Identifying Markers of Differentiation in Monocyte-Derived-Macrophages

by

Matthew Gibson

1482146

Dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

August 2024

Declaration

I, Matthew Gibson (1482146), am a student registered for the degree of Master of Science in the academic year 2024.

I hereby declare the following:

● I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.

● I confirm that the research dissertation submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.

● I have followed the required conventions in referencing the thoughts and ideas of others.

● I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature 

6 August 2024

Acknowledgements

My thanks go to my supervisor, Dr Nikki Gentle, for her guidance. I would further like to thank my friends and family for the endless support and encouragement. To my mom, thank you for the invaluable words of wisdom and unwavering optimism. To my father, thank you for the practical advice and understanding. To my siblings, for being there in my times of need. To my Maxine, who has been my rock when times were tough, and who has been a source of joy and light always.

Finally, my thanks to all the members of the Functional Genomics and Immunogenetics laboratory at WITS, headed by Drs Gentle and Meyer. To my fellow lab mates, especially Darisia, who have been there to share ideas and offer helpful insights.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BP - Biological Processes

CC - Cellular Components

CD - Cluster of Differentiation

DAMP - Damage-Associated Molecular Pattern

DEG - Differentially Expressed Gene

DGE - Differential Gene Expression

DO - Disease Ontology

DTU - Differential Transcript Usage

ECM - ExtraCellular Matrix

FEA - Functional Enrichment Analysis

GLM - Generalised Linear Model

GM-CSF - Granular Macrophage Colony Stimulating Factor

GO - Gene Ontology

HSCs - Haematopoietic Stem Cells

IL - Interleukin

KEGG - Kyoto Encyclopedia of Genes and Genomes

LFC - Log2 Fold Change

M-CSF - Macrophage Colony Stimulating Factor

MDM - Monocyte-Derived Macrophage

MF - Molecular Functions

NB - Negative Binomial

NO - Nitric Oxide

ORA - Over-representation Analysis

PAMP - Pathogen-Associated Molecular Pattern

PBMC - Peripheral Blood Mononuclear Cell

PCA - Principal Components Analysis

PPI - Protein-Protein Interaction

PRR - Pattern Recognition Receptors

QC - Quality Control

RNA - Ribonucleic Acid

RNA-seq - Ribonucleic Acid Sequencing

RTM - Resident Tissue Macrophage

TNF - Tumour Necrosis Factor

VