performed integrated clustering on MNP cells from all patients to define 3 macrophage subsets and 3 dendritic cell subsets (Figure 4.3C-D, S4.3). Three macrophage clusters were identified by the lineage-specific markers CD68, CEBPB, CD14, and CD69, (Figure4.3C-D). The dendritic cells were annotated using markers described previously [384] to identify a CD1C positive subset, CLEC9A positive subset, and plasmacytoid DC (pDC) subset (Figure 4.3C-D). All macrophage populations were represented in each patient (Figure 4.3E, S4.3). Together, these findings suggest common macrophage subsets may arise from environmental cues in the setting of cholestasis.

We next sought to characterize the cross-disease transcriptional signature of each inflammatory macrophage subset and defined MΦ1 as lipid-associated macrophages (LAM), MΦ2 as monocyte-like macrophages (MLM), and MΦ3 as adaptive macrophages (AM) (Figure 4.4A - B). LAM demonstrated the highest expression of genes associated with lipid metabolism including APOC1, APOE, LGMN, FABP5. There was also high overlap with genes previously reported in LAM from human adipose tissue including TREM2 (Figure S4.4). [385] MLM were defined by genes previously identified in monocytes, including S100A8, S100A9, VCAN [386-388]. Finally, AM were enriched for genes associated with lymphocyte activation including CD2, CD7, CCL5, CCL4, CD3D, IL7R. As we have previously defined these immune cells as macrophages, the increased expression of genes involved in adaptive immunity suggest these cells may have engulfed lymphocytes or play a role in regulation of lymphocyte response.

To validate these three populations across cholestatic liver disease, we performed immunofluorescence on a large cohort of patients. We chose markers for each population based on their differential gene expression by scRNA-seq (Figure 4.4B). Using these markers, we demonstrated the presence of all subsets across the fixed BA and ALGS samples from Figure 4.1 through overlap with CD68 expression (Figure 4.4C). Since not all individual cells in a population expressed the relevant marker, we expect this approach to have lower sensitivity than specificity as supported by differences between histology and gene expression analyses for the BASM, iBA, and ALGS samples (Figure 4.4D). Thus, the percent of each population is likely to be an underestimate and may explain the proportion of CD68+ cells not assigned to any population. Despite these differences, comparing the number of cells in each population between 6 BA and 6 ALGS patients shows that LAM tends to account for a greater proportion of macrophages in BA

(Figure 4.4E). In contrast, the AM population is larger on average in ALGS patients. Further study is required to determine whether this difference reflects disease pathogenesis.

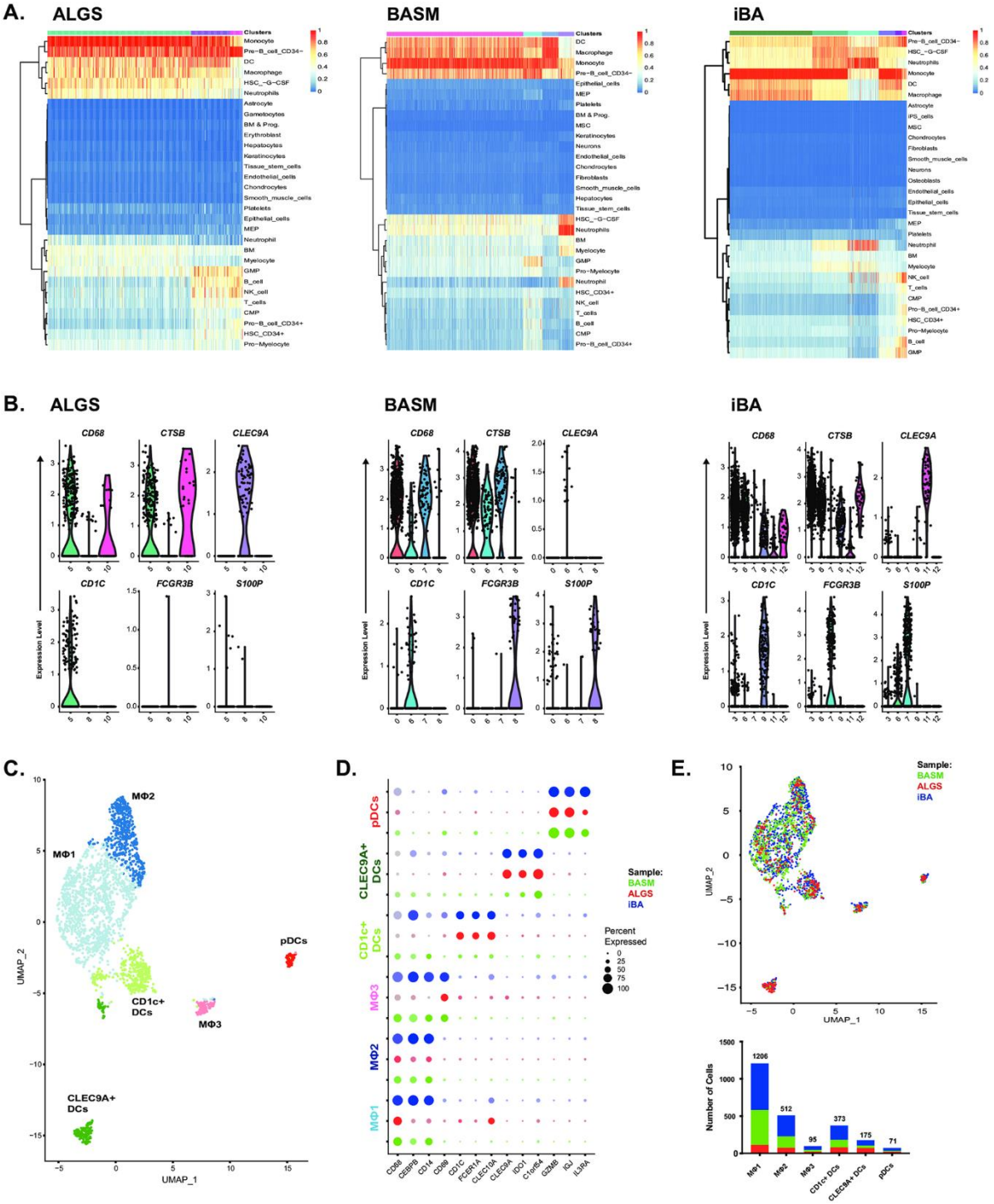


Figure 4.3 Integrated analysis of myeloid cells across patients identifies 3 distinct macrophage subsets in cholestasis. (A) Comparison of annotated ALGS (5, 8, and 10, left), BASM (0, 6, 7, and 8, middle) and iBA (3, 6, 7, 9, 11, and 12, right) myeloid cell clusters with reference data set identified neutrophil clusters in BASM and iBA. (B) While all other clusters expressed macrophage and/or dendritic cell markers, BASM cluster 8 and iBA cluster 7 expressed neutrophil markers FCGR3B and S100P. (C) UMAP of remaining MNP cells showed 6 integrated clusters. (D) Expression of key markers enabled identification of CD1c+ DCs (light green), CLEC9A+ DCs (dark green), pDCs (red), and 3 macrophage subsets (light blue, dark blue, and pink) across ALGS (red), BASM (green), and iBA (blue) patients. (E) BASM (green) and iBA (blue) cells represented the majority in the macrophage clusters.

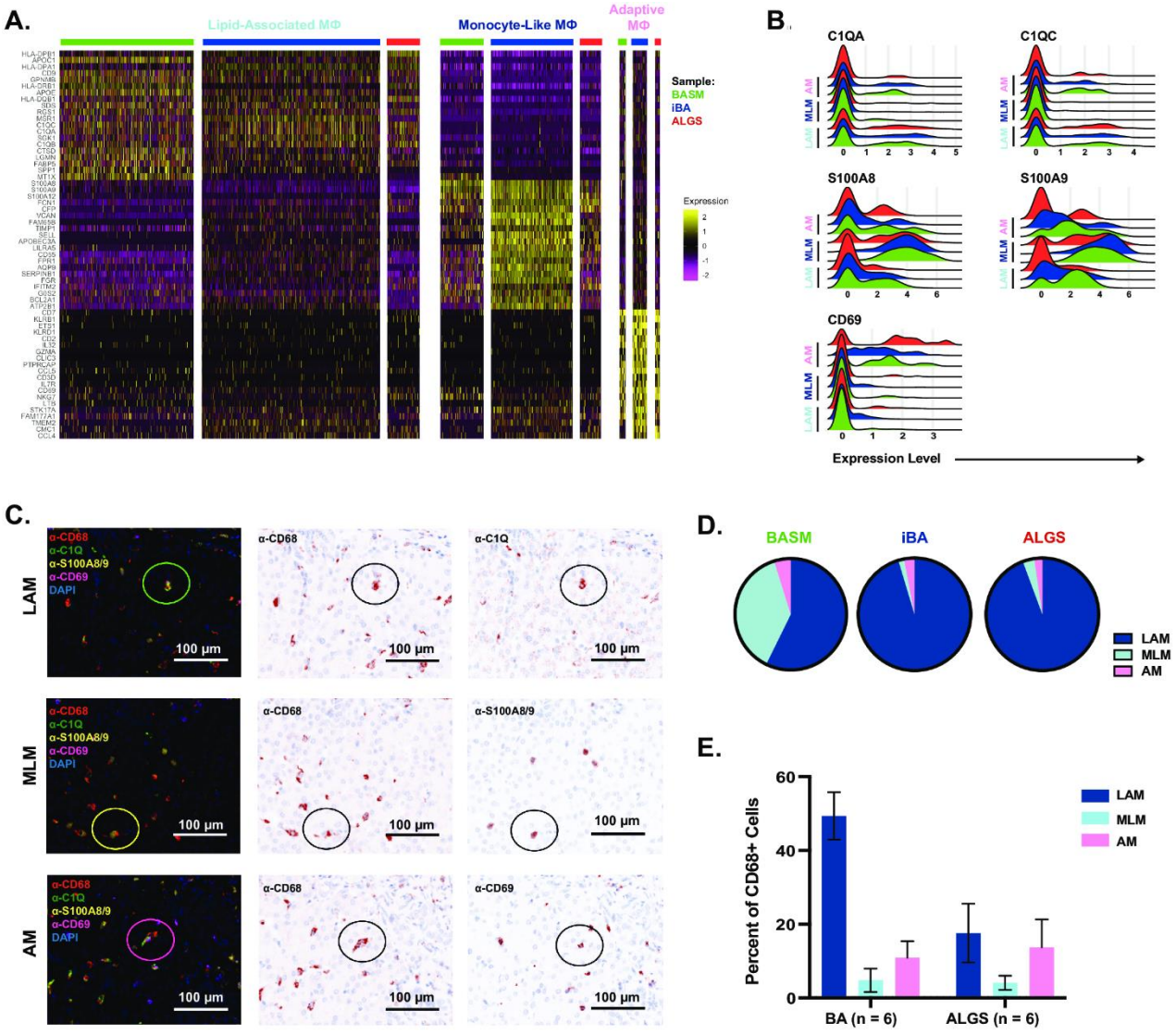


Figure 4.4. The transcriptional signature of macrophage subsets is conserved across patients. (A) Each macrophage subset exhibited a unique transcriptional signature that was similar between ALGS (red), BASM (green), and iBA (blue) patients. (B) Ridge plot demonstrates the expression of genes upregulated in each subset including C1QC and C1QA in lipid-associated macrophages, S100A8 and S100A9 in monocyte-like macrophages, and CD69 in adaptive macrophages. (C) Representative images of immunofluorescence identifying co-localization of each protein marker with anti-CD68 to identify LAM, MLM, and AM. (D) The relative contribution of each subset to

total cholestatic macrophages was compared between the BASM, iBA and ALGS patients with corresponding scRNA-seq data (left to right). (E) Average percent of total CD68+ cells with standard error of mean for each macrophage subset in 6 BA and 6 ALGS patients with prior CD68 quantification.

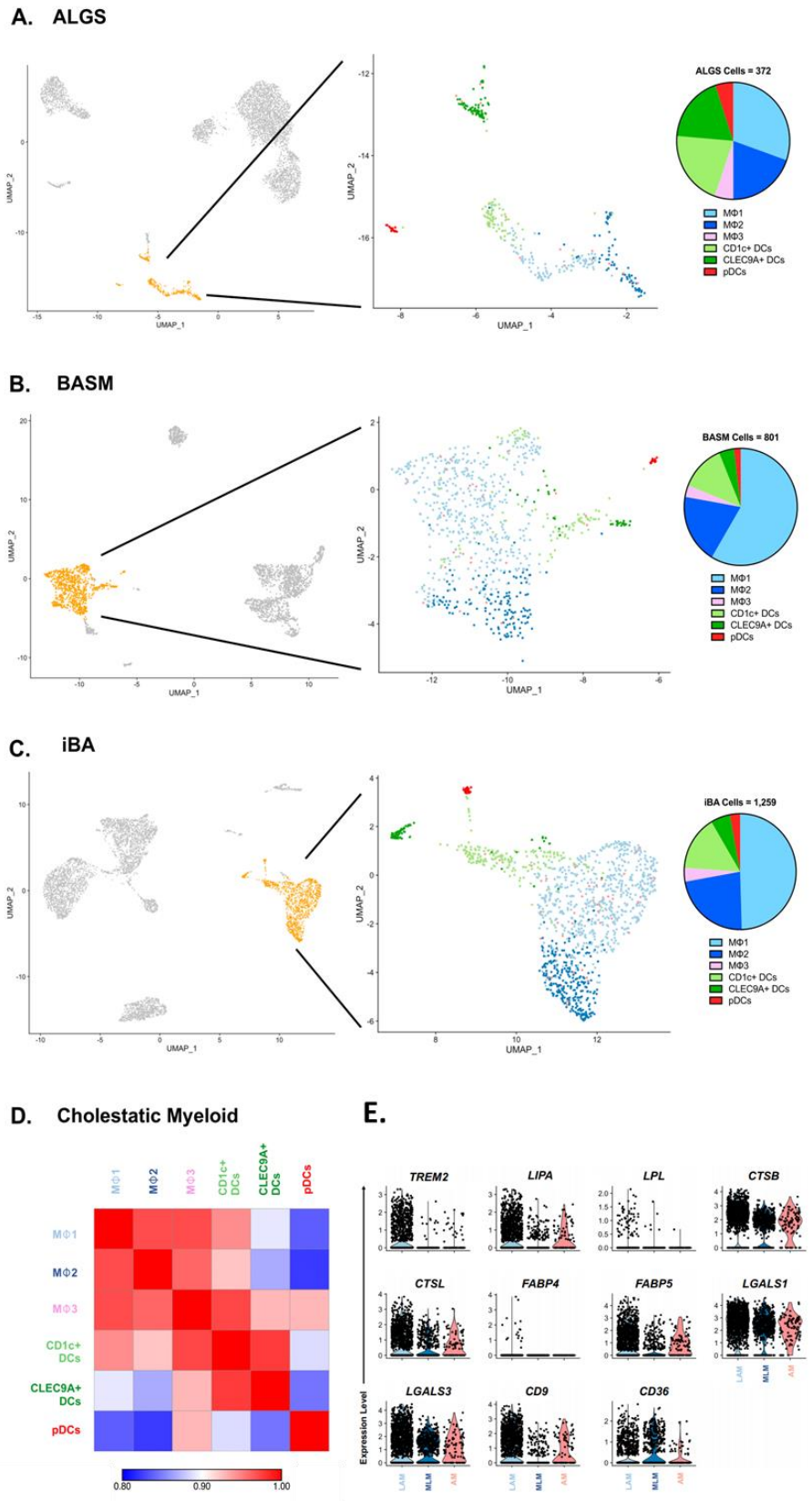

A. ALGS
ALGS Cells = 372
B. BASM
BASM Cells = 801
C. iBA
iBA Cells = 1,259
MΦ1
MΦ2
MΦ3
CD1c+ DCs
CLEC9A+ DCs
pDCs
UMAP_1
UMAP_2
D. Cholestatic Myeloid
0.80
0.90
1.00
E.
TREM2
LIPA
LPL
CTSB
CTSL
FABP4
FABP5
LGALS1
LGALS3
CD9
CD36
Expression Level
LAM
MLM
AM

Figure S4.4 (A) Visualization of clusters from integrated analysis of myeloid cells on original UMAP from Fig 2 and the proportion of MNP cells for ALGS, (B) BASM, and (C) iBA. (D) Pairwise Pearson's correlation of average gene expression between integrated cholestatic myeloid clusters. (E) Violin plots showing expression of genes associated with lipid-associated macrophages in human adipose tissue are most highly expressed in lipid-associated macrophages from cholestatic livers [385].

Reduced expression of immune-regulatory genes in obstructive cholestasis as compared to non-diseased human liver

We took advantage of single-cell data that was previously published using non-diseased adult livers [381] to determine how macrophages from cholestatic livers compared to those from healthy livers. We reproduced the 20 clusters from the original study of which 2 were labelled as "inflammatory" (IM) and "non-inflammatory" (NM) macrophages (Figure S4.5A). Although this data included all cell types, not just CD45+ cells, annotation of immune cell types using lineage-specific markers led to analogous results (Figure S4.5B). To overcome technical variability between datasets limiting the utility of co-clustering, we used SingleR, Correlation analysis, and single gene and gene set comparisons to evaluate similarities and differences between macrophage subsets. All 3 populations of cholestatic macrophages were more similar to the IM than NM (Figure 4.5A-B); AM was the least correlated overall (0.84) compared to LAM (0.89) and MLM (0.89) (Figure 4.5B).

To determine whether the differences between datasets was due to older age of the controls, we performed scRNA-seq on a pediatric non-cholestatic (NC) pediatric liver sample. The NC case

was an 11-year-old female whose explanted liver demonstrated some areas of necrosis consistent with changes after chemotherapy and chronic inflammation with margins negative for tumor. Through a comparable scRNA-seq analysis workflow, we identified two populations of macrophages, which we label Ped1 and Ped2 (Figure 4.5B and S4.5C-G). Unlike the cholestatic macrophages, these populations clearly recapitulate the dichotomy of adult NM and IM (Figure 4.5B). Moreover, while all cholestatic macrophages demonstrated decreased expression of immunoregulatory genes (MARCO, HMOX1, and CD5L), Ped2 expressed these genes at comparable levels to NM (Figure 4.5C). The cholestatic populations, LAM and AM, exhibited distinct transcriptional signatures from both adult and pediatric macrophages subsets (Figure 4.5C). In contrast, the genes that defined MLM were also increased in adult IM and Ped2 (Figure 4.5C). Interestingly, expression of NR1H4, which encodes FXR and is thought to play a role in macrophage inflammasome activation in cholestasis, is negligible across all macrophages (Figure S4.6A). Taken together, our findings support the emergence of disease-specific macrophages in cholestasis that may mediate inflammation via different pathways than FXR signaling.

Although transcriptionally distinct, macrophages in the diseased liver may be derived from their healthy counterparts. Using Monocle, we defined a pseudotime trajectory beginning in NM (pseudotime 0) and ending in IM (pseudotime 25) (Fig 5D, Figure S4.6B). We then identified 2 modules associated with high expression at these endpoints: the non-inflammatory module included genes such as CD5L, MARCO, and VCAM1 whereas the inflammatory module included LYZ, S100A8, and VCAN (Figure 4.5E, S4.6). In support of the limited effect of age on healthy macrophage heterogeneity, the former modules were highest in Ped1, while the latter was highest in Ped2. In contrast, we found that no cholestatic macrophage subset expressed high levels of the

non-inflammatory module (Figure 4.5F). However, MLM exhibited high expression of the inflammatory module, possibly indicating a common origin with IM (Figure 4.5F). This analysis demonstrates the transcriptional variability across cholestatic macrophages beyond the dichotomy of healthy liver macrophages.

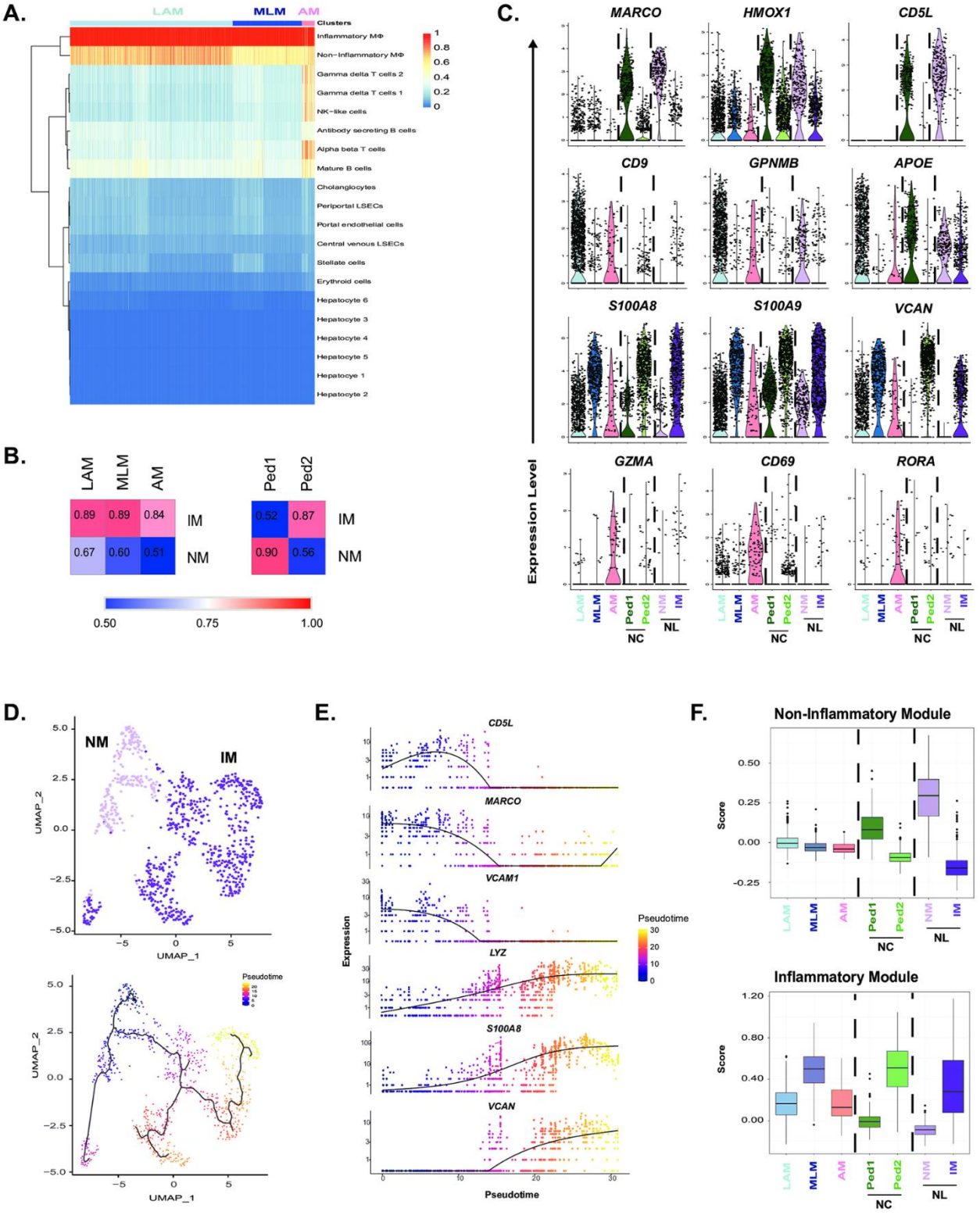

A.
LAM
MLM
AM
Clusters
Inflammatory MΦ
Non−Inflammatory MΦ
Gamma delta T cells 2
Gamma delta T cells 1
NK−like cells
Antibody secreting B cells
Alpha beta T cells
Mature B cells
Cholangiocytes
Periportal LSECs
Portal endothelial cells
Central venous LSECs
Stellate cells
Erythroid cells
Hepatocyte 6
Hepatocyte 3
Hepatocyte 4
Hepatocyte 5
Hepatocye 1
Hepatocyte 2
1
0.8
0.6
0.4
0.2
0
B.
LAM
MLM
AM
Ped1
Ped2
0.89
0.89
0.84
0.52
0.87
IM
0.67
0.60
0.51
0.90
0.56
NM
0.50
0.75
1.00
C.
MARCO
HMOX1
CD5L
CD9
GPNMB
APOE
S100A8
S100A9
VCAN
GZMA
CD69
RORA
Expression Level
NC
NL
D.
NM
IM
UMAP_2
UMAP_1
Pseudotime
E.
CD5L
MARCO
VCAM1
LYZ
S100A8
VCAN
Expression
Pseudotime
F.
Non-Inflammatory Module
Inflammatory Module
Score
NC
NL

Figure 4.5 Cholestatic macrophages are distinct from non-diseased hepatic macrophages. (A) All three cholestatic macrophage subsets were primarily assigned by SingleR to previously published inflammatory macrophages (IM) in non-diseased liver. (B) The mean gene expression of each cholestatic macrophage subset was more highly correlated with inflammatory macrophages (IM) than non-inflammatory macrophages (NM) (left). Of the two macrophage subsets in pediatric non-cholestatic (NC) liver, Ped1 was most similar to NM, while Ped2 was similar to IM (right). (C) Violin plots of individual genes that define similarities and differences in expression between cholestatic macrophage subsets as compared to healthy adult and pediatric NC macrophages. (D) Pseudotime analysis of healthy adult macrophages given a beginning (indicated with black circle) at NM (light purple) inferred a trajectory ending in IM (dark purple). (E) The non-inflammatory module contained genes with expression peaking at pseudotime 0–5, while the inflammatory module peaked at pseudotime 20–25. (F) The non-inflammatory module is lowly expressed across all cholestatic macrophages whereas MLM demonstrated high expression of the inflammatory module.

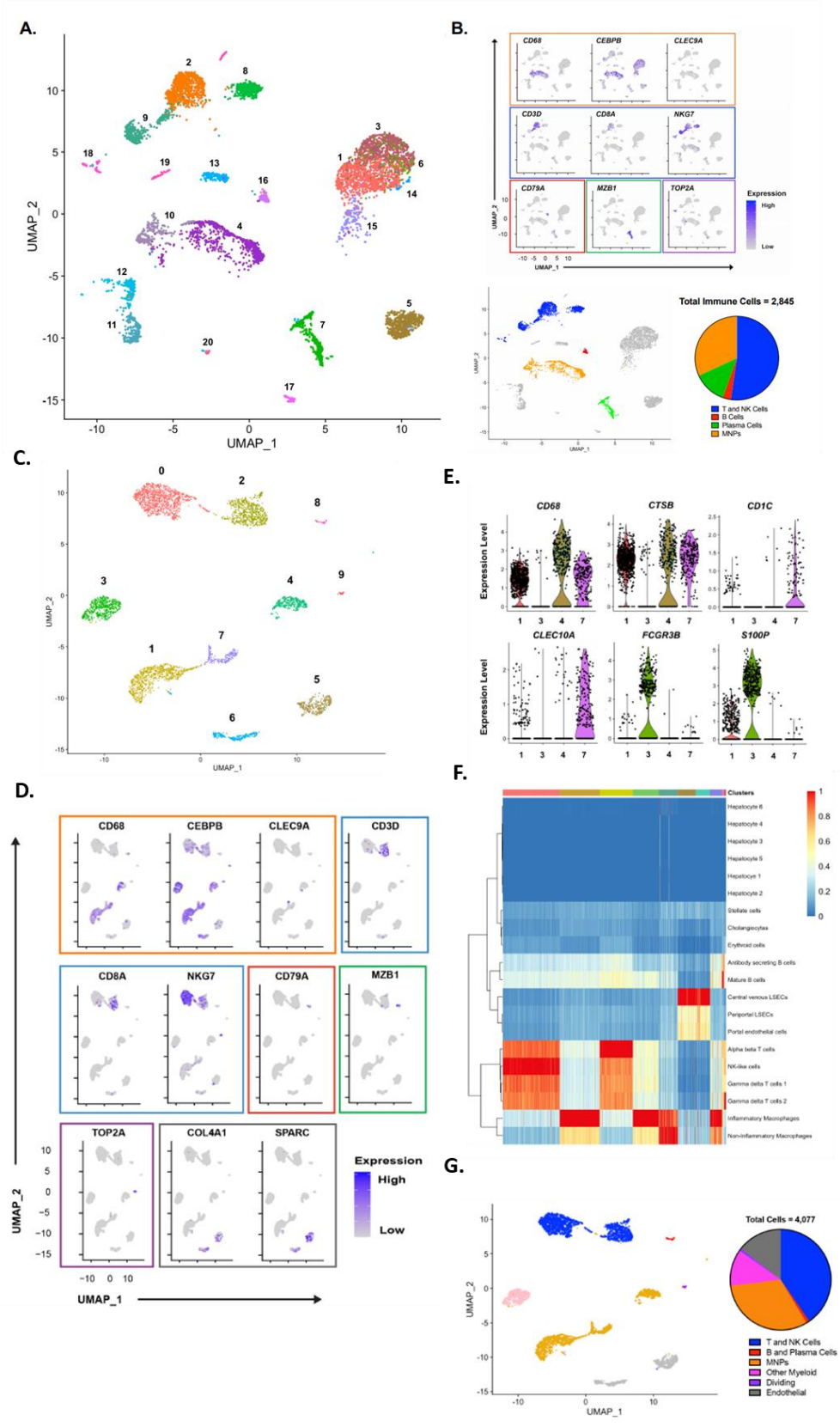

A.
UMAP_2
UMAP_1
B.
CD68
CEBPB
CLEC9A
CD3D
CD8A
NKG7
CD79A
MZB1
TOP2A
Expression
High
Low
Total Immune Cells = 2,845
T and NK Cells
B Cells
Plasma Cells
MNPs
C.
D.
CD68
CEBPB
CLEC9A
CD3D
CD8A
NKG7
CD79A
MZB1
TOP2A
COL4A1
SPARC
Expression
High
Low
UMAP_2
UMAP_1
E.
CD68
CTSB
CD1C
CLEC10A
FCGR3B
S100P
Expression Level
F.
Clusters
Hepatocyte 6
Hepatocyte 4
Hepatocyte 3
Hepatocyte 5
Hepatocyte 1
Hepatocyte 2
Stellate cells
Cholangiocytes
Erythroid cells
Antibody secreting B cells
Mature B cells
Central venous LSECs
Periportal LSECs
Portal endothelial cells
Alpha beta T cells
NK-like cells
Gamma delta T cells 1
Gamma delta T cells 2
Inflammatory Macrophages
Non-inflammatory Macrophages
G.
Total Cells = 4,077
T and NK Cells
B and Plasma Cells
MNPs
Other Myeloid
Dividing
Endothelial
UMAP_2
UMAP_1

Figure S4.5 (A) UMAP reproducing the 20 clusters of cells, including inflammatory macrophages (IM, cluster 4) and non-inflammatory macrophages (NM, cluster 10), from previously published scRNA-seq of non-diseased livers[381]. (B) Expression of lineage-specific genes verifies the identify of immune cells clusters (blue = T/NK cells; red = B cells; green = plasma cells; purple = dividing cells; orange = MNP). The UMAP is recolored by cell type and the proportion of immune cells is shown. (C) UMAP of scRNA-seq data from a pediatric non-cholestatic liver (NC) shows 10 clusters of cells. (D) Feature plot demonstrates expression of lineage-specific genes by cell cluster (blue = T/NK cells; red = B cells; green = plasma cells; orange = MNP; pink-other myeloid cells; purple = dividing cells; gray = endothelial cells). (E) Comparison of gene expression across all myeloid cell clusters identifies cluster 3 as neutrophils expressing FCGR3B and S100P, CD1c+ dendritic cells as cluster 7, and cluster 1 and 4 as macrophage clusters. (F) Single-R analysis using previously published data from adult normal livers as the reference [381] supports our cluster assignments with the addition of neutrophil and dividing cell clusters. (G) Re-colored UMAP by cell type and proportion of immune cells demonstrates high numbers of MNP and T/NK cells with contribution of endothelial cells from possible contamination.

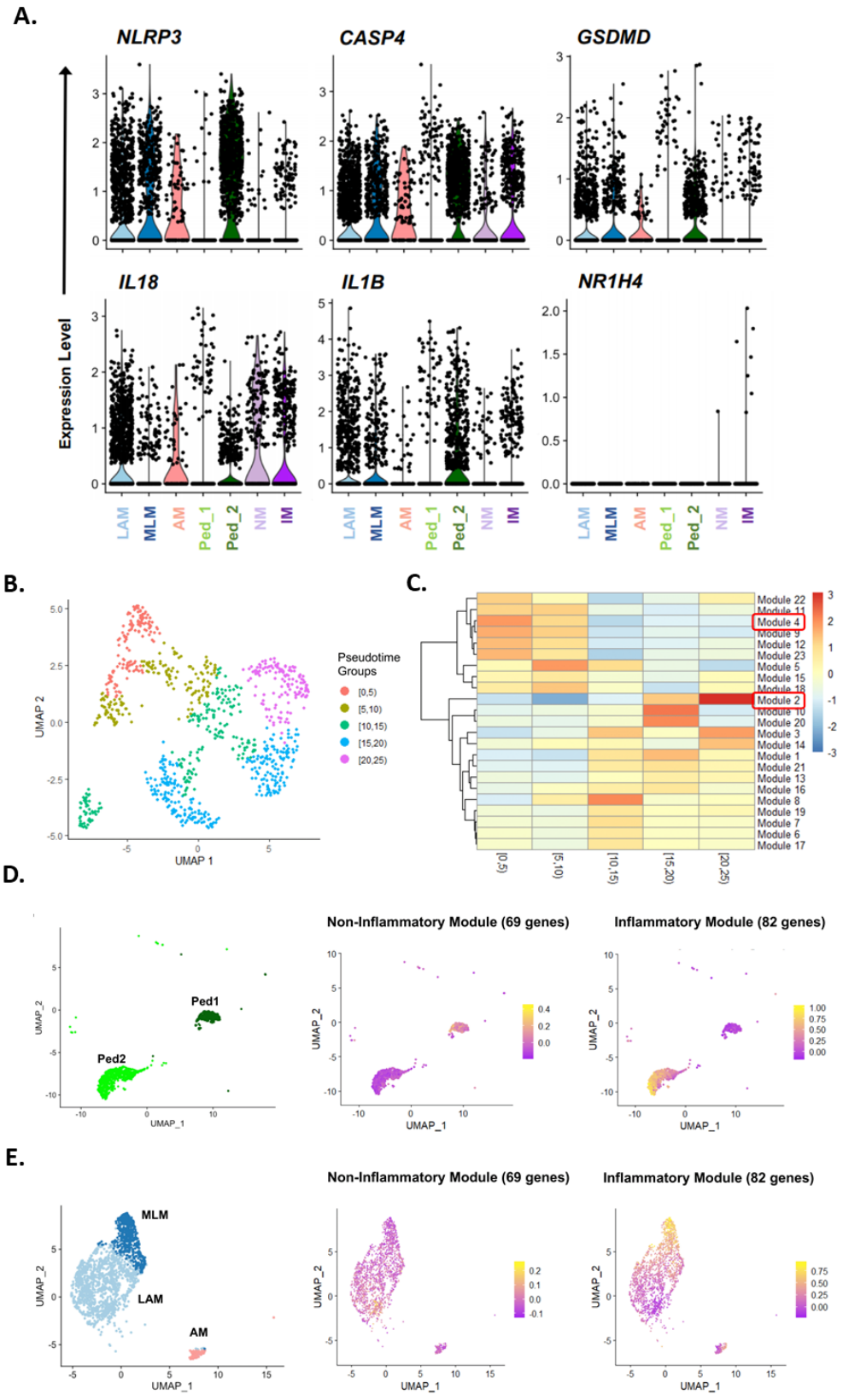

A.
NLRP3
CASP4
GSDMD
IL18
IL1B
NR1H4
Expression Level
LAM
MLM
AM
Ped_1
Ped_2
NM
IM
B.
UMAP 1
UMAP 2
Pseudotime Groups
[0,5)
[5,10)
[10,15)
[15,20)
[20,25)
C.
Module 22
Module 4
Module 12
Module 23
Module 5
Module 15
Module 2
Module 20
Module 3
Module 14
Module 1
Module 21
Module 13
Module 16
Module 8
Module 19
Module 7
Module 6
Module 17
D.
Ped1
Ped2
Non-Inflammatory Module (69 genes)
Inflammatory Module (82 genes)
UMAP_1
UMAP_2
E.
MLM
LAM
AM
Non-Inflammatory Module (69 genes)
Inflammatory Module (82 genes)
UMAP_1
UMAP_2

Figure S4.6 (A) Violin plots showing expression of individual genes involved in inflammasome activation across cholestatic and non-cholestatic macrophage subsets. (B) Macrophages from non-diseased liver (NL) were categorized into 5 groups based on their inferred pseudotime. (C) From 23 modules of genes with pseudotime-associated expression, we chose module 4 with increased expression at pseudotime 0–5 to represent the non-inflammatory module and module 2 with increased expression at pseudotime 20–25 to represent the inflammatory module. (D) Visualization of gene expression for these 2 modules in non-cholestatic pediatric liver macrophages shows that module 4 is upregulated in Ped1 macrophages similar to non-inflammatory adult macrophages and module 2 is upregulated in Ped2 macrophages similar to inflammatory adult macrophages (IM). (E) Comparison to cholestatic macrophages demonstrated low expression of the non-inflammatory module across all subsets whereas MLM demonstrated high expression of the inflammatory module.

4.4 Discussion

We are the first to perform scRNA-seq on pediatric cholestatic liver to define the macrophage transcriptional profile in obstructive cholestasis. Hepatic macrophages play a critical role in maintaining immune tolerance in the setting of persistent exposure to bacterial antigens from the intestine. Loss of this tolerogenic phenotype in the setting of inflammation may be of particular importance in ongoing hepatic injury in obstructive cholestasis. Here, we identify three populations of pathogenic macrophages independent of underlying etiology that may contribute to liver injury in obstructive cholestasis. No cholestatic macrophage subset was characterized by expression of immune regulatory genes as seen in normal adult NM and a subset of macrophages in non-cholestatic pediatric liver (Ped1). Our data thereby suggest that tissue resident macrophages,

such as NM previously reported [381], may be absent or transcriptionally altered by the tissue microenvironment in cholestatic liver injury. Instead, all 3 cholestatic macrophage subsets were most similar to IM, supporting a role for monocyte-derived macrophages or inflammatory polarization of tissue resident macrophages in disease. The population of monocyte-like macrophages had the greatest upregulation of genes encoding the S100 proteins in addition to TREM1, known to amplify the innate immune response [389], suggesting this population may have recently infiltrated [390]. We also defined a subset of lipid-associated macrophages that had the highest expression of genes involved in TLR signaling (e.g. GPNMB [391], MT1G and MT1X [392, 393]). Lastly, we demonstrate the presence of a novel adaptive macrophage subset with increased RORA gene expression, which has been shown to promote anti-inflammatory polarization of hepatic macrophages in a murine model of nonalcoholic steatohepatitis [394] and a human monocyte cell line [395]. The transcriptional profiling of these distinct subsets may identify macrophage-specific targets to ultimately inhibit monocyte recruitment, block TLR-mediated macrophage activation, or re-program macrophages to an anti-inflammatory phenotype.

While macrophages have been implicated in immune-mediated hepatic injury from cholestasis [370-374, 376-380], the exact mechanism is not well known. Current medical therapies for cholestatic liver disease include FXR agonists, which in addition to regulating the bile acid pool may also inhibit macrophage inflammasome activation based on prior studies [374]. However, we demonstrate an absence of NR1H4 encoding FXR in cholestatic macrophages despite evidence of TLR signaling and inflammasome activation. As macrophages play a role in cholestatic liver injury, this finding highlights the lack of current cell-specific immune-modulatory strategies and the need for a deeper understanding of the immune response to cholestasis.

LAM in our samples were characterized by C1Q expression, however, non-cholestatic macrophages, in particular NM, also exhibited C1Q staining on immunofluorescence suggesting LAM may arise from inflammatory polarization of tissue-resident macrophages. In contrast to normal macrophages, LAM had a gene signature that was more similar to recently published data on TREM2+ lipid-associated macrophages in murine adipose tissue [385]. This gene signature was also similar to human hepatic macrophages during obesity and may represent a conserved response to loss of metabolic homeostasis [385]. As hypercholesterolemia is a common sequela of cholestasis, the TREM2+ LAM in our samples may arise from C1Q+ normal macrophages in response to similar metabolic derangements. However, LAM in our study differed in that they had an overall inflammatory gene signature despite expression of TREM2 previously shown to promote anti-inflammatory macrophage polarization [385, 396]. Targeting the TREM2 molecular pathway may be an important therapeutic target to re-program hepatic macrophages to an immune regulatory phenotype and reduce the consequences of hypercholesterolemia in cholestasis.

The mechanism of disease pathogenesis in BA is hypothesized to be multifactorial, including an aberrant immune response to a cognate antigen [397] whereas ALGS is a genetic disease resulting in bile duct paucity. Thus, while the aim of the current study was to identify a common cholestatic macrophage phenotype, there are likely etiology-specific differences in the immune pathways for macrophage activation that require further investigation in larger studies. However, despite this limitation, we provide important insight into hepatic macrophage heterogeneity in cholestatic liver disease compared to healthy livers. Despite age differences, it is worth noting that macrophages from a non-cholestatic pediatric patient demonstrated a similar dichotomy as adult

hepatic macrophages. This finding suggests that the distinct transcriptional signature of cholestatic macrophages is not a result of age-specific differences or technical differences between datasets such as variation in sample isolation, processing and digestion protocols, or experimental design. Lastly, we acknowledge that our findings may not be limited to obstructive cholestasis and may overlap with other causes of end-stage liver disease characterized by cirrhosis and portal hypertension. A recent study on adult cirrhotic livers described a population of scar-associated macrophages in cirrhosis [398] that appear most similar to cholestatic LAM and express higher levels of TREM2, CD9, LGALS3, and SPP1. Future studies will more clearly define the similarities and differences in subset-specific macrophage function by patient age, stage of cholestatic liver disease, and etiology of cirrhosis.

In this study, we have used single-cell transcriptional analysis of pediatric cholestatic liver samples to identify macrophage subsets at greater resolution than previously described characterization by ontogeny or M1 versus M2 polarization [292]. With ongoing work, we will strengthen conclusions on the hepatic macrophage transcriptional signature in different cholestatic liver diseases and identify common therapeutic targets to reprogram macrophages and slow disease progression. More specifically, we highlight expression of the immune regulatory genes RORA and TREM2 within these inflammatory subsets that may be potential therapeutic targets to ameliorate inflammatory injury in obstructive cholestasis. Future work to correlate our findings to the immune cell subsets present earlier in disease will provide important insight into cell-specific therapeutic strategies to improve prognosis shortly after disease onset. Identifying molecular targets to reprogram hepatic macrophages in cholestasis may also have therapeutic implications for other etiologies of liver diseases and reduce the medical burden of end-stage liver disease.

CHAPTER 5

MAGNET: A Web-based Application for Gene Set Enrichment Analysis Using Macrophage Data Sets

Chapter 5 is adapted from an original research article entitled “MAGNET: A Web-based Application for Gene Set Enrichment Analysis Using Macrophage Data Sets” (Chen et al., In revisions).

5.1 Introduction

Analysis of next-generation sequencing (NGS) experiments, such as RNA-seq, often produces long lists of genes as output, such as those differentially expressed between two or more conditions. It is therefore a logical and critical next step to identify the biological relevance associated with these genes. Typically, this is achieved through functional enrichment analyses that utilize standardized biological knowledge repositories, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes Pathways (KEGG) [148] and Molecular Signatures Database (MsigDB) [152]. These repositories annotate sets of genes with defined biological terms. These biological terms are then associated with input gene lists by calculating the overlap and performing statistical enrichment tests to assess significance. A large number of computational tools have been developed for performing this type of enrichment analysis by querying these repositories. Some of the most popular tools includes GOrilla [149], DAVID [150, 399], IPA [400], and BiNGO [401]. Although these applications provide an effective way to characterize user-supplied gene lists and are now considered an essential step in typical bioinformatic workflows, they are often limited to generic results that do offer new insights to domain-specific questions. There are a number of reasons for this limitation including: the attempt to provide terms that encompass all of biology; the static nature of the source repositories that do not account for the latest research; and the broad nature of the annotation scheme.

These issues are exemplified when endeavoring to perform gene set enrichment analysis on the results of genomic experiments on macrophages. Macrophages are highly plastic immune cells that are found in virtually every tissue in health as well as having an essential role in in innate

immune response [8, 29]. They exhibit highly specialized functions, depending on their tissue of residence and exhibit divergent responses to environmental and pathogenic stimuli [47, 54]. They have been implicated in numerous pathological models and are being investigated as potential therapeutic targets in various diseases [402]. For this reason, their genomic landscape has been the subject of many studies across multiple biomedical disciplines [403]. However, in our experience, typical gene set enrichment analysis tools using standardized repositories will primarily return generic terms related to the role of macrophages in immune response and inflammation, regardless of the context of the original experiment. Alternatively, when a set of ubiquitously expressed genes is included as background to account for the macrophage transcriptome, the tools may return no significant results at all. This is because the standardized repositories do not include annotation terms to describe the novel and specialized functions of macrophages. Instead, many canonical macrophage genes are associated with the role of macrophages in innate immunity despite their relevance to other biological processes and gene pathways. The plasticity of macrophages exacerbates this limitation, but the same issue arises across domains in studies that investigate the condition-specific function of particular cell types. Thus, there is a great demand for an application that can provide precise, relevant results in accordance with the latest research.

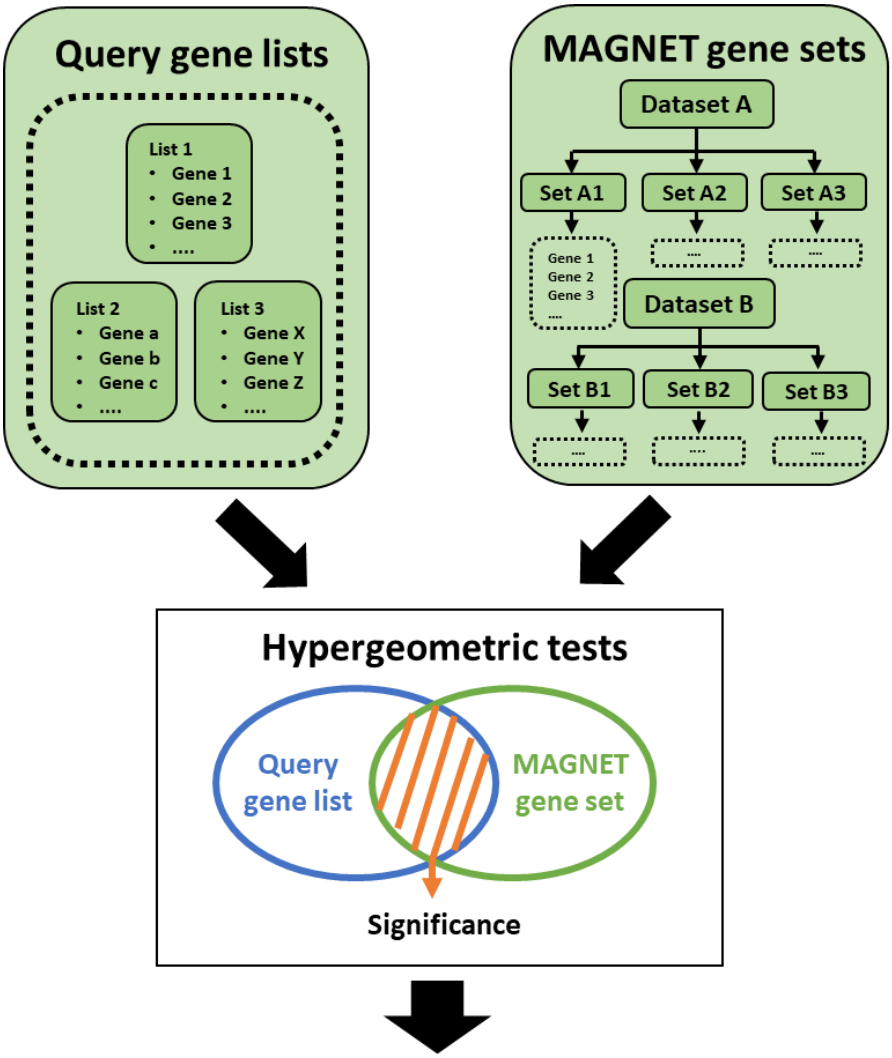
These challenges inspired us to develop MAGNET (Macrophage Annotation of Gene Network Enrichment Tool), an interactive web application based in Python, by utilizing the curated annotations from prior macrophage studies instead of terms from standardized biological knowledge repositories. These annotations can be curated from published manuscripts and data repositories, such as Gene Expression Omnibus (GEO) [404], that describe gene expression patterns defined by comparing macrophages across experimental conditions, such as different

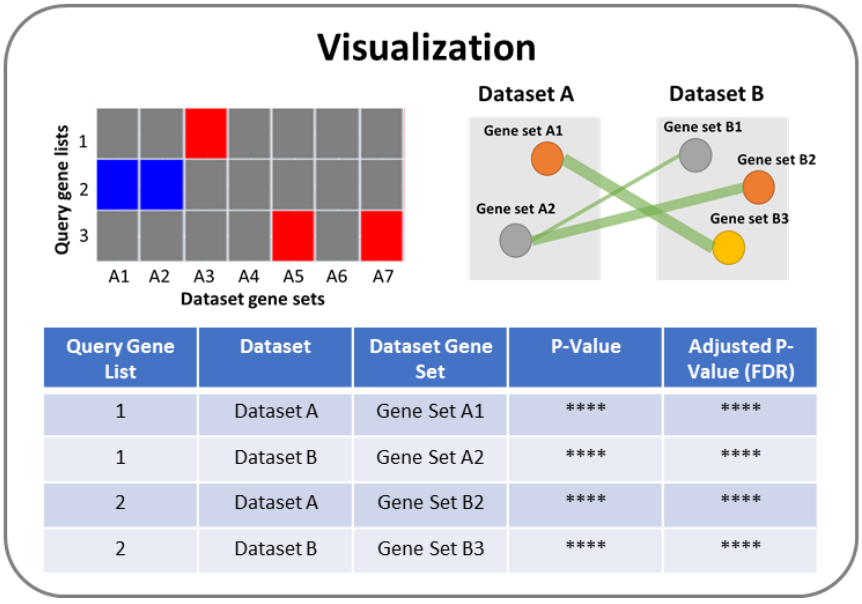
tissues, disease status, or time. Although a wealth of information has been published on macrophage function and identity using genomic assays, it is not always easy for a researcher outside the original study to utilize these results. MAGNET overcomes the obstacles associated with retrieving the data and performing bioinformatic analysis by providing the user with a user-friendly graphical interface to compare their data with multiple other studies in parallel. By outputting the results of gene set enrichment analyses against these updated and domain-specific annotations, we show that MAGNET provides a relevant and unique characterization of user-supplied gene lists in an accessible and flexible manner.

5.2 Methods

Overview of MAGNET

We developed MAGNET, a web-based interactive application for assessing and visualizing enrichments of user-supplied lists of genes against annotations curated from published literature on murine macrophage gene expression. The application is implemented using Python/Django framework with PostgreSQL database integration. It is publicly available at https://magnet-winterlab.herokuapp.com. The overview schematic of MAGNET is shown in Figure 5.1. Th The user must input at least one gene list to query as well as a second gene list specifying the background set. Typically, we would expect the lists to originate from an RNA-seq or similar experiment, but it is not necessary. The overlap of the query gene list with MAGNET's annotations is then compared against the overlap with the background gene set using the hypergeometric distribution. The significant annotations are visualized in multiple formats – as a heatmap of enrichment/depletion, table of significantly enriched results, and network of intersecting annotations – to facilitate user interpretation of results.





Query Gene List	Dataset	Dataset Gene Set	P-Value	Adjusted P-Value (FDR)
1	Dataset A	Gene Set A1	****	****
1	Dataset B	Gene Set A2	****	****
2	Dataset A	Gene Set B2	****	****
2	Dataset B	Gene Set B3	****	****

Figure 5.1 Overview schematic of MAGNET workflow. The user-supplied gene list(s) are compared with annotated gene sets that have been curated from published macrophage datasets. The significance of the overlap is assessed using the hypergeometric distribution to identify enriched (and depleted) MAGNET annotations. The results are visualized in heatmap, table, and network formats.

Table 5 Datasets currently included in the MAGNET database.

Title	Authors	Year	Journal	Summary	Citation
Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment	Lavin et al.	2014	Cell	**11 gene sets:** bulk RNA-seq on tissue resident macrophages from 7 tissues	[47]
Gene expression profiles and transcriptional regulatory pathways underlying mouse tissue identity and diversity	Gautier et al.	2012	Nature Immunology	**4 gene sets:** microarray on tissue-resident macrophages from 4 tissue/organs	[51]
Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors	Paul et al.	2015	Cell	**19 gene sets:** single-cell RNA-seq on myeloid progenitors in bone marrow	[405]
Microglia development follows a stepwise program to regulate brain homeostasis	Matcovitch-Natan et al.	2016	Science	**7 gene sets:** bulk RNA-seq on developmental stages of microglia	[157]
Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span	Misharin et al.	2017	Journal of Experimental Medicine	**5 gene sets:** bulk RNA-seq on monocytes and macrophages in the lung during pulmonary fibrosis	[80]
Microbiome Influences Prenatal and Adult Microglia in a Sex-	Thion et al.	2018	Cell	**7 gene sets:** microarray on	[158]

Specific Manner				developmental stages of microglia	
Locally renewing resident synovial macrophages provide a protective barrier for the joint	Culemann et al.	2019	Nature	**7 gene sets:** single-cell RNA-seq on synovial macrophages in the arthritic joint	[220]

5.2.3 User Interfaces

MAGNET implements two modes of input for the query gene list input (Figure 5.1A). In the traditional "single" mode, the user submits a single column of genes via the input box or file upload. This mode operates in an analogous manner to typical gene set enrichment analysis with MAGNET calculating the enrichment of this gene list against all database annotations. In contrast, MAGNET is the first application, of which we are aware, to offer the option of multiple parallel queries. In 'multiple' mode, the user can analyze multiple gene lists simultaneously by uploading a comma-separated file that consists of two columns: the first contains the genes while the second assigns them to different lists. The purpose of this feature is to enable the user to perform parallel enrichment analyses across related gene lists without requiring multiple queries and visualize the results as a single output. For example, the multiple query mode is particularly useful for analyzing a gene list that has been clustered and provides the enrichment for each cluster independently. Regardless of the query mode, MAGNET also requires the user to input a background gene list either through the text box or a file upload. This background is part of the hypergeometric calculation described below. Finally, the user may select which curated datasets from the MAGNET database against which to perform enrichment analysis. In addition to the datasets currently included, users also have the ability to upload custom datasets against which to test their

query. This original feature enables users to perform meta-analysis between any two datasets, greatly increasing the flexibility of the application.

MAGNET also includes two further interfaces for exploring datasets and genes in the database. Each curated dataset is linked to a documentation page listing essential information on the source, including the full citation information, abstract, link to PubMed, and details on the annotated gene sets (Figure 5.2B). For genes, we have implemented an interface that enables the user to query the MAGNET annotations for individual genes (Figure 5.2C). One or more genes can be entered as input and a table of the associated gene sets across all datasets is returned as output. Through this interface, the user can explore the MAGNET database without performing a full enrichment analysis.

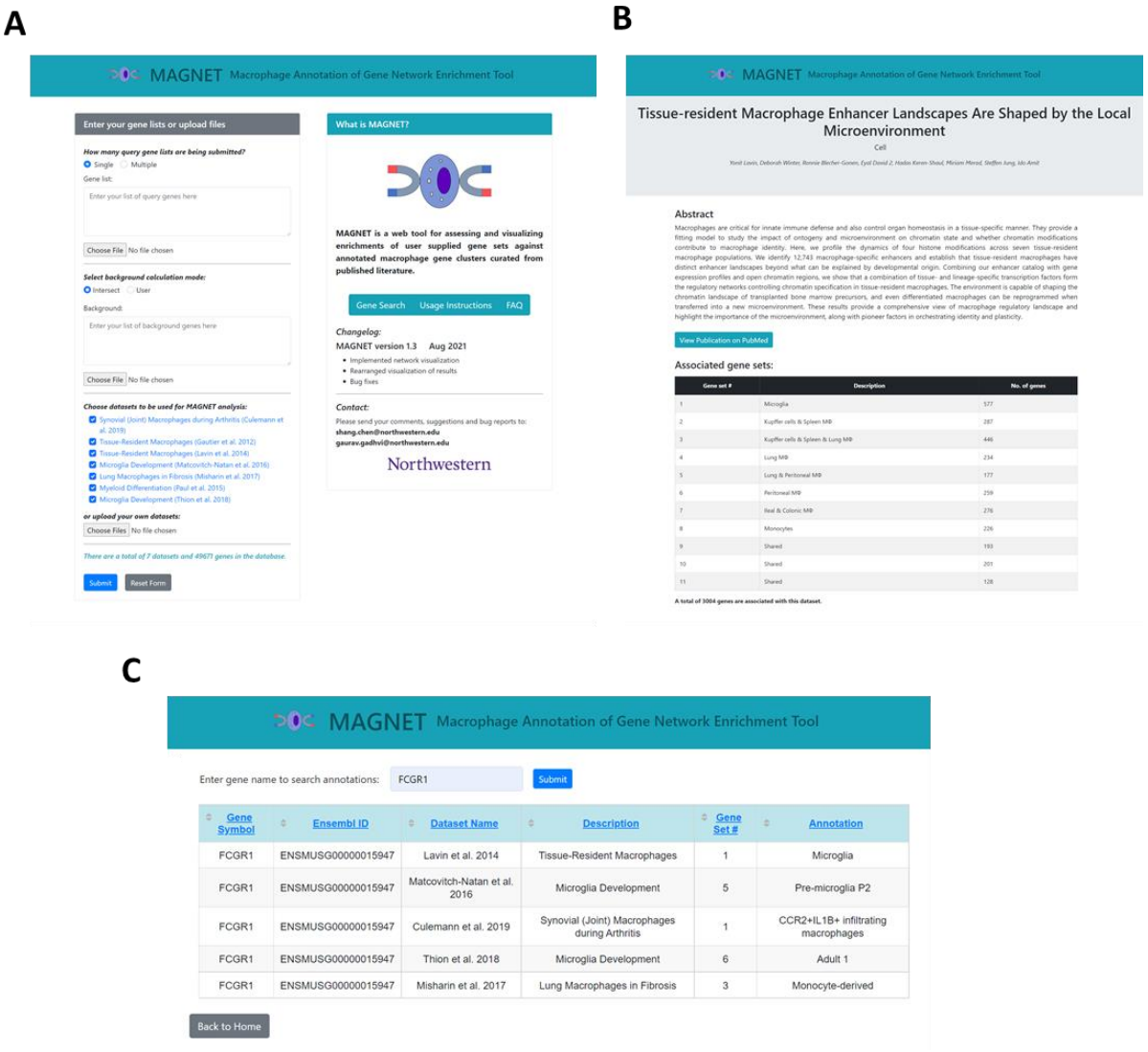


Figure 5.2 MAGNET user interfaces. A. Gene Set Enrichment Interface: The user is required to upload query and background gene lists, select background calculation mode, and choose the datasets to be included in the analysis. There is also the option for the user to upload one or more custom datasets. **B**. Dataset Interface: This interface provides information on the source of each dataset. Lavin et al., 2014 is shown as an example. **C**. Individual Gene Interface: The user can enter genes to query against the MAGNET database of annotations. FCGR1, the gene encoding

the macrophage surface marker CD64, is shown as an example.

Hypergeometric Enrichment Analysis

Like other common gene set enrichment analysis tools, such as GOrilla (4) and DAVID (5, 6), MAGNET utilizes the hypergeometric distribution to identify annotation terms that are significantly enriched among the query genes. MAGNET calculates the number of overlapping genes (k) independently between the user-supplied gene list (s) and each annotation. Then, the significance of that value is calculated using the hypergeometric distribution which gives the probability of observing an overlap of k or more genes by chance given the size of the query and annotation:

$N = number\ of\ genes\ in\ the\ background\ list$

$n = number\ of\ genes\ in\ the\ query\ list$

$K = number\ of\ genes\ associated\ found\ in\ the\ MAGNET\ annotated\ gene\ set$

$k = number\ of\ genes\ in\ the\ overlap\ between\ the\ query\ list\ and$
$MAGNET\ annotated\ gene\ set$

$$P(x \geq k) = \ hg(k; N, K, n) = \sum_{i=k}^{\min(n,K)} \frac{(C_i^n)(C_{K-i}^{N-n})}{C_K^N} \ where\ C_b^a = \frac{a!}{b!\,(a-b)!}$$

The resulting probability is reported as the p-value of enrichment. Likewise, the complementary probability (1-p), reflects the significance of depletion. Since each calculation of overlap

represents a different hypothesis, MAGNET also outputs an adjusted p-value to account for multiple comparisons using the Benjamini-Hochberg method for False Discovery Rate (FDR) [406]. The total number of comparisons is the number of query gene lists multiplied by the number of annotations across all datasets.

MAGNET offers two settings for determining the number of genes in the background which affects the value of the parameters N and n (Figure 5.3):

1. *INTERSECT*: The background N is given as the intersection between the user-supplied background and the total genes in the dataset containing the annotation in question. Consequently, n is given as the intersection between user-supplied query list and the total dataset. This is the default mode of calculation.

2. *USER*: N and n are simply defined as the user-supplied background and query lists, respectively. This option may be preferable when there is limited overlap between the user-supplied background and the MAGNET dataset.

The hypergeometric tests are performed sequentially on each pair of query gene list(s) and annotated gene sets from curated datasets selected or uploaded by the user.

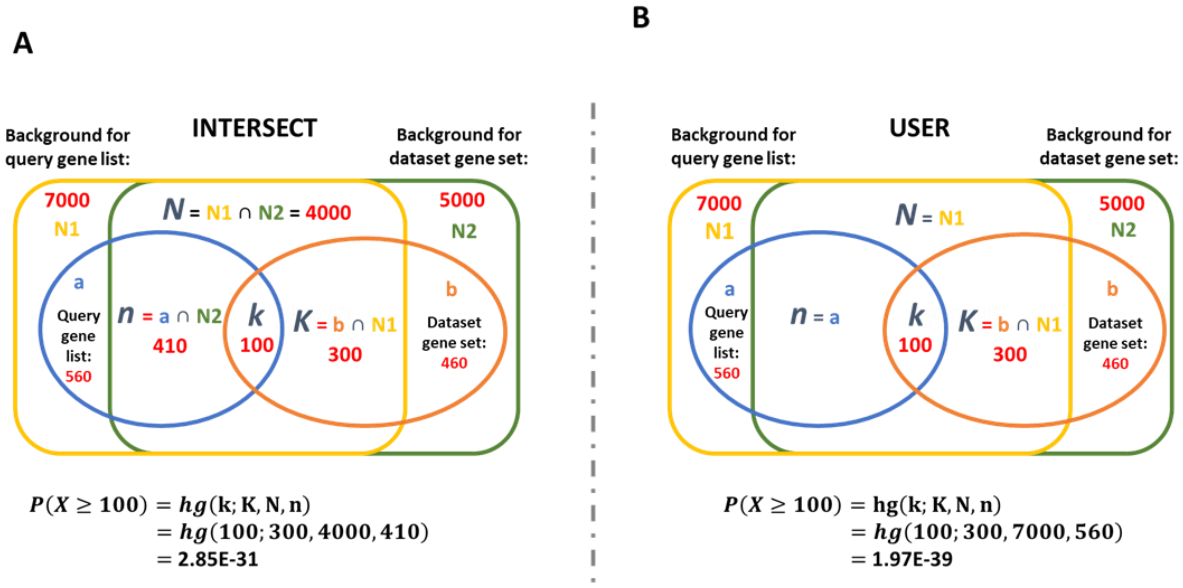


Figure 5.3 Comparison of the two settings for background in the hypergeometric test performed by MAGNET. In this example, there are 560 and 460 genes in a query gene list and annotated gene set, respectively, with an overlap of 100 between the two. There are 7000 genes in the user-supplied background with 4000 overlapping the 5000 genes from the dataset. The overlap calculations and hypergeometric p-values are illustrated for the two background settings: **A**. *INTERSECT* and **B**. *USER*.

Visualization of results

MAGNET generates three types of outputs to facilitate easy interpretation and visualization of results for the user: HEATMAP, TABLE, and NETWORK:

HEATMAP: The first output consists of a heatmap corresponding to each dataset selected for analysis, which are color-coded red or blue to reflect significant enrichments and depletions,

respectively (Fig 5.4A). Within each heatmap, the row(s) represent the query gene list(s) while the columns represent annotated gene sets from MAGNET or custom datasets. A slider widget is provided to enable the user to select their desired p-value cutoffs and update the heatmaps interactively. A mouse-over function allows the user to see additional information on the overlap, including description of the gene set, raw and adjusted p-values, and parameters used for hypergeometric test. Utilizing heatmaps to visualize significance allows the user to easily assess all the comparisons in a data set simultaneously. Each heatmap may be downloaded separately as a png file.

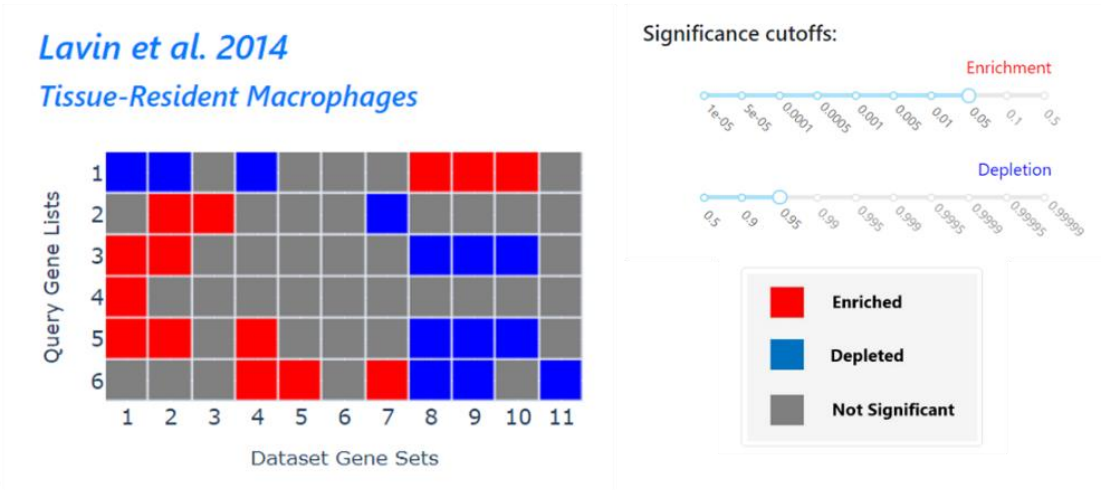
TABLE: The second output is a table listing all the significantly enriched gene sets (Fig 5.4B). The table includes sortable columns for the name of the source dataset, the name of the annotation, the p-value, the FDR adjusted p-value, and the parameters used in the hypergeometric calculation. The final column reports the full list of genes in the overlap. The contents in the table are updated dynamically depending on the p-value cutoff chosen. The table can be downloaded as a csv file.

NETWORK: The third output is a graph depicting the network of shared genes between different MAGNET annotations with the enrichment results superimposed (Figure 5.4C). The widths of the edges correspond to the proportion of genes shared between two annotated gene sets, whereas the color of the nodes represent the significance by p-value. The underlying graph remains constant and can be visualized to understand the overlap between datasets (**Fig S5.1**). In the multiple query mode, the user must specify which gene list to visualize. The network visualization is generated through an interactive Cytoscape [407] plugin that enables the user to set the position of different elements. This novel method of visualizing these results allows the user to appreciate how the

different annotations are related across multiple datasets.

Any user-supplied query, background, or custom dataset genes that are not found in the MAGNET database are reported at the bottom of the output page to allow the user to catch inconsistencies in the nomenclature.

A



B

Query Gene List (n)	Dataset	Description	Dataset Gene Set (K)	P-value	Adjusted P-value (FDR)	Parameters (N, K, n, k)	Overlapping Genes
filter data...							
1	Culemann et al. 2019	Synovial (Joint) Macrophages during Arthritis	CCR2+ARG1+ infiltrating macrophages (5)	0.00018	0.0014	1673, 252, 853, 154	Cstb, Txn1...
1	Culemann et al. 2019	Synovial (Joint) Macrophages during Arthritis	STMN1+ proliferating cells (7)	3e-95	3.5e-93	1673, 611, 853, 508	Birc5, Ube...
1	Gautier et al. 2012	Tissue-Resident Macrophages	Peritoneal (3)	0.038	0.15	119, 14, 18, 4	Acaca, Lrg...
1	Lavin et al. 2014	Tissue-Resident Macrophages	Monocytes (8)	4.4e-19	1.4e-17	1703, 161, 509, 99	Tubb5, G6...
1	Lavin et al. 2014	Tissue-Resident Macrophages	Multiple (9)	4.5e-12	8.5e-11	1703, 120, 509, 70	Rpl32, Rpl...
1	Lavin et al. 2014	Tissue-Resident Macrophages	Multiple (10)	3e-7	0.0000031	1703, 142, 509, 69	Rps8, Uqc...
1	Matcovitch-Natan et al. 2016	Microglia Development	Early microglia E1 (2)	1.7e-40	1e-38	1916, 580, 937, 415	Ube2c, E...
1	Matcovitch-Natan et al. 2016	Microglia Development	Early microglia E2 (3)	6e-49	5.4e-47	1916, 376, 937, 307	Cenpa, Cd...
1	Misharin et al. 2017	Lung Macrophages in Fibrosis	Mo-AM & TR-AM (1)	8.1e-17	2.1e-15	6822, 2183, 2259, 873	Cdc45, Co...
1	Misharin et al. 2017	Lung Macrophages in Fibrosis	IM & Mo-AM (3)	0.0000067	0.000063	6822, 677, 2259, 275	Gna12, Sc...

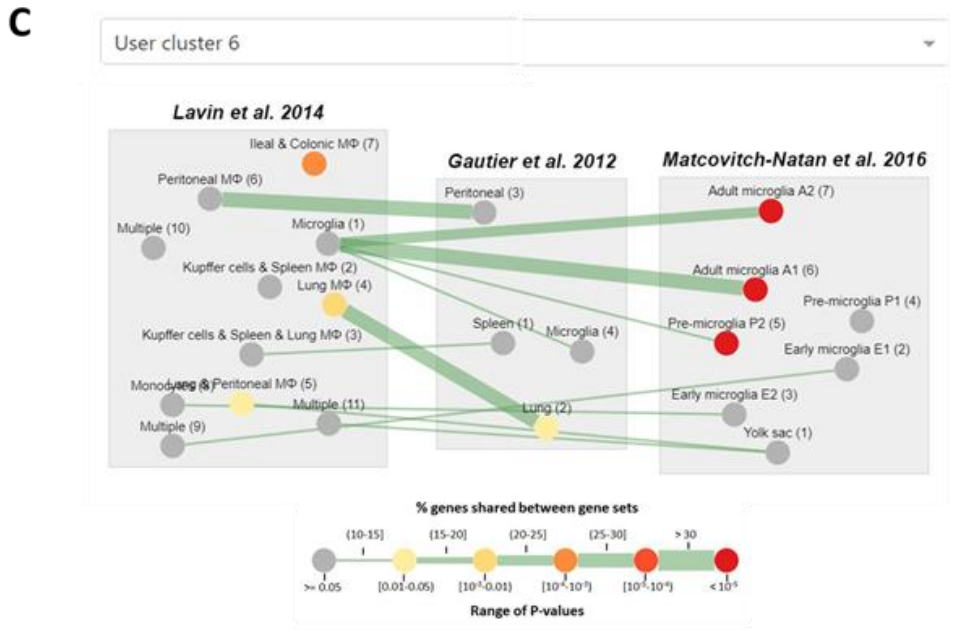


Figure 5.4 Example output of MAGNET using gene clusters reported in Koch et al., 2018 as input. A. Heatmap output of results against curated dataset Lavin et al. 2014 with p-value cutoffs of 0.05 and 0.95 for enrichment and depletion, respectively. **B.** Table output of significantly enriched annotations. Only the first 10 rows of the table are shown. **C**. Network output for visualizing Cluster 1 results with the three datasets from **A.**

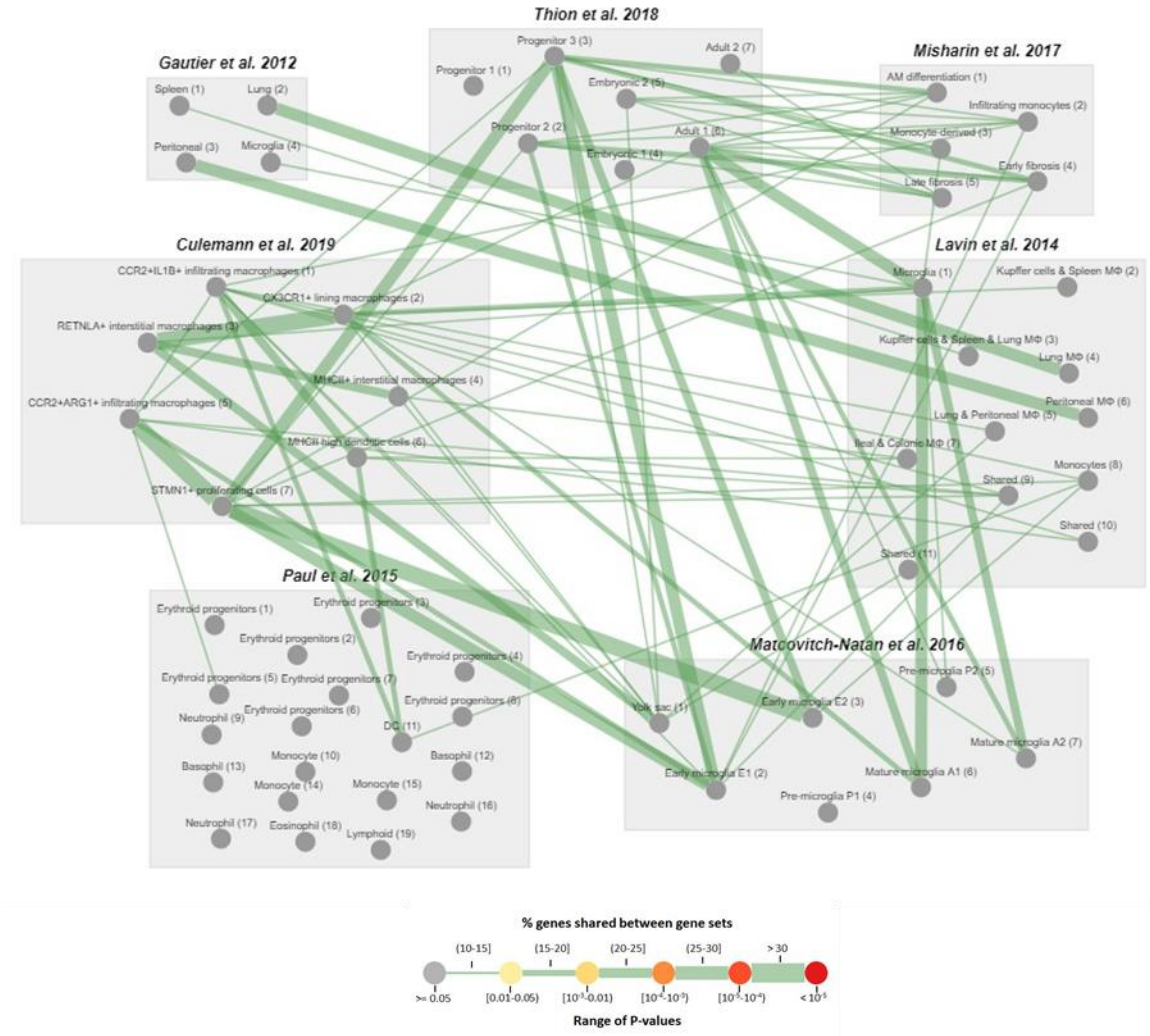


Figure S5.1 Network visualization of all gene sets curated in MAGNET.

5.3 Results

To assess the performance of MAGNET on a real-world example, we used an independent RNA-seq dataset that compared gene expression of microglia, brain-resident macrophages, between mice bred with no microbiota (germ-free) and control adult mice [157]. The experiment was designed to assess the effect of the microbiome on microglia development in the brain. The original publication reported a total of 322 microbiome-dependent genes with decreased

expression in germ-free mice. As reported in the original study, the top hits from GO analysis of these genes largely comprises generic biological terms associated with inflammation and overlapping terms reflecting the same small set of genes (Figure 5.5A). Using the 7764 expressed genes documented in the publication as background, we analyzed the 322 genes as a single query with INTERSECT setting in MAGNET (Table S5.2). Then, we summarized the results as a bar graph of enrichment (given by the ratio of overlapping genes in each gene set) to compare annotations within a dataset. MAGNET reproduced the association of these genes with mature microglia genes as reported in the original publication (*Matcovitch-Natan et al. 2016 - Mature microglia A2 (7) ,* p = 0.0292), suggesting that microglia maturity is dependent on an intact microbiome (Figure 5.5B) [157]. However, we also observed that yolk sac genes are significantly enriched among the microbiome-dependent gene list (*Matcovitch-Natan et al. 2016 - Yolk sac (1),* p = 0.0001): this novel result may indicate another means by which microglia development is perturbed in germ free mice. We next compared these results with MAGNET annotations stemming from a second more recent study of microglia development (*Thion et al 2018*) and find that the *Progenitor 1 (1)* (p = 0.0200) and *Embryonic* 2 (5) (p = 0.0007) genes sets are enriched rather than either adult gene set (Figure 5.5B) [158]. While these results appear contradictory to the prior publication, there is room for interpretation as the gene sets do not entirely line up between datasets (Figure S5.2A). In addition, we ran MAGNET with the INTERSECT setting to query the enrichment specifically across microglia development genes but we observed similar results when MAGNET is run with the wider gene context in USER setting (Fig S5.2B)**.** Aside from small technical differences, the likely explanation for the discrepancy between data sets is the difference in the sex of the mice used in each study: Matcovich-Natan et al. used only female mice while Thion et al. utilized both males and females.

Thus, the user may infer that sex affects the role of the microbiome in microglia development – which is indeed supported by data in the latter study demonstrating that loss of microbiota disproportionately perturbs the microglial phenotype of female mice – and can plan follow-up experiments to test this hypothesis. By simplifying comparison of user-supplied gene lists with published datasets, MAGNET enables further insight into macrophage biology.

A

GO term ID	Description	# in Overlap	# in Gene Set	P-value
GO:0002476	antigen processing and presentation of endogenous peptide antigen via MHC class Ib	6	9	3.57E-7
GO:0002484	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway	6	9	3.57E-7
GO:0002428	antigen processing and presentation of peptide antigen via MHC class Ib	6	10	8.6E-7
GO:0006955	immune response	34	353	2.9E-6
GO:0009617	response to bacterium	18	129	5.07E-6
GO:0002486	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent	5	8	5.86E-6
GO:0002475	antigen processing and presentation via MHC class Ib	6	13	6.33E-6
GO:0002483	antigen processing and presentation of endogenous peptide antigen	6	14	1.07E-5
GO:0019885	antigen processing and presentation of endogenous peptide antigen via MHC class I	6	14	1.07E-5
GO:0035455	response to interferon-alpha	6	16	2.66E-5

Overlapping terms Generic inflammation terms

B

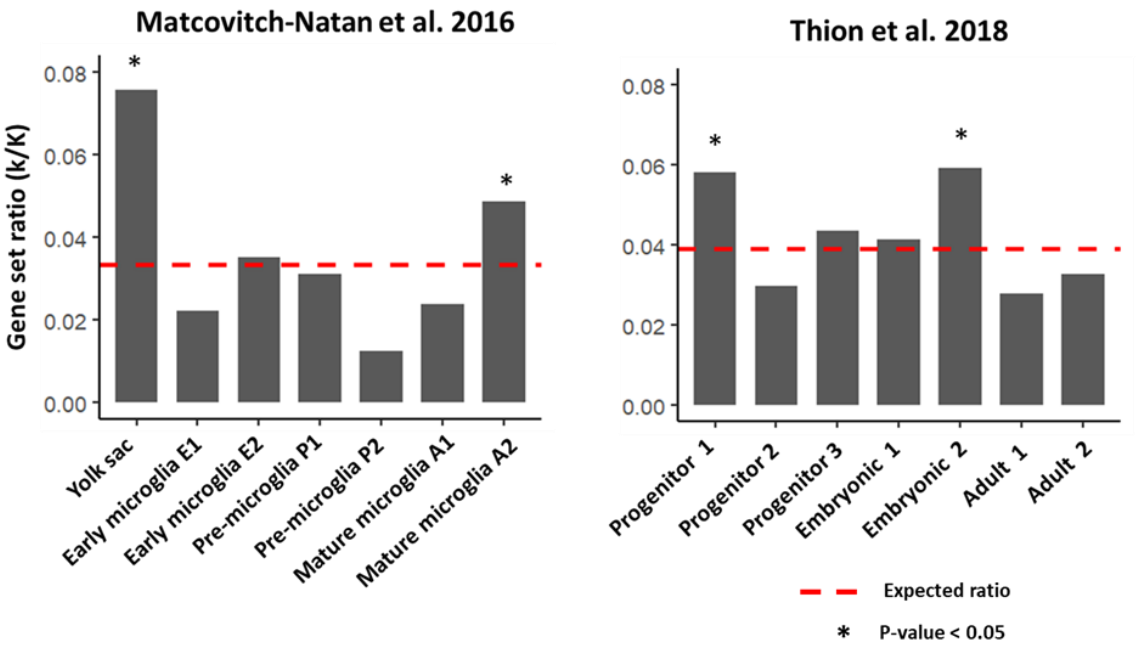
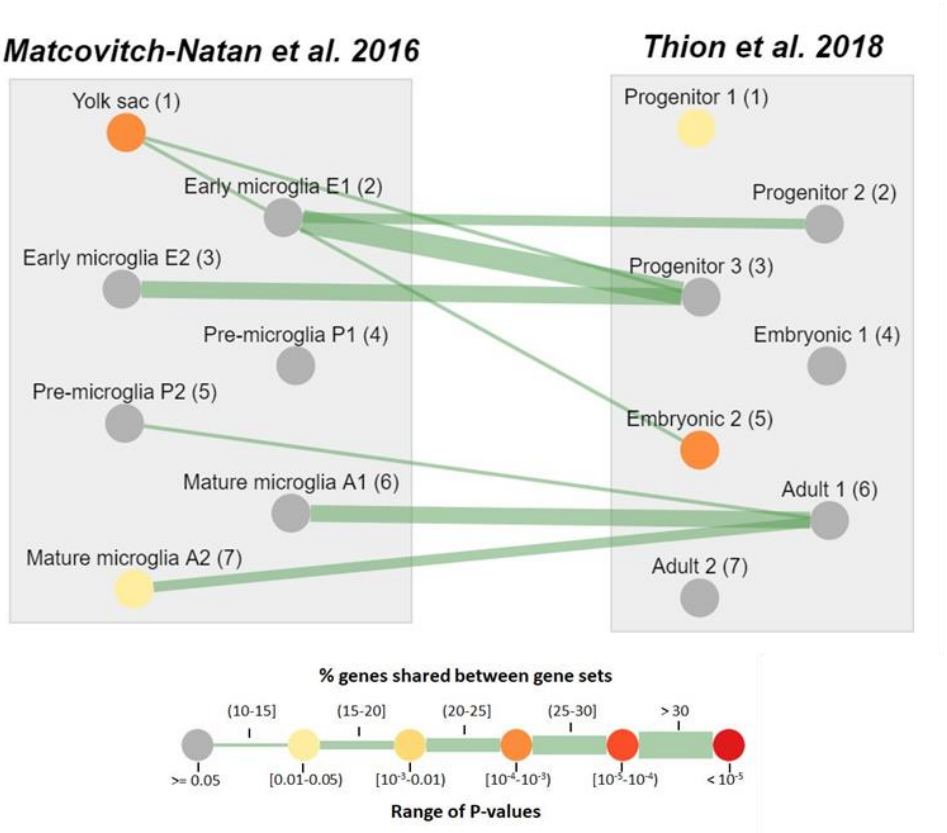


Figure 5.5 Running MAGNET with single query mode. A. Significant GO processes enriched for the 322 microbiome-related genes [157]**. B.** Ratio of overlapped microbiome-associated genes (k) to each reference gene set (K) from Matcovich-Natan et al. 2016 [157] and Thion et al. 2018 [158]. The dashed line indicates the expected distribution of genes, which is defined as the proportion of microbiome-associated genes in all expressed genes (n = 7764).

A


Matcovitch-Natan et al. 2016
Thion et al. 2018
Yolk sac (1)
Early microglia E1 (2)
Early microglia E2 (3)
Pre-microglia P1 (4)
Pre-microglia P2 (5)
Mature microglia A1 (6)
Mature microglia A2 (7)
Progenitor 1 (1)
Progenitor 2 (2)
Progenitor 3 (3)
Embryonic 1 (4)
Embryonic 2 (5)
Adult 1 (6)
Adult 2 (7)
% genes shared between gene sets
(10-15]
(15-20]
(20-25]
(25-30]
> 30
>= 0.05
[0.01-0.05)
[10-3-0.01)
[10-4-10-3)
[10-5-10-4)
< 10-5
Range of P-values

B

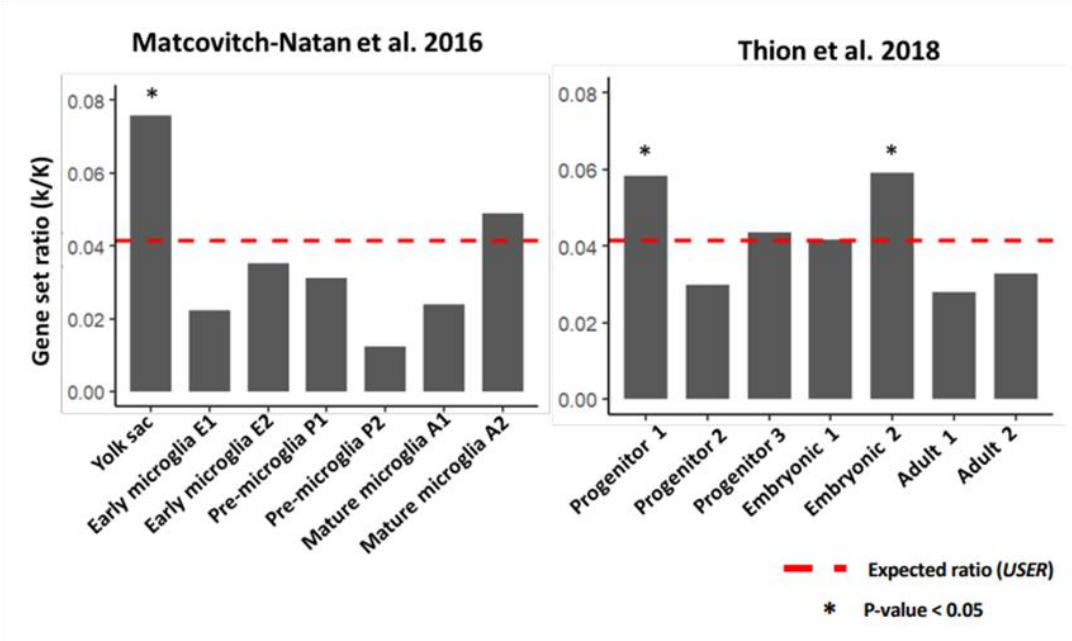

Matcovitch-Natan et al. 2016
Thion et al. 2018
Gene set ratio (k/K)
0.08
0.06
0.04
0.02
0.00
*
Yolk sac
Early microglia E1
Early microglia E2
Pre-microglia P1
Pre-microglia P2
Mature microglia A1
Mature microglia A2
Progenitor 1
Progenitor 2
Progenitor 3
Embryonic 1
Embryonic 2
Adult 1
Adult 2
Expected ratio (USER)
* P-value < 0.05

Figure S5.2 Alternate options for MAGNET with single query mode. A. Network visualization from MAGNET showing enrichment of Matcovitch-Natan et al. 2016 and Thion et al. 2018 datasets using the query list of microbiome-dependent genes with *INTERSECT* setting**. B.** The ratio of genes in each annotated gene set from the Matcovich-Natan et al. 2016 [157] and Thion et al. 2018 [158] datasets that overlap the microbiome-dependent gene list as calculated by MAGNET with *USER* setting. The dashed line indicates the expected ratio based on the total number of microbiome-dependent and dataset genes. * indicates P-value < 0.05 for the significance of enrichment.

Next, we tested the multiple query mode in MAGNET using another independent RNA-seq dataset that profiled gene expression in alveolar macrophages from a murine model of lung transplantation [139]. The experiment was designed to compare naïve alveolar macrophages with those isolated from newly implanted lungs at 2- and 24-hours post-reperfusion. The processed dataset consists of 7166 differentially expressed genes that clustered into 6 different expression patterns using k-means (Figure 5.6A). The original publication reported the top GO hit for each cluster; however, because the enrichment was calculated separately for each cluster, it is not possible to compare terms in order to build a cohesive narrative. To better understand how these processes vary across the whole dataset, we used them as custom datasets in MAGNET with these 6 clusters as input in multiple query mode and the 7166 genes as background (Figure 5.6B and Table S5.3). We find that some of these processes are, in fact, shared across clusters. Moreover, we also input the 6 clusters in the multiple query mode of MAGNET (*INTERSECT*) using the built-in datasets (Fig S5.3 and Table S5.3). We then chose select annotations across datasets to feature as a bubble plot (Fig 5.6C). For instance, the genes that are preferentially expressed at 24

hours (Cluster 1) are highly enriched for *STMN1+ proliferating cells (7)* from the *Culemann et al. 2019* dataset *($p=2.95x10^{-95}$).* This is consistent with the unique enrichment of the GO process *Cell cycle* in this cluster. Other MAGNET annotations that share genes with *Culemann 7* (Fig S5.1) and are associated with cell cycle in their respective datasets – such as *Thion et al 2018 - Progenitor 3 (3)* and *Matcovitch-Natan et al 2016 - Early microglia E1/E2 (2/3)* – were also enriched (Fig S5.3). In addition, we find that Cluster 1 genes significantly overlap *Misharin et al 2017 – AM differentiation ($p=8.15x10^{-95}$)* which suggests that there is similarity in gene expression after transplantation and as monocytes differentiate into alveolar macrophages during fibrosis. These results fit with the current paradigm of infiltrating monocytes replacing tissue-resident macrophages following a disruption to the niche [55]. The hypothesis that the AM compartment in this dataset comprises monocyte-derived cells starting at 2 hours post-reperfusion is further supported by the enrichment of monocyte-related annotations in Cluster 3 ((*Culemann et al. 2019 - CCR2+IL1B+ infiltrating macrophages (1)*; *Misharin et al. 2017 - Infiltrating monocytes (2)*). Similarly, the shift from annotations representing more mature phenotypes (*Thion et al 2018 - Adult 2 (7); Matcovitch-Natan et al. 2016 - Mature microglia A1 (6)*); *Culemann et al. 2019 - CX3CR1+ lining macrophages (2)*), particularly in Cluster 5, suggests that the original tissue-resident macrophages are replaced. Finally, MAGNET allows the user to determine how the genes in a given annotation distribute across the query gene lists. As an example, we visualized the gene set specific to *Lung macrophages (4)* from *Lavin et al 2014* and found that nearly half the genes that overlap this dataset are in Clusters 5 and 6 (Fig 5.6D). Taken together, we demonstrated that MAGNET's multiple query mode enables parallel comparison of gene lists to expand conventional enrichment results and leverage the latest macrophage research for a more complete picture of the underlying biology.

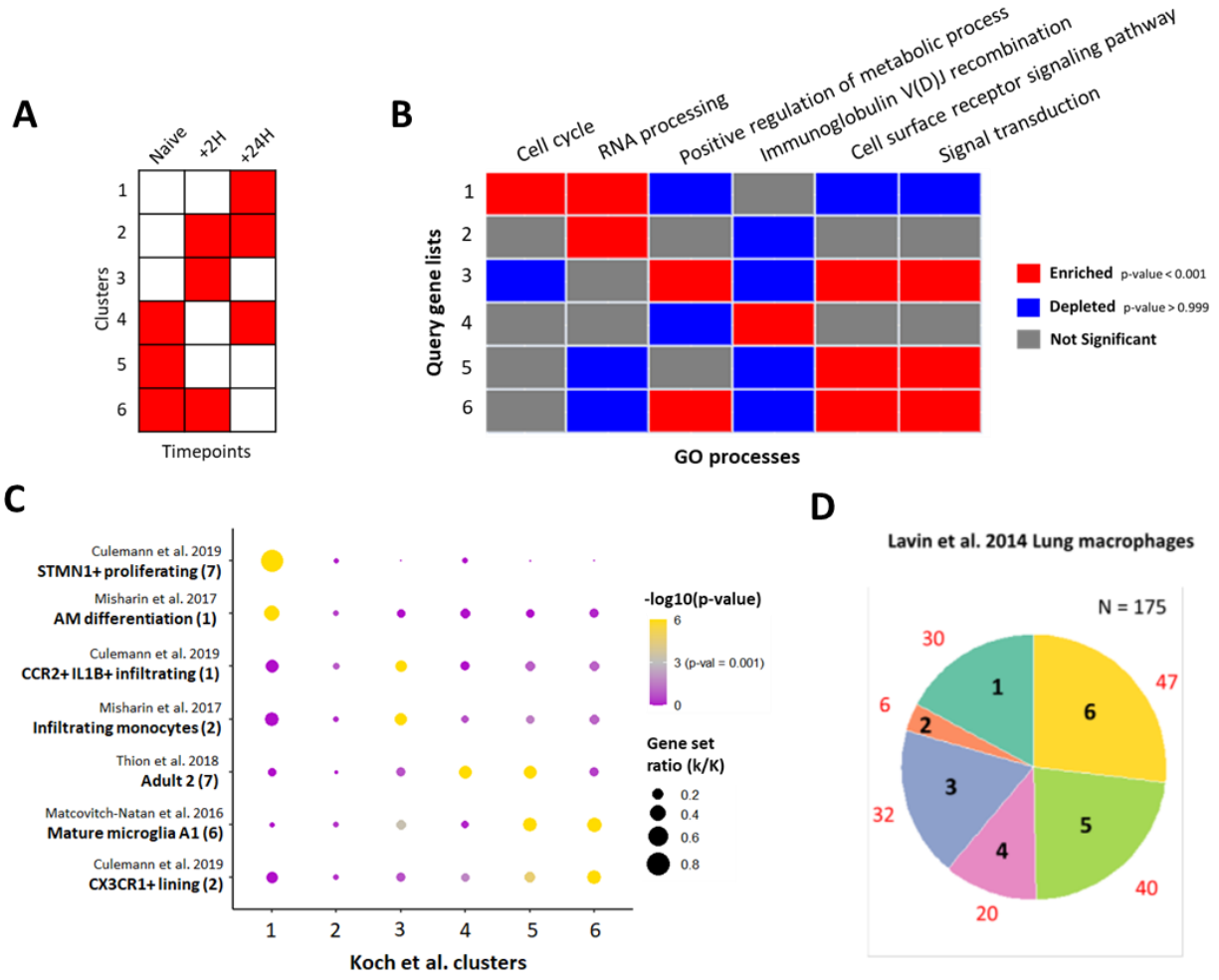
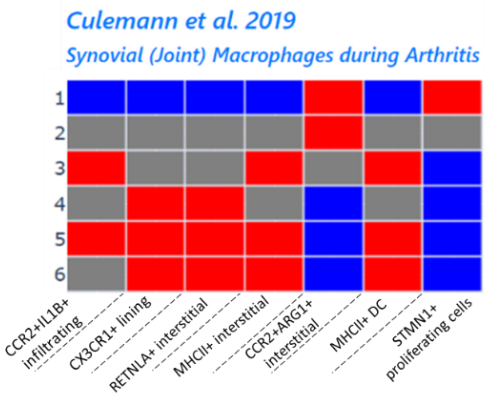
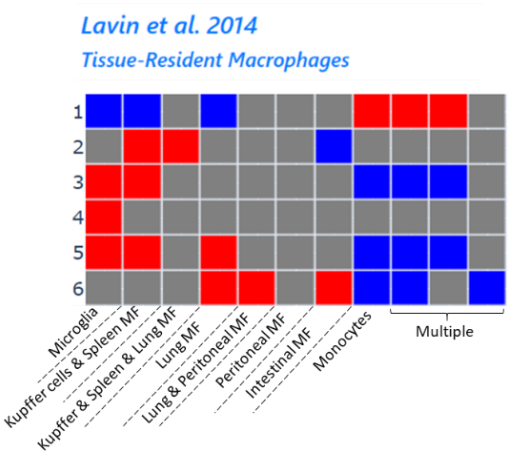
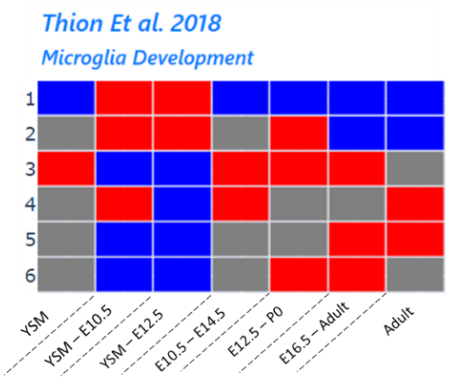
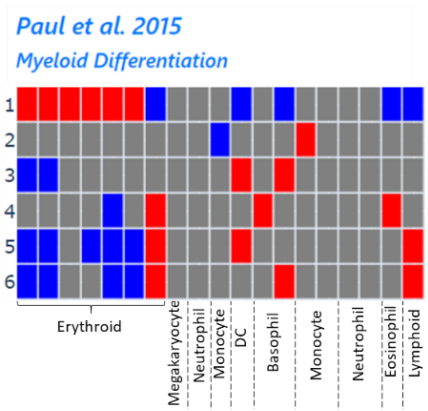


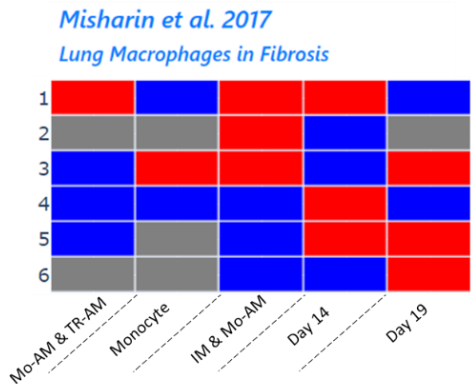
Fig 6. Running MAGNET with multiple query mode. A. Schematic of the 6 gene expression clusters defined in Koch et al 2018. Red in each row indicates the cluster(s) with the highest relative expression. **B.** Heatmap visualization from MAGNET showing enrichment of GO processes across the 6 clusters from Koch et al. 2018 using a custom user dataset with *USER* setting[139]. **C.** Enrichment of select MAGNET gene sets across the 6 clusters from Koch et al. 2018 with *INTERSECT* setting. The color scale reflects p-values (-log10) whereas the size indicates the ratio of genes in each annotated gene set that overlap the 6 clusters from Koch et al. 2018

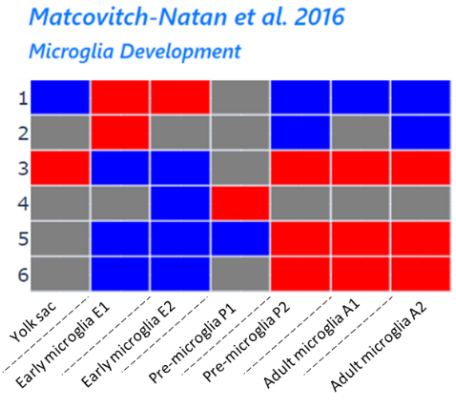

Culemann et al. 2019
Synovial (Joint) Macrophages during Arthritis
1
2
3
4
5
6
CCR2+IL1B+ infiltrating
CX3CR1+ lining
RETNLA+ interstitial
MHCII+ interstitial
CCR2+ARG1+ interstitial
MHCII+ DC
STMN1+ proliferating cells


Lavin et al. 2014
Tissue-Resident Macrophages
1
2
3
4
5
6
Microglia
Kupffer cells & Spleen MF
Kupffer & Spleen & Lung MF
Lung MF
Lung & Peritoneal MF
Peritoneal MF
Intestinal MF
Monocytes
Multiple


Thion Et al. 2018
Microglia Development
1
2
3
4
5
6
YSM
YSM – E10.5
YSM – E12.5
E10.5 – E14.5
E12.5 – P0
E16.5 – Adult
Adult


Paul et al. 2015
Myeloid Differentiation
1
2
3
4
5
6
Erythroid
Megakaryocyte
Neutrophil
Monocyte
DC
Basophil
Monocyte
Neutrophil
Eosinophil
Lymphoid


Misharin et al. 2017
Lung Macrophages in Fibrosis
1
2
3
4
5
6
Mo-AM & TR-AM
Monocyte
IM & Mo-AM
Day 14
Day 19


Matcovitch-Natan et al. 2016
Microglia Development
1
2
3
4
5
6
Yolk sac
Early microglia E1
Early microglia E2
Pre-microglia P1
Pre-microglia P2
Adult microglia A1
Adult microglia A2

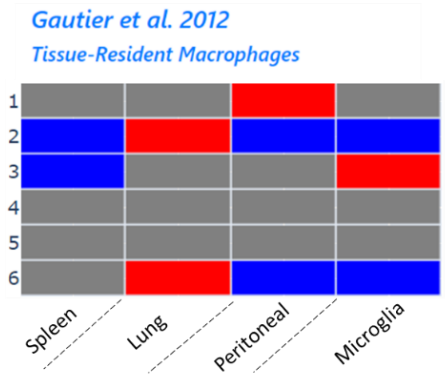

Gautier et al. 2012
Tissue-Resident Macrophages
1
2
3
4
5
6
Spleen
Lung
Peritoneal
Microglia

Figure S5.3 Heatmap visualization from MAGNET with multiple query mode. The results show enrichment for all current datasets in multiple query mode of 6 gene expression clusters from Koch et al. 2018 with *INTERSECT* setting.

5.4 Discussion

The explosion of high-throughput genomic data has facilitated rapid developments in bioinformatic tools for gene set enrichment analysis in the past couple decades. The majority of available enrichment tools aim to be all-encompassing to maximize the ability to handle data across the full range of different biological experiments. However, given the breadth and depth of biological domains, it is difficult to achieve this goal and still return results that are relevant to specific questions. In macrophage biology, this issue is exacerbated as macrophages demonstrate an astounding plasticity and play a role in multiple developmental, homeostatic, and disease processes [29, 402]. The inability of conventional enrichment tools to capture this variability is illustrated by the preponderance of generic terms among the results. For example, many contemporary studies focus on interrogating macrophage heterogeneity within a tissue in a disease model compared with steady state [80, 97, 99, 284, 408]. These studies are commonly interested in differences in ontogeny and function between macrophage subpopulations (i.e., monocyte-derived vs. tissue-resident). Thus, a tool that returns the term “inflammation” or “defense response” is unlikely to lead to meaningful biological insights. Since conventional gene set enrichment analysis is not designed for domain-specific research questions, their potential to foster cutting-edge research is limited.

We therefore designed MAGNET to address this issue by utilizing the vast amount of

annotated macrophage datasets available rather than standardized knowledge repositories. MAGNET represents an innovative approach to gene set enrichment analysis for domain-specific questions. Using Matcovitch-Natan et al. 2016 and Koch et al 2018 as examples, we demonstrated that MAGNET provides relevant and unique results when characterizing macrophage-related gene lists. Furthermore, MAGNET offers several novel functionalities that have not previously been incorporated into conventional enrichment tools including: multiple query mode, a gene search interface, alternate visualizations (heatmap and network), and the ability to upload custom dataset annotations. As an easily accessible online application with modifiable parameters, MAGNET is an approachable and flexible tool. We expect MAGNET to serve as a valuable addition to the bioinformatic workflow associated with genomic analysis in the field of macrophage biology.

Because it is based on a manually curated database, MAGNET exhibits some inherent limitations. First, the selection and availability of datasets might lead in unintentional bias in the annotations towards certain tissues, disease states, and other comparisons. This shortcoming is somewhat ameliorated by enabling the user to upload their own dataset(s), allowing for meta-analysis between any two experiments. In addition, manual curation means that the time required to incorporate a new dataset is the major bottleneck for scaling up the database. Due to differences in availability of data, some datasets are more difficult in process that others. As we continue to expand MAGNET, we will explore the possibility of automating the curation process using web scraping methods, with the goal of striking a balance between quality versus quantity. Finally, since the value of each dataset is dependent on the experimental and analytical approach of the original study, there may be variability in the relevance the output annotations. Moreover, different datasets may be associated with similar annotations (ex. tissue-resident macrophages in Lavin et

al., 2014 and Gautier et al., 2012), but result in divergent enrichment results. Thus, the user must use their best judgement to assess results and resolve any discrepancies.

We plan to continuously optimize and introduce new functionalities to MAGNET. For example, we plan to include additional metrics to improve the interpretability of enrichment results. The Jaccard index, which provides an intuitive assessment of similarity between gene sets, would extend the multiple query mode by enabling comparison of all user-supplied query gene lists against an entire MAGNET dataset. Another possibility is to output an odds ratio that, unlike p-values for which the magnitude should not be compared across tests, would provide a metric for ranking and prioritization of enriched annotations. In the current implementation, MAGNET may be used to annotate gene lists from experiments on other species by first converting into the murine orthologs. However, as the number of studies performed increases (for example, in humans), we intend to implement independent species-specific versions of the application. Furthermore, beyond gene sets, a similar approach based on published datasets could be used to annotate genomic regions, such as promoters and enhancers, as demonstrated by other tools [409, 410]. In the larger context, the software behind MAGNET could easily be adapted to other domains where conventional gene set enrichment tools are similarly limited. We envision that the introduction of MAGNET will catalyze the developments of similar and more targeted applications for different cell types, tissues, or fields, such as cancer and neurobiology.

Chapter 6. Concluding remarks

6.1 Summary

The development of NGS technology revolutionized the research of many biological domains, and the biology of macrophage is no exception. With their genetic blueprints being increasingly accessible and affordable, we now know that macrophages display remarkable diversity in their gene regulatory programs and functions, which are shaped primarily by their surrounding tissue environments. The introduction of single cell genomic assays further enables researchers to unbiasedly investigate and characterize the heterogeneity of macrophage populations that are previously thought to be homogenous within specific organs. However, the unprecedented throughput of NGS experiments also meant that datasets are now larger and more complex than ever. Rigorous development and application of novel bioinformatic algorithms is therefore essential to transform raw data generated from sequencing machines into interpretable biological insights. In this thesis, I provided examples on how integrative analysis of bulk and single cell RNA-seq data can contribute to our understanding of macrophage heterogeneity across different biological conditions.

In Chapter 2, we uncovered the existence of distinct macrophage subsets residing within the synovial tissue. In Section 2.2, we first utilized scRNA-seq to define the four potential subpopulations. Guided by information from single cell data, the four populations are confirmed and isolated through FACS and subjected to bulk RNA-seq for more thorough characterization, revealing distinct gene regulatory programs. The ability to discover novel cell populations of scRNA-seq and the higher throughput and less noisy nature of bulk RNA-seq complement each

other well, making their combination a desirable strategy to investigate macrophage heterogeneity. It is important to emphasize that we did not define novel subsets based solely on the resulting single cell clustering labels, which have been the case in many published studies. Results from clustering algorithms should be treated with caution, as they often depend on user-supplied parameters and can easily lead to arbitrary grouping of cells. We therefore employed multiple methods, including computing cell-cell correlations, examining top marker genes, and automatic assignment of identities by external references to rigorously assess the quality of clustering and merge the ones that likely came from homogenous populations. Utilizing two mouse models of arthritis, we next documented the proportional changes of the four subsets and the convergence of their transcriptional signatures towards a monocytic phenotype during pathological conditions. Finally, we were able to project the characterized mouse macrophage subsets onto human arthritic patients by developing a novel computational framework based on gene set scoring. The correlations between macrophage subset proportions and disease severity in patients may inform on potential innovative therapeutic strategies that modulate the relative proportions of subsets. In Section 2.3, following up on the observation that steady state macrophage heterogeneity is disrupted under diseased conditions, we utilized a novel mouse model of arthritis (FLIP deletion) developed in house to investigate the dynamics of monocyte infiltration and their differentiation into synovial resident macrophages. Leveraging the power of RNA-sequencing and traditional wet lab experiments, we uncovered evidence of impaired differentiation into a distinct terminal subset in FLIP-/- mice. This particular macrophage subset exhibited transcriptional signatures indicative of protective phenotypes against chronic inflammation, again highlighting the roles of tissue resident macrophages in disease pathogenesis and their potential for being therapeutic targets.

In chapter 3, utilizing a similar strategy of performing scRNA-seq followed by FACS and bulk RNA-seq, we showed that cellular heterogeneity that is previously underappreciated extends beyond tissue resident macrophages. We identified and characterized two populations of synovial resident monocytes that differs in their localization relative to vasculatures. We provided evidence that they are long-lived and maintained independently from circulation by performing computational integrative analysis of monocytes from control and CCR2/NR4A1 KO mice. Finally, we demonstrated that extravascular synovial monocytes contribute to the development of inflammatory arthritis in a LFA1-dependent manner by examining their transcriptional profiles. In chapter 4, we shifted our focus away from the joint to another organ, liver. We employed scRNA-seq to define three distinct macrophages across pediatric liver samples of BA and ALGS with distinctive transcriptional profiles. Efforts are currently underway to identify similarities and differences in these macrophage subsets across etiologies of cholestatic liver disease by investigating their counterparts in mouse model. In summary, these two chapters showcased the ability of single cell/bulk transcriptional profiling and computational analysis in advancing our knowledge of cellular heterogeneity across different cell and tissue types.

After analyzing multiple macrophage transcriptional datasets in the preceding chapters, a recurring and essential step encountered in the workflow is to attribute biological significance to the identified genes of interest. However, current approaches that utilizes standardized knowledge repositories to perform gene set/pathway enrichment statistical tests often fails to produce results relevant to macrophage biology. Therefore, in chapter 5, I developed a web application called MAGNET for performing functional enrichment tests specifically for macrophage-related genomic data. The main motive behind the application is that expert-annotated gene sets in

published literature related to macrophage biology, which are often hidden in supplemental materials and underutilized, would be a more suitable choice as reference for functional enrichment tests than standardized repositories that attempts to encompass all facets of biology. We also introduced novel methods for visualizing enrichment results for multiple gene sets at once. It is envisioned that the introduction of MAGNET would encourage the development of similar applications for more targeted and improved functional enrichment analysis in other biological domains.

6.2 Limitations and future directions

Even though the findings presented in this dissertation demonstrated the power of transcriptional profiling in investigating macrophage heterogeneity, it is not surprising that several overarching limitations exists. First of all, although the identification of tissue macrophage subsets in mouse models provided evidence of their differing functions in disease, their existence and pathogenic roles remains to be validated in humans. The novel algorithm of orthologous gene signature projection introduced in Chapter 2.2 showed potential in identifying human counterparts, but the small sample size (n=7) rendered it impossible to confidently investigate the functions in RA for each projected macrophage subset. Fortunately, as NGS assays become increasingly affordable and integrated with the standard workflows of patient tissue collection in clinical studies, we are optimistic that the amount of available sequencing data from human patients will increase substantially within the near future. Another issue in defining murine tissue macrophage subsets is the biological variations that exist between different mouse models, which can potentially alter the proportions and phenotypes of constituting cell populations, including macrophages. Differences in flow cytometry gating strategies also renders it challenging to compare and reconcile the

definitions of potential tissue resident macrophage subpopulations identified across different studies. For instance, it remains to be investigated how the four subsets identified in Chapter 2.2 relates to the two resident subsets isolated in Chapter 2.3.

Another major limitation in this dissertation is that biological processes entail a lot more than gene expression. Characterization of transcriptional landscapes is merely scratching the surfaces of the complex regulation operating behind macrophage heterogeneity. One of the major future direction, which is already underway in our group, is to employ other functional genomic NGS assays such as ATAC-seq [210] and ChIP-seq [411]. These two assays enable the mapping of epigenomic features, including open chromatin regions, promoter and enhancer locations, and binding sites of specific TFs, thus would allow us to infer the potential regulatory mechanisms underlying the diversified transcriptional profiles and functions of tissue macrophages more confidently. The recently introduced single cell versions of these epigenomic assays, scATAC-seq and scChIP-seq [412], can also be of great utility for interrogating the heterogeneity among macrophages. Another exciting potential application is spatial genomics, which is a newly developed technology that aims to profile the genomic information of single cells while simultaneously recording their relative localization within their native tissue environments [413]. Given the major roles of local environments in shaping the functional plasticity of tissue resident macrophages, employing spatial genomics can directly aid in elucidating the interactions between macrophages and the surrounding cells that potentially confer their identities. In summary, performing genomics assays that profiles different biological information and integrating their findings, also known as the “multi-omics” approach, can potentially enable us to gain an even clearer picture of how tissue resident macrophages function and interact with the surround

environments in both health and disease.

Finally, although bioinformatics as a field has prospered thanks to the development of various high-throughput sequencing assays. The development of computational methods to handle the processing and interpretation of sequencing data is still in its infant stage, especially for single cell experiments. This leads to numerous analytical methods being developed for the same bioinformatic tasks, with no consensus on whether one is more superior over others. For instance, more than 20 software packages for annotating the identities of single cells have been developed [414], while over 40 algorithms, a staggering number, are available for tackling the problem of trajectory inference [415]. Therefore, albeit we executed rigorous and robust computational analyses throughout this thesis, it is definitely possible that slightly different results may arise if we switched out some components in the workflow. Efforts for benchmarking different bioinformatic algorithms often relies on either computationally generated references or the so-called "gold-standard" datasets, where pre-determined labels are compared to results generated from algorithms being benchmarked [416]. Reference datasets generated in silico offer the advantage of being completely objective with known ground truths but cannot be trusted to emulate real biological variations. On the other hand, the "gold-standard" datasets, while representing real biological variations, is highly susceptible to selection bias and the fallacy of circular logic – the reference labeling are often determined by some of the methods being benchmarked. Therefore, I believe one of the most important future directions for the bioinformatics field is to strive for improved benchmarking datasets, both quantitatively and qualitatively. For quantitative improvement, central repositories that collect large amounts of benchmarking data can be very beneficial. Currently available examples include VariBench and VariSNP, where datasets of

validated human genetic variants are curated [417]. A similar application of transcriptomic datasets collaboratively maintained by the community would facilitate the streamlining and standardization of benchmarking efforts. The development of cross-dataset integration methods can also potentially aid in improving the power and robustness of benchmarking data by combining similar datasets into one. In regard to qualitative improvement, obtaining multi-omics information for each dataset by performing multiple genomic assays can possibly improve the labeling accuracy significantly. A major endeavor is being spearheaded ENCODE Consortium, with the goal of identifying and curating all functional regulatory elements in human genome. The emergence of the accompanying multi-omics data integration algorithms [418] also further assist in the creation of high quality "gold-standard" datasets. Collectively, collaborative efforts by the research community to improve the quantity and quality of benchmarking datasets is crucial for future development and evaluation of bioinformatic applications and algorithms.

In summary, the incorporation of multi-omics NGS assays, development of novel algorithms, and improvement in benchmarking methodologies holds great promise to address the limitations discussed above and further revolutionize understanding of macrophage heterogeneity in humans.

References

1. Metchnikoff, E., *Untersuchungen über die mesodermalen Phagocyten einiger Wirbeltiere.* Biol Zentralbl, 1883. **3**: p. 560-5.
2. Metschnikoff, E., *Lecture on phagocytosis and immunity.* British medical journal, 1891. **1**(1570): p. 213.
3. van Furth, R., et al., *The mononuclear phagocyte system: a new classification of macrophages, monocytes, and their precursor cells.* Bulletin of the World Health Organization, 1972. **46**(6): p. 845.
4. Aderem, A. and D.M. Underhill, *Mechanisms of phagocytosis in macrophages.* Annual review of immunology, 1999. **17**(1): p. 593-623.
5. Hirayama, D., T. Iida, and H. Nakase, *The Phagocytic Function of Macrophage-Enforcing Innate Immunity and Tissue Homeostasis.* Int J Mol Sci, 2017. **19**(1).
6. Unanue, E.R., *Antigen-presenting function of the macrophage.* Annual review of immunology, 1984. **2**: p. 395-428.
7. Arango Duque, G. and A. Descoteaux, *Macrophage cytokines: involvement in immunity and infectious diseases.* Frontiers in immunology, 2014. **5**: p. 491.
8. Davies, L.C., et al., *Tissue-resident macrophages.* Nature immunology, 2013. **14**(10): p. 986.
9. Davies, L.C. and P.R. Taylor, *Tissue-resident macrophages: then and now.* Immunology, 2015. **144**(4): p. 541-8.
10. Paolicelli, R.C., et al., *Synaptic pruning by microglia is necessary for normal brain development.* science, 2011. **333**(6048): p. 1456-1458.
11. Li, Q. and B.A. Barres, *Microglia and macrophages in brain homeostasis and disease.*

Nature Reviews Immunology, 2018. **18**(4): p. 225-242.

12. Willekens, F.L., et al., *Liver Kupffer cells rapidly remove red blood cell–derived vesicles from the circulation by scavenger receptors.* Blood, 2005. **105**(5): p. 2141-2145.
13. Krenkel, O. and F. Tacke, *Liver macrophages in tissue homeostasis and disease.* Nature Reviews Immunology, 2017. **17**(5): p. 306-321.
14. Lambrecht, B.N., *Alveolar macrophage in the driver's seat.* Immunity, 2006. **24**(4): p. 366-8.
15. Hussell, T. and T.J. Bell, *Alveolar macrophages: plasticity in a tissue-specific context.* Nature Reviews Immunology, 2014. **14**(2): p. 81-93.
16. Chakarov, S., et al., *Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches.* Science, 2019. **363**(6432).
17. Blériot, C. and F. Ginhoux, *Understanding the Heterogeneity of Resident Liver Macrophages.* Front Immunol, 2019. **10**: p. 2694.
18. Bain, C.C. and A. Schridde, *Origin, Differentiation, and Function of Intestinal Macrophages.* Frontiers in Immunology, 2018. **9**(2733).
19. Shaw, T.N., et al., *Tissue-resident macrophages in the intestine are long lived and defined by Tim-4 and CD4 expression.* J Exp Med, 2018. **215**(6): p. 1507-1518.
20. Goldmann, T., et al., *Origin, fate and dynamics of macrophages at central nervous system interfaces.* Nature Immunology, 2016. **17**(7): p. 797-805.
21. Masuda, T., et al., *Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution.* Nature, 2019. **566**(7744): p. 388-392.
22. Ajami, B., et al., *Local self-renewal can sustain CNS microglia maintenance and function throughout adult life.* Nature Neuroscience, 2007. **10**(12): p. 1538-1543.

23. Tarling, J.D., H.s. Lin, and S. Hsu, *Self-renewal of pulmonary alveolar macrophages: evidence from radiation chimera studies.* Journal of leukocyte biology, 1987. **42**(5): p. 443-446.

24. Merad, M., et al., *Langerhans cells renew in the skin throughout life under steady-state conditions.* Nature immunology, 2002. **3**(12): p. 1135-1141.

25. Ginhoux, F., et al., *Fate mapping analysis reveals that adult microglia derive from primitive macrophages.* Science, 2010. **330**(6005): p. 841-845.

26. Yona, S., et al., *Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis.* Immunity, 2013. **38**(1): p. 79-91.

27. Hoeffel, G., et al., *Adult Langerhans cells derive predominantly from embryonic fetal liver monocytes with a minor contribution of yolk sac–derived macrophages.* Journal of Experimental Medicine, 2012. **209**(6): p. 1167-1181.

28. Palis, J., et al., *Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse.* Development, 1999. **126**(22): p. 5073-84.

29. Ginhoux, F. and M. Guilliams, *Tissue-resident macrophage ontogeny and homeostasis.* Immunity, 2016. **44**(3): p. 439-449.

30. Palis, J. and M.C. Yoder, *Yolk-sac hematopoiesis: the first blood cells of mouse and man.* Exp Hematol, 2001. **29**(8): p. 927-36.

31. Medvinsky, A. and E. Dzierzak, *Definitive hematopoiesis is autonomously initiated by the AGM region.* Cell, 1996. **86**(6): p. 897-906.

32. Blériot, C., S. Chakarov, and F. Ginhoux, *Determinants of Resident Tissue Macrophage Identity and Function.* Immunity, 2020. **52**(6): p. 957-970.

33. Cumano, A. and I. Godin, *Ontogeny of the hematopoietic system.* Annu Rev Immunol,

2007. **25**: p. 745-85.

34. Katsumoto, A., et al., *Ontogeny and functions of central nervous system macrophages.* Journal of immunology (Baltimore, Md. : 1950), 2014. **193**(6): p. 2615-2621.

35. Tan, S.Y. and M.A. Krasnow, *Developmental origin of lung macrophage diversity.* Development, 2016. **143**(8): p. 1318-27.

36. Guillot, A. and F. Tacke, *Liver Macrophages: Old Dogmas and New Insights.* Hepatol Commun, 2019. **3**(6): p. 730-743.

37. Molawi, K., et al., *Progressive replacement of embryo-derived cardiac macrophages with age.* J Exp Med, 2014. **211**(11): p. 2151-8.

38. Bain, C.C. and A.M. Mowat, *The monocyte-macrophage axis in the intestine.* Cellular Immunology, 2014. **291**(1): p. 41-48.

39. Cassado Ados, A., M.R. D'Império Lima, and K.R. Bortoluci, *Revisiting mouse peritoneal macrophages: heterogeneity, development, and function.* Front Immunol, 2015. **6**: p. 225.

40. Okabe, Y. and R. Medzhitov, *Tissue-specific signals control reversible program of localization and functional polarization of macrophages.* Cell, 2014. **157**(4): p. 832-44.

41. Kurotaki, D., T. Uede, and T. Tamura, *Functions and development of red pulp macrophages.* Microbiol Immunol, 2015. **59**(2): p. 55-62.

42. Wang, S., et al., *Functions of Macrophages in the Maintenance of Intestinal Homeostasis.* Journal of Immunology Research, 2019. **2019**: p. 1512969.

43. Clayton, K., et al., *Langerhans Cells—Programmed by the Epidermis.* Frontiers in Immunology, 2017. **8**(1676).

44. Munro, D.A.D. and J. Hughes, *The Origins and Functions of Tissue-Resident*

Macrophages in Kidney Development. Frontiers in physiology, 2017. **8**: p. 837-837.

45. Hulsmans, M., et al., *Macrophages Facilitate Electrical Conduction in the Heart.* Cell, 2017. **169**(3): p. 510-522.e20.

46. van de Laar, L., et al., *Yolk Sac Macrophages, Fetal Liver, and Adult Monocytes Can Colonize an Empty Niche and Develop into Functional Tissue-Resident Macrophages.* Immunity, 2016. **44**(4): p. 755-68.

47. Lavin, Y., et al., *Tissue-Resident Macrophage Enhancer Landscapes Are Shaped by the Local Microenvironment.* Cell, 2014. **159**(6): p. 1312-1326.

48. Celada, A., et al., *The transcription factor PU.1 is involved in macrophage proliferation.* J Exp Med, 1996. **184**(1): p. 61-9.

49. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Molecular cell, 2010. **38**(4): p. 576-589.

50. Hamada, M., et al., *Role of MafB in macrophages.* Exp Anim, 2020. **69**(1): p. 1-10.

51. Gautier, E.L., et al., *Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages.* Nature Immunology, 2012. **13**(11): p. 1118-1128.

52. Pradhan, P., et al., *Interplay of Heme with Macrophages in Homeostasis and Inflammation.* Int J Mol Sci, 2020. **21**(3).

53. Schneider, C., et al., *Induction of the nuclear receptor PPAR-γ by the cytokine GM-CSF is critical for the differentiation of fetal monocytes into alveolar macrophages.* Nat Immunol, 2014. **15**(11): p. 1026-37.

54. Gosselin, D., et al., *Environment drives selection and function of enhancers controlling*

tissue-specific macrophage identities. Cell, 2014. **159**(6): p. 1327-40.

55. Guilliams, M. and C.L. Scott, *Does niche competition determine the origin of tissue-resident macrophages?* Nature Reviews Immunology, 2017. **17**(7): p. 451-460.

56. Guilliams, M., et al., *Establishment and maintenance of the macrophage niche.* Immunity, 2020. **52**(3): p. 434-451.

57. Wang, Y., et al., *IL-34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia.* Nat Immunol, 2012. **13**(8): p. 753-60.

58. Kana, V., et al., *CSF-1 controls cerebellar microglia and is required for motor function and social interaction.* J Exp Med, 2019. **216**(10): p. 2265-2281.

59. Easley-Neal, C., et al., *CSF1R Ligands IL-34 and CSF1 Are Differentially Required for Microglia Development and Maintenance in White and Gray Matter Brain Regions.* Frontiers in Immunology, 2019. **10**(2199).

60. Bessis, A., et al., *Microglial control of neuronal death and synaptic properties.* Glia, 2007. **55**(3): p. 233-8.

61. Bonnardel, J., et al., *Stellate Cells, Hepatocytes, and Endothelial Cells Imprint the Kupffer Cell Identity on Monocytes Colonizing the Liver Macrophage Niche.* Immunity, 2019. **51**(4): p. 638-654.e9.

62. Sakai, M., et al., *Liver-Derived Signals Sequentially Reprogram Myeloid Enhancers to Initiate and Maintain Kupffer Cell Identity.* Immunity, 2019. **51**(4): p. 655-670.e8.

63. Sierro, F., et al., *A Liver Capsular Network of Monocyte-Derived Macrophages Restricts Hepatic Dissemination of Intraperitoneal Bacteria by Neutrophil Recruitment.* Immunity, 2017. **47**(2): p. 374-388.e6.

64. Borges da Silva, H., et al., *Splenic Macrophage Subsets and Their Function during*

Blood-Borne Infections. Frontiers in Immunology, 2015. **6**: p. 480.

65. Medzhitov, R., *Origin and physiological roles of inflammation.* Nature, 2008. **454**(7203): p. 428-435.

66. Blériot, C., et al., *Liver-resident macrophage necroptosis orchestrates type 1 microbicidal inflammation and type-2-mediated tissue repair during bacterial infection.* Immunity, 2015. **42**(1): p. 145-158.

67. Machiels, B., et al., *A gammaherpesvirus provides protection against allergic asthma by inducing the replacement of resident alveolar macrophages with regulatory monocytes.* Nature immunology, 2017. **18**(12): p. 1310.

68. Shi, C. and E.G. Pamer, *Monocyte recruitment during infection and inflammation.* Nat Rev Immunol, 2011. **11**(11): p. 762-74.

69. Scott, C.L., et al., *Bone marrow-derived monocytes give rise to self-renewing and fully differentiated Kupffer cells.* Nat Commun, 2016. **7**: p. 10321.

70. Louwe, P.A., et al., *Recruited macrophages that colonize the post-inflammatory peritoneal niche convert into functionally divergent resident cells.* Nature Communications, 2021. **12**(1): p. 1770.

71. Saeed, S., et al., *Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity.* Science, 2014. **345**(6204): p. 1251086.

72. Netea, M.G., et al., *Defining trained immunity and its role in health and disease.* Nature Reviews Immunology, 2020. **20**(6): p. 375-388.

73. Ziegler-Heitbrock, L., *Monocyte subsets in man and other species.* Cell Immunol, 2014. **289**(1-2): p. 135-9.

74. Arnold, L., et al., *Inflammatory monocytes recruited after skeletal muscle injury switch*

into antiinflammatory macrophages to support myogenesis. J Exp Med, 2007. **204**(5): p. 1057-69.

75. Westman, J., S. Grinstein, and P.E. Marques, *Phagocytosis of Necrotic Debris at Sites of Injury and Inflammation.* Frontiers in Immunology, 2020. **10**(3030).

76. Oishi, Y. and I. Manabe, *Macrophages in inflammation, repair and regeneration.* Int Immunol, 2018. **30**(11): p. 511-528.

77. Ma, Y., A.J. Mouton, and M.L. Lindsey, *Cardiac macrophage biology in the steady-state heart, the aging heart, and following myocardial infarction.* Translational Research, 2018. **191**: p. 15-28.

78. Arango Duque, G. and A. Descoteaux, *Macrophage cytokines: involvement in immunity and infectious diseases.* Front Immunol, 2014. **5**: p. 491.

79. Wynn, T.A. and K.M. Vannella, *Macrophages in Tissue Repair, Regeneration, and Fibrosis.* Immunity, 2016. **44**(3): p. 450-462.

80. Misharin, A.V., et al., *Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span.* J Exp Med, 2017. **214**(8): p. 2387-2404.

81. Moore, K.J., F.J. Sheedy, and E.A. Fisher, *Macrophages in atherosclerosis: a dynamic balance.* Nat Rev Immunol, 2013. **13**(10): p. 709-21.

82. Mosser, D.M. and J.P. Edwards, *Exploring the full spectrum of macrophage activation.* Nature reviews. Immunology, 2008. **8**(12): p. 958-969.

83. Orecchioni, M., et al., *Macrophage Polarization: Different Gene Signatures in M1(LPS+) vs. Classically and M2(LPS–) vs. Alternatively Activated Macrophages.* Frontiers in Immunology, 2019. **10**(1084).

84. Xue, J., et al., *Transcriptome-based network analysis reveals a spectrum model of human*

macrophage activation. Immunity, 2014. **40**(2): p. 274-88.

85. Martinez, F.O. and S. Gordon, *The M1 and M2 paradigm of macrophage activation: time for reassessment.* F1000prime reports, 2014. **6**: p. 13-13.

86. Rigamonti, E., et al., *Macrophage plasticity in skeletal muscle repair.* BioMed research international, 2014. **2014**: p. 560629-560629.

87. O'Rourke, S.A., A. Dunne, and M.G. Monaghan, *The Role of Macrophages in the Infarcted Myocardium: Orchestrators of ECM Remodeling.* Frontiers in cardiovascular medicine, 2019. **6**: p. 101-101.

88. Dick, S.A., et al., *Self-renewing resident cardiac macrophages limit adverse remodeling following myocardial infarction.* Nature Immunology, 2019. **20**(1): p. 29-39.

89. Tetley, T.D., *Macrophages and the pathogenesis of COPD.* Chest, 2002. **121**(5 Suppl): p. 156s-159s.

90. Kazankov, K., et al., *The role of macrophages in nonalcoholic fatty liver disease and nonalcoholic steatohepatitis.* Nat Rev Gastroenterol Hepatol, 2019. **16**(3): p. 145-159.

91. Liu, Y.C., et al., *Macrophage polarization in inflammatory diseases.* Int J Biol Sci, 2014. **10**(5): p. 520-9.

92. Ginhoux, F., et al., *New insights into the multidimensional concept of macrophage ontogeny, activation and function.* Nat Immunol, 2016. **17**(1): p. 34-40.

93. Adan, A., et al., *Flow cytometry: basic principles and applications.* Critical Reviews in Biotechnology, 2017. **37**(2): p. 163-176.

94. Hulett, H.R., et al., *Cell sorting: automated separation of mammalian cells as a function of intracellular fluorescence.* Science, 1969. **166**(3906): p. 747-749.

95. Shields, C.W.t., C.D. Reyes, and G.P. López, *Microfluidic cell sorting: a review of the*

advances in the separation of cells from debulking to rare cell isolation. Lab Chip, 2015. **15**(5): p. 1230-49.

96. Misharin, A.V., et al., *Flow cytometric analysis of macrophages and dendritic cell subsets in the mouse lung.* Am J Respir Cell Mol Biol, 2013. **49**(4): p. 503-10.

97. Epelman, S., et al., *Embryonic and adult-derived resident cardiac macrophages are maintained through distinct mechanisms at steady state and during inflammation.* Immunity, 2014. **40**(1): p. 91-104.

98. Evren, E., et al., *Distinct developmental pathways from blood monocytes generate human lung macrophage diversity.* Immunity, 2021. **54**(2): p. 259-275.e7.

99. Seidman, J.S., et al., *Niche-Specific Reprogramming of Epigenetic Landscapes Drives Myeloid Cell Diversity in Nonalcoholic Steatohepatitis.* Immunity, 2020. **52**(6): p. 1057-1074.e7.

100. Liddelow, S.A., et al., *Neurotoxic reactive astrocytes are induced by activated microglia.* Nature, 2017. **541**(7638): p. 481-487.

101. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-563.

102. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proceedings of the national academy of sciences, 1977. **74**(12): p. 5463-5467.

103. Collins, F.S., M. Morgan, and A. Patrinos, *The Human Genome Project: lessons from large-scale biology.* Science, 2003. **300**(5617): p. 286-290.

104. Stranneheim, H. and J. Lundeberg, *Stepping stones in DNA sequencing.* Biotechnology journal, 2012. **7**(9): p. 1063-1073.

105. Liu, L., et al., *Comparison of next-generation sequencing systems.* Journal of Biomedicine and Biotechnology, 2012. **2012**.

106. Illumina, *An introduction to Next-Generation Sequencing Technology.*

107. Pease, A.C., et al., *Light-generated oligonucleotide arrays for rapid DNA sequence analysis.* Proceedings of the National Academy of Sciences, 1994. **91**(11): p. 5022-5026.

108. DeRisi, J., et al., *Use of a cDNA microarray to analyse gene expression patterns in human cancer.* Nat Genet, 1996. **14**(4): p. 457-60.

109. Wu, L., P. Mickey Williams, and W.H. Koch, *Clinical applications of microarray-based diagnostic tests.* Biotechniques, 2005. **39**(4S): p. S557-S582.

110. Mortazavi, A., et al., *Mapping and quantifying mammalian transcriptomes by RNA-Seq.* Nature Methods, 2008. **5**(7): p. 621-628.

111. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

112. Illumina, *Advantages of paired-end and single-read sequencing.*

113. Zhao, S., et al., *Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion.* Scientific Reports, 2018. **8**(1): p. 4781.

114. Tarazona, S., et al., *Differential expression in RNA-seq: a matter of depth.* Genome Res, 2011. **21**(12): p. 2213-23.

115. Ching, T., S. Huang, and L.X. Garmire, *Power analysis and sample size estimation for RNA-Seq differential expression.* Rna, 2014. **20**(11): p. 1684-1696.

116. Illumina, *bcl2fastq2 Conversion Software v2.20.*

117. S, A., *FASTQC. A quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

118. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina*

sequence data. Bioinformatics, 2014. **30**(15): p. 2114-2120.

119. Wang, L., S. Wang, and W. Li, *RSeQC: quality control of RNA-seq experiments.* Bioinformatics, 2012. **28**(16): p. 2184-2185.

120. Burrows, M. and D. Wheeler. *A block-sorting lossless data compression algorithm*. in *Digital SRC Research Report*. 1994. Citeseer.

121. Needleman, S.B. and C.D. Wunsch, *A general method applicable to the search for similarities in the amino acid sequence of two proteins.* Journal of Molecular Biology, 1970. **48**(3): p. 443-453.

122. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature methods, 2012. **9**(4): p. 357.

123. Li, H., *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.* arXiv preprint arXiv:1303.3997, 2013.

124. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

125. Kim, D., et al., *Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.* Nature biotechnology, 2019. **37**(8): p. 907-915.

126. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome biology, 2013. **14**(4): p. 1-13.

127. Haas, B.J., et al., *De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.* Nature Protocols, 2013. **8**(8): p. 1494-1512.

128. Schulz, M.H., et al., *Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels.* Bioinformatics, 2012. **28**(8): p. 1086-1092.

129. Anders, S., P.T. Pyl, and W. Huber, *HTSeq—a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-169.

130. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-930.

131. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nature Biotechnology, 2010. **28**(5): p. 511-515.

132. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC bioinformatics, 2011. **12**(1): p. 1-16.

133. Pertea, M., et al., *StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.* Nature biotechnology, 2015. **33**(3): p. 290-295.

134. Chandramohan, R., et al., *Benchmarking RNA-Seq quantification tools.* Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2013. **2013**: p. 647-650.

135. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nature biotechnology, 2016. **34**(5): p. 525-527.

136. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nature methods, 2017. **14**(4): p. 417-419.

137. Wu, D.C., et al., *Limitations of alignment-free tools in total RNA-seq quantification.* BMC genomics, 2018. **19**(1): p. 1-14.

138. Conesa, A., et al., *A survey of best practices for RNA-seq data analysis.* Genome Biology,

2016. **17**(1): p. 13.

139. Koch, C.M., et al., *A beginner's guide to analysis of RNA sequencing data.* American journal of respiratory cell and molecular biology, 2018. **59**(2): p. 145-157.

140. Jolliffe, I.T. and J. Cadima, *Principal component analysis: a review and recent developments.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2016. **374**(2065): p. 20150202.

141. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 1-21.

142. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biology, 2010. **11**(10): p. R106.

143. Hardcastle, T.J. and K.A. Kelly, *baySeq: empirical Bayesian methods for identifying differential expression in sequence count data.* BMC bioinformatics, 2010. **11**(1): p. 1-14.

144. Tarazona, S., et al., *Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package.* Nucleic acids research, 2015. **43**(21): p. e140-e140.

145. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic acids research, 2015. **43**(7): p. e47-e47.

146. Pertea, M., et al., *Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.* Nature Protocols, 2016. **11**(9): p. 1650-1667.

147. Ashburner, M., et al., *Gene ontology: tool for the unification of biology.* Nature genetics, 2000. **25**(1): p. 25-29.

148. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.

149. Eden, E., et al., *GOrilla: a tool for discovery and visualization of enriched GO terms in*

ranked gene lists. BMC bioinformatics, 2009. **10**(1): p. 1-7.

150. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

151. Mi, H. and P. Thomas, *PANTHER pathway: an ontology-based pathway database coupled with data analysis tools.* Methods Mol Biol, 2009. **563**: p. 123-40.

152. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.

153. Mootha, V.K., et al., *PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nature Genetics, 2003. **34**(3): p. 267-273.

154. Subramanian, A., 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**(43): p. 15545-15550.

155. Heng, T.S. and M.W. Painter, *The Immunological Genome Project: networks of gene expression in immune cells.* Nat Immunol, 2008. **9**(10): p. 1091-4.

156. Jojic, V., et al., *Identification of transcriptional regulators in the mouse immune system.* Nat Immunol, 2013. **14**(6): p. 633-43.

157. Matcovitch-Natan, O., et al., *Microglia development follows a stepwise program to regulate brain homeostasis.* Science, 2016. **353**(6301): p. aad8670.

158. Thion, M.S., et al., *Microbiome Influences Prenatal and Adult Microglia in a Sex-Specific Manner.* Cell, 2018. **172**(3): p. 500-516.e16.

159. Tang, F., et al., *mRNA-Seq whole-transcriptome analysis of a single cell.* Nature Methods, 2009. **6**(5): p. 377-382.

160. Picelli, S., et al., *Full-length RNA-seq from single cells using Smart-seq2.* Nature Protocols, 2014. **9**(1): p. 171-181.

161. Hashimshony, T., et al., *CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq.* Genome Biology, 2016. **17**(1): p. 77.

162. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-1214.

163. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nature communications, 2017. **8**(1): p. 1-12.

164. Sasagawa, Y., et al., *Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads.* Genome Biology, 2018. **19**(1): p. 29.

165. Jaitin, D.A., et al., *Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types.* Science, 2014. **343**(6172): p. 776.

166. Ding, J., et al., *Systematic comparative analysis of single cell RNA-sequencing methods.* bioRxiv, 2019: p. 632216.

167. Islam, S., et al., *Quantitative single-cell RNA-seq with unique molecular identifiers.* Nature Methods, 2014. **11**(2): p. 163-166.

168. Cochain, C., et al., *Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis.* Circ Res, 2018. **122**(12): p. 1661-1674.

169. Ochocka, N., et al., *Single-cell RNA sequencing reveals functional heterogeneity of glioma-associated brain macrophages.* Nature Communications, 2021. **12**(1): p. 1151.

170. Schyns, J., et al., *Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung.* Nature

Communications, 2019. **10**(1): p. 3964.

171. Joshi, N., et al., *A spatially restricted fibrotic niche in pulmonary fibrosis is sustained by M-CSF/M-CSFR signalling in monocyte-derived alveolar macrophages.* Eur Respir J, 2020. **55**(1).

172. Saviano, A., N.C. Henderson, and T.F. Baumert, *Single-cell genomics and spatial transcriptomics: Discovery of novel cell states and cellular interactions in liver physiology and disease biology.* Journal of Hepatology, 2020. **73**(5): p. 1219-1230.

173. Lun, A.T.L., et al., *EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data.* Genome Biology, 2019. **20**(1): p. 63.

174. Stuart, T., et al., *Comprehensive integration of single-cell data.* Cell, 2019. **177**(7): p. 1888-1902. e21.

175. McCarthy, D.J., et al., *Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R.* Bioinformatics, 2017. **33**(8): p. 1179-1186.

176. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome Biology, 2018. **19**(1): p. 15.

177. Lun, A.T., D.J. McCarthy, and J.C. Marioni, *A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor.* F1000Res, 2016. **5**: p. 2122.

178. Hafemeister, C. and R. Satija, *Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression.* Genome Biology, 2019. **20**(1): p. 296.

179. Maaten, L.v.d. and G. Hinton, *Visualizing data using t-SNE.* Journal of machine learning research, 2008. **9**(Nov): p. 2579-2605.

180. McInnes, L., J. Healy, and J. Melville, *Umap: Uniform manifold approximation and*

projection for dimension reduction. arXiv preprint arXiv:1802.03426, 2018.

181. Becht, E., et al., *Dimensionality reduction for visualizing single-cell data using UMAP.* Nature biotechnology, 2019. **37**(1): p. 38.

182. Blondel, V.D., et al., *Fast unfolding of communities in large networks.* Journal of statistical mechanics: theory and experiment, 2008. **2008**(10): p. P10008.

183. Kiselev, V.Y., et al., *SC3: consensus clustering of single-cell RNA-seq data.* Nat Methods, 2017. **14**(5): p. 483-486.

184. Van Gassen, S., et al., *FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data.* Cytometry A, 2015. **87**(7): p. 636-45.

185. Pasquini, G., et al., *Automated methods for cell type annotation on scRNA-seq data.* Computational and Structural Biotechnology Journal, 2021. **19**: p. 961-969.

186. Zhang, A.W., et al., *Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling.* Nature methods, 2019. **16**(10): p. 1007-1015.

187. Aran, D., et al., *Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage.* Nature immunology, 2019. **20**(2): p. 163-172.

188. Pliner, H.A., J. Shendure, and C. Trapnell, *Supervised classification enables rapid annotation of cell atlases.* Nature methods, 2019. **16**(10): p. 983-986.

189. Luecken, M.D. and F.J. Theis, *Current best practices in single-cell RNA-seq analysis: a tutorial.* Mol Syst Biol, 2019. **15**(6): p. e8746.

190. Finak, G., et al., *MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data.* Genome Biology, 2015. **16**(1): p. 278.

191. Kharchenko, P.V., L. Silberstein, and D.T. Scadden, *Bayesian approach to single-cell*

differential expression analysis. Nature methods, 2014. **11**(7): p. 740-742.

192. Miao, Z., et al., *DEsingle for detecting three types of differential expression in single-cell RNA-seq data.* Bioinformatics, 2018. **34**(18): p. 3223-3224.

193. Chen, G., B. Ning, and T. Shi, *Single-Cell RNA-Seq Technologies and Related Computational Data Analysis.* Frontiers in Genetics, 2019. **10**(317).

194. Soneson, C. and M.D. Robinson, *Bias, robustness and scalability in single-cell differential expression analysis.* Nat Methods, 2018. **15**(4): p. 255-261.

195. Huang, M., et al., *SAVER: gene expression recovery for single-cell RNA sequencing.* Nature Methods, 2018. **15**(7): p. 539-542.

196. van Dijk, D., et al., *Recovering Gene Interactions from Single-Cell Data Using Data Diffusion.* Cell, 2018. **174**(3): p. 716-729.e27.

197. Talwar, D., et al., *AutoImpute: Autoencoder based imputation of single-cell RNA-seq data.* Scientific Reports, 2018. **8**(1): p. 16329.

198. Buettner, F., et al., *f-scLVM: scalable and versatile factor analysis for single-cell RNA-seq.* Genome Biology, 2017. **18**(1): p. 212.

199. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species.* Nature Biotechnology, 2018. **36**(5): p. 411-420.

200. Haghverdi, L., et al., *Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors.* Nature biotechnology, 2018. **36**(5): p. 421-427.

201. Korsunsky, I., et al., *Fast, sensitive and accurate integration of single-cell data with Harmony.* Nature methods, 2019. **16**(12): p. 1289-1296.

202. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-883.

203. Cao, J., et al., *The single-cell transcriptional landscape of mammalian organogenesis.* Nature, 2019. **566**(7745): p. 496-502.

204. Ji, Z. and H. Ji, *TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis.* Nucleic Acids Research, 2016. **44**(13): p. e117-e117.

205. Street, K., et al., *Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics.* BMC Genomics, 2018. **19**(1): p. 477.

206. Wolf, F.A., et al., *PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells.* Genome Biology, 2019. **20**(1): p. 59.

207. Aibar, S., et al., *SCENIC: single-cell regulatory network inference and clustering.* Nature Methods, 2017. **14**(11): p. 1083-1086.

208. Cabello-Aguilar, S., et al., *SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics.* Nucleic Acids Research, 2020. **48**(10): p. e55-e55.

209. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nature Methods, 2017. **14**(9): p. 865-868.

210. Buenrostro, J.D., et al., *Single-cell chromatin accessibility reveals principles of regulatory variation.* Nature, 2015. **523**(7561): p. 486-490.

211. Zhu, C., S. Preissl, and B. Ren, *Single-cell multimodal omics: the power of many.* Nature Methods, 2020. **17**(1): p. 11-14.

212. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* bioRxiv, 2020: p. 2020.10.12.335331.

213. Johnston Jr, R.B., *Monocytes and macrophages.* New England Journal of Medicine, 1988. **318**(12): p. 747-752.

214. Ginhoux, F. and S. Jung, *Monocytes and macrophages: developmental pathways and*

tissue homeostasis. Nature Reviews Immunology, 2014. **14**(6): p. 392-404.

215. Epelman, S., K.J. Lavine, and G.J. Randolph, *Origin and functions of tissue macrophages.* Immunity, 2014. **41**(1): p. 21-35.

216. Li, Q. and B.A. Barres, *Microglia and macrophages in brain homeostasis and disease.* Nature Reviews Immunology, 2018. **18**(4): p. 225.

217. Smith, M.D., *The normal synovium.* Open Rheumatol J, 2011. **5**: p. 100-6.

218. Kurowska-Stolarska, M. and S. Alivernini, *Synovial tissue macrophages: friend or foe?* RMD Open, 2017. **3**(2): p. e000527.

219. Culemann, S., A. Grüneboom, and G. Krönke, *Origin and function of synovial macrophage subsets during inflammatory joint disease*, in *Advances in immunology*. 2019, Elsevier. p. 75-98.

220. Culemann, S., et al., *Locally renewing resident synovial macrophages provide a protective barrier for the joint.* Nature, 2019.

221. Hettinger, J., et al., *Origin of monocytes and macrophages in a committed progenitor.* Nature immunology, 2013. **14**(8): p. 821-830.

222. Schulz, C., et al., *A lineage of myeloid cells independent of Myb and hematopoietic stem cells.* Science, 2012. **336**(6077): p. 86-90.

223. Gomez Perdiguero, E., et al., *Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors.* Nature, 2015. **518**(7540): p. 547-551.

224. Morales-Nebreda, L., et al., *The heterogeneity of lung macrophages in the susceptibility to disease.* European Respiratory Review, 2015. **24**(137): p. 505-509.

225. Misharin, A.V., et al., *Nonclassical Ly6C− monocytes drive the development of inflammatory arthritis in mice.* Cell reports, 2014. **9**(2): p. 591-604.

226. Tu, J., et al., *Ontogeny of synovial macrophages and the roles of synovial macrophages from different origins in arthritis.* Frontiers in immunology, 2019. **10**: p. 1146.

227. Zhang, F., et al., *Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry.* Nature immunology, 2019. **20**(7): p. 928-942.

228. Alivernini, S., et al., *Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis.* Nature Medicine, 2020.

229. Tran, C.N., S.K. Lundy, and D.A. Fox, *Synovial biology and T cells in rheumatoid arthritis.* Pathophysiology, 2005. **12**(3): p. 183-9.

230. Firestein, G.S., *Evolving concepts of rheumatoid arthritis.* Nature, 2003. **423**(6937): p. 356-361.

231. McInnes, I.B. and G. Schett, *The Pathogenesis of Rheumatoid Arthritis.* New England Journal of Medicine, 2011. **365**(23): p. 2205-2219.

232. Guo, Q., et al., *Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies.* Bone Res, 2018. **6**: p. 15.

233. Kennedy, A., et al., *Macrophages in synovial inflammation.* Frontiers in immunology, 2011. **2**: p. 52.

234. Siouti, E. and E. Andreakos, *The many facets of macrophages in rheumatoid arthritis.* Biochem Pharmacol, 2019. **165**: p. 152-169.

235. Udalova, I.A., A. Mantovani, and M. Feldmann, *Macrophage heterogeneity in the context of rheumatoid arthritis.* Nature Reviews Rheumatology, 2016. **12**(8): p. 472-485.

236. Croft, A.P., et al., *Distinct fibroblast subsets drive inflammation and damage in arthritis.* Nature, 2019. **570**(7760): p. 246-251.

237. Blom, A.B., et al., *Crucial role of macrophages in matrix metalloproteinase – mediated cartilage destruction during experimental osteoarthritis: involvement of matrix metalloproteinase 3.* Arthritis & Rheumatism, 2007. **56**(1): p. 147-157.

238. Orr, C., et al., *Synovial tissue research: a state-of-the-art review.* Nature Reviews Rheumatology, 2017. **13**(8): p. 463-475.

239. Huang, Q.-Q., et al., *FLIP: a novel regulator of macrophage differentiation and granulocyte homeostasis.* Blood, 2010. **116**(23): p. 4968-4977.

240. Huang, Q.-Q., et al., *CD11c-mediated deletion of Flip promotes autoreactivity and inflammatory arthritis.* Nature communications, 2015. **6**(1): p. 1-15.

241. Qiu, X., et al., *Reversed graph embedding resolves complex single-cell trajectories.* Nat Methods, 2017. **14**(10): p. 979-982.

242. Hänzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-Seq data.* BMC Bioinformatics, 2013. **14**(1): p. 7.

243. Garcia-Alonso, L., et al., *Benchmark and integration of resources for the estimation of human transcription factor activities.* Genome research, 2019. **29**(8): p. 1363-1375.

244. Mandelin, A.M., 2nd, et al., *Transcriptional Profiling of Synovial Macrophages Using Minimally Invasive Ultrasound-Guided Synovial Biopsies in Rheumatoid Arthritis.* Arthritis Rheumatol, 2018. **70**(6): p. 841-854.

245. Durinck, S., et al., *Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt.* Nature protocols, 2009. **4**(8): p. 1184.

246. Tirosh, I., et al., *Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.* Science, 2016. **352**(6282): p. 189-96.

247. Kay, J. and L. Calabrese, *The role of interleukin-1 in the pathogenesis of rheumatoid*

arthritis. Rheumatology (Oxford), 2004. **43 Suppl 3**: p. iii2-iii9.

248. Lee, K.M., et al., *CCL17 in Inflammation and Pain.* J Immunol, 2020. **205**(1): p. 213-222.

249. Igarashi, H., et al., *A pro-inflammatory role for A20 and ABIN family proteins in human fibroblast-like synoviocytes in rheumatoid arthritis.* Immunol Lett, 2012. **141**(2): p. 246-53.

250. Blecher-Gonen, R., et al., *Single-Cell Analysis of Diverse Pathogen Responses Defines a Molecular Roadmap for Generating Antigen-Specific Immunity.* Cell Syst, 2019. **8**(2): p. 109-121.e6.

251. Liu, Z., et al., *Thrombospondin-1 (TSP1) contributes to the development of vascular inflammation by regulating monocytic cell motility in mouse models of abdominal aortic aneurysm.* Circulation research, 2015. **117**(2): p. 129-141.

252. Potthoff, M.J. and E.N. Olson, *MEF2: a central regulator of diverse developmental programs.* Development, 2007. **134**(23): p. 4131-40.

253. Aude-Garcia, C., et al., *Dual roles for MEF2A and MEF2D during human macrophage terminal differentiation and c-Jun expression.* Biochem J, 2010. **430**(2): p. 237-44.

254. Arnold, M.A., et al., *MEF2C transcription factor controls chondrocyte hypertrophy and bone development.* Dev Cell, 2007. **12**(3): p. 377-89.

255. Tugal, D., X. Liao, and M.K. Jain, *Transcriptional control of macrophage polarization.* Arterioscler Thromb Vasc Biol, 2013. **33**(6): p. 1135-44.

256. Liao, X., et al., *Kruppel-like factor 4 regulates macrophage polarization.* J Clin Invest, 2011. **121**(7): p. 2736-49.

257. Chinenov, Y., et al., *Glucocorticoid receptor coordinates transcription factor-dominated*

regulatory network in macrophages. BMC genomics, 2014. **15**(1): p. 656-656.

258. Zhang, W., et al., *Klf10 inhibits IL-12p40 production in macrophage colony-stimulating factor-induced mouse bone marrow-derived macrophages.* European journal of immunology, 2013. **43**(1): p. 258-269.

259. Liu, Y., et al., *Exosomal KLF3-AS1 from hMSCs promoted cartilage repair and chondrocyte proliferation in osteoarthritis.* Biochem J, 2018. **475**(22): p. 3629-3638.

260. Chintakuntlawar, A.V. and J. Chodosh, *Chemokine CXCL1/KC and its receptor CXCR2 are responsible for neutrophil chemotaxis in adenoviral keratitis.* J Interferon Cytokine Res, 2009. **29**(10): p. 657-66.

261. Remmerie, A., et al., *Osteopontin Expression Identifies a Subset of Recruited Macrophages Distinct from Kupffer Cells in the Fatty Liver.* Immunity, 2020. **53**(3): p. 641-657.e14.

262. Dorrington, M.G. and I.D.C. Fraser, *NF-κB Signaling in Macrophages: Dynamics, Crosstalk, and Signal Integration.* Frontiers in Immunology, 2019. **10**: p. 705.

263. Iwanaszko, M. and M. Kimmel, *NF-κB and IRF pathways: cross-regulation on target genes promoter level.* BMC genomics, 2015. **16**(1): p. 307-307.

264. Christianson, C.A., et al., *K/BxN Serum Transfer Arthritis as a Model of Inflammatory Joint Pain*. 2012, Humana Press. p. 249-260.

265. Wu, D.J. and I.E. Adamopoulos, *Loss of WDFY3 ameliorates severity of serum transfer-induced arthritis independently of autophagy.* Cell Immunol, 2017. **316**: p. 61-69.

266. Martin, P., et al., *Disease severity in K/BxN serum transfer-induced arthritis is not affected by IL-33 deficiency.* Arthritis Research & Therapy, 2013. **15**(1): p. R13.

267. Xue, M., et al., *Endogenous MMP-9 and not MMP-2 promotes rheumatoid synovial*

fibroblast survival, inflammation and cartilage degradation. Rheumatology, 2014. **53**(12): p. 2270-2279.

268. McInnes, I.B. and G. Schett, *Cytokines in the pathogenesis of rheumatoid arthritis.* Nature Reviews Immunology, 2007. **7**(6): p. 429-442.

269. Burrage, P.S., K.S. Mix, and C.E. Brinckerhoff, *Matrix metalloproteinases: role in arthritis.* Front Biosci, 2006. **11**: p. 529-43.

270. Muskardin, T.L.W. and T.B. Niewold, *Type I interferon in rheumatic diseases.* Nat Rev Rheumatol, 2018. **14**(4): p. 214-228.

271. Castañeda-Delgado, J.E., et al., *Type I Interferon Gene Response Is Increased in Early and Established Rheumatoid Arthritis and Correlates with Autoantibody Production.* Frontiers in Immunology, 2017. **8**(285).

272. Cooles, F.A.H., et al., *The interferon gene signature is increased in patients with early treatment-naive rheumatoid arthritis and predicts a poorer response to initial therapy.* J Allergy Clin Immunol, 2018. **141**(1): p. 445-448.e4.

273. Jacobson, L.S., et al., *Cathepsin-mediated necrosis controls the adaptive immune response by Th2 (T helper type 2)-associated adjuvants.* J Biol Chem, 2013. **288**(11): p. 7481-7491.

274. Kemble, S. and A.P. Croft, *Critical Role of Synovial Tissue–Resident Macrophage and Fibroblast Subsets in the Persistence of Joint Inflammation.* Frontiers in Immunology, 2021. **12**(3579).

275. McArdel, S.L., C. Terhorst, and A.H. Sharpe, *Roles of CD48 in regulating immunity and tolerance.* Clin Immunol, 2016. **164**: p. 10-20.

276. McAllister, K., et al., *Identification of BACH2 and RAD51B as rheumatoid arthritis*

susceptibility loci in a meta-analysis of genome-wide data. Arthritis Rheum, 2013. **65**(12): p. 3058-62.

277. Zhou, Y., et al., *The Bach Family of Transcription Factors: A Comprehensive Review.* Clin Rev Allergy Immunol, 2016. **50**(3): p. 345-56.
278. Roberts, C.A., A.K. Dickinson, and L.S. Taams, *The Interplay Between Monocytes/Macrophages and CD4(+) T Cell Subsets in Rheumatoid Arthritis.* Front Immunol, 2015. **6**: p. 571.
279. Watts, G.M., et al., *Manifestations of inflammatory arthritis are critically dependent on LFA-1.* J Immunol, 2005. **174**(6): p. 3668-75.
280. Neumann, E., et al., *Tetraspanin CD82 affects migration, attachment and invasion of rheumatoid arthritis synovial fibroblasts.* Ann Rheum Dis, 2018. **77**(11): p. 1619-1626.
281. Wiese, M.D., A.T. Manning-Bennett, and A.Y. Abuhelwa, *Investigational IRAK-4 inhibitors for the treatment of rheumatoid arthritis.* Expert Opin Investig Drugs, 2020. **29**(5): p. 475-482.
282. Thalhamer, T., M.A. McGrath, and M.M. Harnett, *MAPKs and their relevance to arthritis and inflammation.* Rheumatology, 2008. **47**(4): p. 409-414.
283. Raker, V.K., C. Becker, and K. Steinbrink, *The cAMP Pathway as Therapeutic Target in Autoimmune and Inflammatory Diseases.* Front Immunol, 2016. **7**: p. 123.
284. Huang, Q.Q., et al., *Critical role of synovial tissue-resident macrophage niche in joint homeostasis and suppression of chronic inflammation.* Sci Adv, 2021. **7**(2).
285. Cui, J., et al., *p38 MAPK contributes to CD54 expression and the enhancement of phagocytic activity during macrophage development.* Cell Immunol, 2009. **256**(1-2): p. 6-11.

286. Wilson, N.J., et al., *cAMP inhibits CSF-1-stimulated tyrosine phosphorylation but augments CSF-1R-mediated macrophage differentiation and ERK activation.* Febs j, 2005. **272**(16): p. 4141-52.

287. Liu, L., et al., *Proinflammatory signal suppresses proliferation and shifts macrophage metabolism from Myc-dependent to HIF1α-dependent.* Proc Natl Acad Sci U S A, 2016. **113**(6): p. 1564-9.

288. Zhu, J., J. Blenis, and J. Yuan, *Activation of PI3K/Akt and MAPK pathways regulates Myc-mediated transcription by phosphorylating and promoting the degradation of Mad1.* Proc Natl Acad Sci U S A, 2008. **105**(18): p. 6584-9.

289. Zhang, W. and H.T. Liu, *MAPK signal pathways in the regulation of cell proliferation in mammalian cells.* Cell Research, 2002. **12**(1): p. 9-18.

290. Hutchison, M.R., M.H. Bassett, and P.C. White, *SCF, BDNF, and Gas6 are regulators of growth plate chondrocyte proliferation and differentiation.* Mol Endocrinol, 2010. **24**(1): p. 193-203.

291. Yu, Y.Y., et al., *Bone morphogenetic protein 2 stimulates endochondral ossification by regulating periosteal cell fate during bone repair.* Bone, 2010. **47**(1): p. 65-73.

292. Murray, P.J., et al., *Macrophage activation and polarization: nomenclature and experimental guidelines.* Immunity, 2014. **41**(1): p. 14-20.

293. Mun, S.H., P.S.U. Park, and K.-H. Park-Min, *The M-CSF receptor in osteoclasts and beyond.* Experimental & Molecular Medicine, 2020. **52**(8): p. 1239-1254.

294. Cichoń, M.A., et al., *The receptor tyrosine kinase Axl regulates cell-cell adhesion and stemness in cutaneous squamous cell carcinoma.* Oncogene, 2014. **33**(32): p. 4185-92.

295. Brand, D.D., A.H. Kang, and E.F. Rosloniec, *The mouse model of collagen-induced*

arthritis. Methods Mol Med, 2004. **102**: p. 295-312.

296. Jia, W., et al., *Galectin-3 accelerates M2 macrophage infiltration and angiogenesis in tumors.* Am J Pathol, 2013. **182**(5): p. 1821-31.

297. Cohen, A.W., et al., *Role of caveolin-1 in the modulation of lipolysis and lipid droplet formation.* Diabetes, 2004. **53**(5): p. 1261-70.

298. Cadigan, K.M., J.G. Heider, and T.Y. Chang, *Isolation and characterization of Chinese hamster ovary cell mutants deficient in acyl-coenzyme A:cholesterol acyltransferase activity.* J Biol Chem, 1988. **263**(1): p. 274-82.

299. Kubler, A., et al., *Cathepsin K Contributes to Cavitation and Collagen Turnover in Pulmonary Tuberculosis.* J Infect Dis, 2016. **213**(4): p. 618-27.

300. Zigrino, P., et al., *Fibroblast-Derived MMP-14 Regulates Collagen Homeostasis in Adult Skin.* J Invest Dermatol, 2016. **136**(8): p. 1575-1583.

301. Geiss-Friedlander, R. and F. Melchior, *Concepts in sumoylation: a decade on.* Nature Reviews Molecular Cell Biology, 2007. **8**(12): p. 947-956.

302. Dehnavi, S., et al., *The role of protein SUMOylation in rheumatoid arthritis.* Journal of autoimmunity, 2019. **102**: p. 1-7.

303. Wang, Y.G., et al., *Stimulating lymphotoxin beta receptor on the dendritic cells is critical for their homeostasis and expansion.* J Immunol, 2005. **175**(10): p. 6997-7002.

304. Lagali, N.S., et al., *Dendritic cell maturation in the corneal epithelium with onset of type 2 diabetes is associated with tumor necrosis factor receptor superfamily member 9.* Scientific Reports, 2018. **8**(1): p. 14248.

305. Santiago-Schwarz, F., et al., *Dendritic cells (DCs) in rheumatoid arthritis (RA): progenitor cells and soluble factors contained in RA synovial fluid yield a subset of*

myeloid DCs that preferentially activate Th1 inflammatory-type responses. The Journal of Immunology, 2001. **167**(3): p. 1758-1768.

306. Marzaioli, V., et al., *Monocyte-Derived Dendritic Cell Differentiation in Inflammatory Arthritis Is Regulated by the JAK/STAT Axis via NADPH Oxidase Regulation.* Front Immunol, 2020. **11**: p. 1406.

307. McDonald, L.T., et al., *Increased macrophage-derived SPARC precedes collagen deposition in myocardial fibrosis.* American Journal of Physiology-Heart and Circulatory Physiology, 2018. **315**(1): p. H92-H100.

308. Monach, P.A., D. Mathis, and C. Benoist, *The K/BxN arthritis model.* Curr Protoc Immunol, 2008. **Chapter 15**: p. Unit 15.22.

309. Shi, C. and E.G. Pamer, *Monocyte recruitment during infection and inflammation.* Nature Reviews Immunology, 2011. **11**(11): p. 762-774.

310. Seidman, J.S., et al., *Niche-Specific Reprogramming of Epigenetic Landscapes Drives Myeloid Cell Diversity in Nonalcoholic Steatohepatitis.* Immunity, 2020.

311. Sheng, J., C. Ruedl, and K. Karjalainen, *Most tissue-resident macrophages except microglia are derived from fetal hematopoietic stem cells.* Immunity, 2015. **43**(2): p. 382-393.

312. Bain, C.C., et al., *Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice.* Nature immunology, 2014. **15**(10): p. 929-937.

313. Swirski, F.K., et al., *Identification of splenic reservoir monocytes and their deployment to inflammatory sites.* Science, 2009. **325**(5940): p. 612-616.

314. Watanabe, S., et al., *The role of macrophages in the resolution of inflammation.* The

Journal of clinical investigation, 2019. **129**(7): p. 2619-2628.

315. A-Gonzalez, N., et al., *Phagocytosis imprints heterogeneity in tissue-resident macrophages.* Journal of Experimental Medicine, 2017. **214**(5): p. 1281-1296.

316. Roberts, A.W., et al., *Tissue-resident macrophages are locally programmed for silent clearance of apoptotic cells.* Immunity, 2017. **47**(5): p. 913-927.

317. Robbins, C.S., et al., *Local proliferation dominates lesional macrophage accumulation in atherosclerosis.* Nature medicine, 2013. **19**(9): p. 1166.

318. Perlman, H., et al., *FLICE-inhibitory protein expression during macrophage differentiation confers resistance to fas-mediated apoptosis.* The Journal of experimental medicine, 1999. **190**(11): p. 1679-1688.

319. Hagert, C., et al., *The macrophage mannose receptor regulate mannan-induced psoriasis, psoriatic arthritis, and rheumatoid arthritis-like disease models.* Frontiers in immunology, 2018. **9**: p. 114.

320. Alvarado-Vazquez, P.A., et al., *Macrophage-specific nanotechnology-driven CD163 overexpression in human macrophages results in an M2 phenotype under inflammatory conditions.* Immunobiology, 2017. **222**(8-9): p. 900-912.

321. Perlman, H., et al., *Rheumatoid arthritis synovial macrophages express the Fas-associated death domain–like interleukin-1β–converting enzyme–inhibitory protein and are refractory to Fas-mediated apoptosis.* Arthritis & Rheumatism: Official Journal of the American College of Rheumatology, 2001. **44**(1): p. 21-30.

322. Smiljanovic, B., et al., *Monocyte alterations in rheumatoid arthritis are dominated by preterm release from bone marrow and prominent triggering in the joint.* Annals of the rheumatic diseases, 2018. **77**(2): p. 300-308.

323. Steen, J., et al., *Recognition of amino acid motifs, rather than specific proteins, by human plasma Cell–Derived monoclonal antibodies to posttranslationally modified proteins in rheumatoid arthritis.* Arthritis & Rheumatology, 2019. **71**(2): p. 196-209.

324. Linos, A., et al., *The epidemiology of rheumatoid arthritis in Rochester Minnesota: a study of incidence, prevalence, and mortality.* American journal of epidemiology, 1980. **111**(1): p. 87-98.

325. Ogawa, T., M. Kitagawa, and K. Hirokawa, *Age-related changes of human bone marrow: a histometric estimation of proliferative cells, apoptotic cells, T cells, B cells and macrophages.* Mechanisms of ageing and development, 2000. **117**(1-3): p. 57-68.

326. Wong, C.K., et al., *Aging impairs alveolar macrophage phagocytosis and increases influenza-induced mortality in mice.* The Journal of Immunology, 2017. **199**(3): p. 1060-1068.

327. Paoletti, A., et al., *Monocyte/macrophage abnormalities specific to rheumatoid arthritis are linked to miR-155 and are differentially modulated by different TNF inhibitors.* The Journal of Immunology, 2019. **203**(7): p. 1766-1775.

328. Yona, S. and S. Jung, *Monocytes: subsets, origins, fates and functions.* Curr Opin Hematol, 2010. **17**(1): p. 53-9.

329. Ziegler-Heitbrock, L., et al., *Nomenclature of monocytes and dendritic cells in blood.* Blood, 2010. **116**(16): p. e74-e80.

330. Guilliams, M., A. Mildner, and S. Yona, *Developmental and Functional Heterogeneity of Monocytes.* Immunity, 2018. **49**(4): p. 595-613.

331. Mildner, A., et al., *Genomic Characterization of Murine Monocytes Reveals C/EBPβ Transcription Factor Dependence of Ly6C(-) Cells.* Immunity, 2017. **46**(5): p. 849-

862.e7.

332. Narasimhan, P.B., et al., *Nonclassical monocytes in health and disease.* Annual review of immunology, 2019. **37**: p. 439-456.

333. Schyns, J., et al., *Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung.* Nature Communications, 2019. **10**(1).

334. Solomon, S., et al., *A crucial role for macrophages in the pathology of K/B x N serum-induced arthritis.* Eur J Immunol, 2005. **35**(10): p. 3064-73.

335. Jacobs, J.P., et al., *Deficiency of CXCR2, but not other chemokine receptors, attenuates autoantibody-mediated arthritis in a murine model.* Arthritis Rheum, 2010. **62**(7): p. 1921-32.

336. Puchner, A., et al., *Non-classical monocytes as mediators of tissue destruction in arthritis.* Ann Rheum Dis, 2018. **77**(10): p. 1490-1497.

337. Quinones, M.P., et al., *The complex role of the chemokine receptor CCR2 in collagen-induced arthritis: implications for therapeutic targeting of CCR2 in rheumatoid arthritis.* Journal of molecular medicine, 2005. **83**(9): p. 672-681.

338. Brunet, A., et al., *NR4A1-dependent Ly6C(low) monocytes contribute to reducing joint inflammation in arthritic mice through Treg cells.* Eur J Immunol, 2016. **46**(12): p. 2789-2800.

339. Liebmann, M., et al., *Nur77 serves as a molecular brake of the metabolic switch during T cell activation to restrict autoimmunity.* Proc Natl Acad Sci U S A, 2018. **115**(34): p. E8017-e8026.

340. Kirkling, M.E., et al., *Notch Signaling Facilitates In Vitro Generation of Cross-*

Presenting Classical Dendritic Cells. Cell Rep, 2018. **23**(12): p. 3658-3672.e6.

341. Boulet, S., et al., *The orphan nuclear receptor NR4A3 controls the differentiation of monocyte-derived dendritic cells following microbial stimulation.* Proceedings of the National Academy of Sciences, 2019. **116**(30): p. 15150.

342. Ponichtera, H.E., et al., *CD209a expression on dendritic cells is critical for the development of pathogenic Th17 cell responses in murine schistosomiasis.* J Immunol, 2014. **192**(10): p. 4655-65.

343. Stables, M.J., et al., *Transcriptomic analyses of murine resolution-phase macrophages.* Blood, 2011. **118**(26): p. e192-208.

344. Cheung, A.K., et al., *Cysteine-rich intestinal protein 2 (CRIP2) acts as a repressor of NF-kappaB-mediated proangiogenic cytokine transcription to suppress tumorigenesis and angiogenesis.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8390-5.

345. Mayan, H., Z. Farfel, and S.J.D. Karlish, *Renal Mg handling, FXYD2 and the central role of the Na,K-ATPase.* Physiological reports, 2018. **6**(17): p. e13843-e13843.

346. Li, S., et al., *Ribonuclease 4 protects neuron degeneration by promoting angiogenesis, neurogenesis, and neuronal survival under stress.* Angiogenesis, 2013. **16**(2): p. 387-404.

347. Jakubzick, C., et al., *Minimal differentiation of classical monocytes as they survey steady-state tissues and transport antigen to lymph nodes.* Immunity, 2013. **39**(3): p. 599-610.

348. Monach, P.A., et al., *Neutrophils in a mouse model of autoantibody-mediated arthritis: critical producers of Fc receptor gamma, the receptor for C5a, and lymphocyte function-associated antigen 1.* Arthritis Rheum, 2010. **62**(3): p. 753-64.

349. Lessard, A.J., et al., *Triggering of NOD2 Receptor Converts Inflammatory Ly6C(high) into Ly6C(low) Monocytes with Patrolling Properties.* Cell Rep, 2017. **20**(8): p. 1830-

1843.

350. Bain, C.C., et al., *Long-lived self-renewing bone marrow-derived macrophages displace embryo-derived cells to inhabit adult serous cavities.* Nature Communications, 2016. **7**(1): p. ncomms11852.

351. Ferrer, I.R., et al., *A wave of monocytes is recruited to replenish the long-term Langerhans cell network after immune injury.* Sci Immunol, 2019. **4**(38).

352. Kelly, L.M., et al., *MafB is an inducer of monocytic differentiation.* Embo j, 2000. **19**(9): p. 1987-97.

353. Meredith, M.M., et al., *Expression of the zinc finger transcription factor zDC (Zbtb46, Btbd4) defines the classical dendritic cell lineage.* J Exp Med, 2012. **209**(6): p. 1153-65.

354. Satpathy, A.T., et al., *Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages.* J Exp Med, 2012. **209**(6): p. 1135-52.

355. Atehortúa, L., et al., *Endothelial Alterations in Systemic Lupus Erythematosus and Rheumatoid Arthritis: Potential Effect of Monocyte Interaction.* Mediators Inflamm, 2017. **2017**: p. 9680729.

356. Binstadt, B.A., et al., *Particularities of the vasculature can promote the organ specificity of autoimmune attack.* Nat Immunol, 2006. **7**(3): p. 284-92.

357. Wipke, B.T., et al., *Dynamic visualization of a joint-specific autoimmune response through positron emission tomography.* Nat Immunol, 2002. **3**(4): p. 366-72.

358. Cloutier, N., et al., *Platelets can enhance vascular permeability.* Blood, 2012. **120**(6): p. 1334-43.

359. Niess, J.H., et al., *CX3CR1-mediated dendritic cell access to the intestinal lumen and*

bacterial clearance. Science, 2005. **307**(5707): p. 254-8.

360. Zimmermann, H.W., et al., *Bidirectional transendothelial migration of monocytes across hepatic sinusoidal endothelium shapes monocyte differentiation and regulates the balance between immunity and tolerance in liver.* Hepatology, 2016. **63**(1): p. 233-46.

361. Angelovich, T.A., et al., *Quantification of Monocyte Transmigration and Foam Cell Formation from Individuals with Chronic Inflammatory Conditions.* J Vis Exp, 2017(128).

362. Auffray, C., et al., *Monitoring of blood vessels and tissues by a population of monocytes with patrolling behavior.* Science, 2007. **317**(5838): p. 666-70.

363. Suchard, S.J., et al., *An LFA-1 (alphaLbeta2) small-molecule antagonist reduces inflammation and joint destruction in murine models of arthritis.* J Immunol, 2010. **184**(7): p. 3917-26.

364. Fiorucci, S., et al., *Bile acids activated receptors regulate innate immunity.* Frontiers in immunology, 2018. **9**: p. 1853.

365. Wammers, M., et al., *Reprogramming of pro-inflammatory human macrophages to an anti-inflammatory phenotype by bile acids.* Scientific reports, 2018. **8**(1): p. 1-15.

366. Sato, K., et al., *Pathogenesis of Kupffer cells in cholestatic liver injury.* The American journal of pathology, 2016. **186**(9): p. 2238-2247.

367. Hoeffel, G. and F. Ginhoux, *Fetal monocytes and the origins of tissue-resident macrophages.* Cellular immunology, 2018. **330**: p. 5-15.

368. Tacke, F., *Targeting hepatic macrophages to treat liver diseases.* Journal of hepatology, 2017. **66**(6): p. 1300-1312.

369. Brempelis, K.J. and I.N. Crispe, *Infiltrating monocytes in liver injury and repair.* Clinical

& translational immunology, 2016. **5**(11): p. e113.

370. Rattay, S., et al., *Anti-inflammatory consequences of bile acid accumulation in virus-infected bile duct ligated mice.* Plos one, 2018. **13**(6): p. e0199863.

371. Guicciardi, M.E., et al., *Macrophages contribute to the pathogenesis of sclerosing cholangitis in mice.* Journal of hepatology, 2018. **69**(3): p. 676-686.

372. El Kasmi, K.C., et al., *Toll-like receptor 4–dependent Kupffer cell activation and liver injury in a novel mouse model of parenteral nutrition and intestinal injury.* Hepatology, 2012. **55**(5): p. 1518-1528.

373. El Kasmi, K.C., et al., *Macrophage-derived IL-1β/NF-κB signaling mediates parenteral nutrition-associated cholestasis.* Nature communications, 2018. **9**(1): p. 1-14.

374. Hao, H., et al., *Farnesoid X receptor regulation of the NLRP3 inflammasome underlies cholestasis-associated sepsis.* Cell metabolism, 2017. **25**(4): p. 856-867.

375. Taylor, S.A., et al., *Oligoclonal immunoglobulin repertoire in biliary remnants of biliary atresia.* Scientific reports, 2019. **9**(1): p. 1-11.

376. Kobayashi, H., et al., *Hepatic overexpression of MHC class II antigens and macrophage-associated antigens (CD68) in patients with biliary atresia of poor prognosis.* Journal of pediatric surgery, 1997. **32**(4): p. 590-593.

377. Urushihara, N., et al., *Elevation of serum interleukin-18 levels and activation of Kupffer cells in biliary atresia.* Journal of pediatric surgery, 2000. **35**(3): p. 446-449.

378. Davenport, M., et al., *Immunohistochemistry of the liver and biliary tree in extrahepatic biliary atresia.* Journal of pediatric surgery, 2001. **36**(7): p. 1017-1025.

379. Mack, C.L., et al., *Armed CD4+ Th1 effector cells and activated macrophages participate in bile duct injury in murine biliary atresia.* Clinical Immunology, 2005.

115(2): p. 200-209.

380. Li, J., et al., *Th2 signals induce epithelial injury in mice and are compatible with the biliary atresia phenotype.* The Journal of clinical investigation, 2011. **121**(11).

381. MacParland, S.A., et al., *Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations.* Nature communications, 2018. **9**(1): p. 1-21.

382. Hsiao, C.J., et al., *Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis.* Genome research, 2020. **30**(4): p. 611-621.

383. Pollara, G., et al., *Validation of immune cell modules in multicellular transcriptomic data.* PloS one, 2017. **12**(1): p. e0169271.

384. Villani, A.-C., et al., *Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors.* Science, 2017. **356**(6335).

385. Jaitin, D.A., et al., *Lipid-associated macrophages control metabolic homeostasis in a Trem2-dependent manner.* Cell, 2019. **178**(3): p. 686-698.

386. Shah, R.D., et al., *Expression of calgranulin genes S100A8, S100A9 and S100A12 is modulated by n-3 PUFA during inflammation in adipose tissue and mononuclear cells.* PloS one, 2017. **12**(1): p. e0169614.

387. Masuda, A., et al., *Versican is upregulated in circulating monocytes in patients with systemic sclerosis and amplifies a CCL2-mediated pathogenic loop.* Arthritis research & therapy, 2013. **15**(4): p. 1-15.

388. Sotoodehnejadnematalahi, F., et al., *Mechanisms of hypoxic up-regulation of versican gene expression in macrophages.* PLoS One, 2015. **10**(6): p. e0125799.

389. Dower, K., et al., *Innate immune responses to TREM-1 activation: overlap, divergence, and positive and negative cross-talk with bacterial lipopolysaccharide.* J Immunol, 2008.

180(5): p. 3520-34.

390. Xia, C., et al., *S100 Proteins As an Important Regulator of Macrophage Inflammation.* Front Immunol, 2017. **8**: p. 1908.

391. Ripoll, V.M., et al., *Gpnmb is induced in macrophages by IFN-γ and lipopolysaccharide and acts as a feedback regulator of proinflammatory responses.* The Journal of Immunology, 2007. **178**(10): p. 6557-6566.

392. De, S.K., M.T. McMaster, and G.K. Andrews, *Endotoxin induction of murine metallothionein gene expression.* Journal of Biological Chemistry, 1990. **265**(25): p. 15267-15274.

393. Subramanian Vignesh, K. and G.S. Deepe Jr, *Metallothioneins: emerging modulators in immunity and infection.* International journal of molecular sciences, 2017. **18**(10): p. 2197.

394. Han, Y.-H., et al., *RORα induces KLF4-mediated M2 polarization in the liver macrophages that protect against nonalcoholic steatohepatitis.* Cell reports, 2017. **20**(1): p. 124-135.

395. Nejati Moharrami, N., et al., *RORα controls inflammatory state of human macrophages.* PloS one, 2018. **13**(11): p. e0207374.

396. Turnbull, I.R., et al., *Cutting edge: TREM-2 attenuates macrophage activation.* The Journal of Immunology, 2006. **177**(6): p. 3520-3524.

397. Kilgore, A. and C.L. Mack, *Update on investigations pertaining to the pathogenesis of biliary atresia.* Pediatric surgery international, 2017. **33**(12): p. 1233-1241.

398. Ramachandran, P., et al., *Resolving the fibrotic niche of human liver cirrhosis at single-cell level.* Nature, 2019. **575**(7783): p. 512-518.

399. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.

400. Krämer, A., et al., *Causal analysis approaches in Ingenuity Pathway Analysis.* Bioinformatics, 2013. **30**(4): p. 523-530.

401. Maere, S., K. Heymans, and M. Kuiper, *BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks.* Bioinformatics, 2005. **21**(16): p. 3448-3449.

402. Wynn, T.A., A. Chawla, and J.W. Pollard, *Macrophage biology in development, homeostasis and disease.* Nature, 2013. **496**(7446): p. 445-55.

403. Fonseca, G.J., J.S. Seidman, and C.K. Glass, *Genome-Wide Approaches to Defining Macrophage Identity and Function.* Microbiol Spectr, 2016. **4**(5).

404. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic acids research, 2002. **30**(1): p. 207-210.

405. Paul, F., et al., *Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors.* Cell, 2015. **163**(7): p. 1663-1677.

406. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.

407. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome research, 2003. **13**(11): p. 2498-2504.

408. Mouton, A.J., et al., *Mapping macrophage polarization over the myocardial infarction*

time continuum. Basic research in cardiology, 2018. **113**(4): p. 1-18.

409. Sheffield, N.C. and C. Bock, *LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor.* Bioinformatics, 2016. **32**(4): p. 587-9.

410. Aszódi, A., *MULTOVL: fast multiple overlaps of genomic regions.* Bioinformatics, 2012. **28**(24): p. 3318-3319.

411. Park, P.J., *ChIP–seq: advantages and challenges of a maturing technology.* Nature Reviews Genetics, 2009. **10**(10): p. 669-680.

412. Kelsey, G., O. Stegle, and W. Reik, *Single-cell epigenomics: Recording the past and predicting the future.* Science, 2017. **358**(6359): p. 69.

413. Vickovic, S., et al., *High-definition spatial transcriptomics for in situ tissue profiling.* Nature Methods, 2019. **16**(10): p. 987-990.

414. Abdelaal, T., et al., *A comparison of automatic cell identification methods for single-cell RNA sequencing data.* Genome Biology, 2019. **20**(1): p. 194.

415. Saelens, W., et al., *A comparison of single-cell trajectory inference methods.* Nature Biotechnology, 2019. **37**(5): p. 547-554.

416. Mangul, S., et al., *Systematic benchmarking of omics computational tools.* Nature Communications, 2019. **10**(1): p. 1393.

417. Sarkar, A., Y. Yang, and M. Vihinen, *Variation benchmark datasets: update, criteria, quality and applications.* Database, 2020. **2020**.

418. Subramanian, I., et al., *Multi-omics Data Integration, Interpretation, and Its Application.* Bioinformatics and biology insights, 2020. **14**: p. 1177932219899051-1177932219899051.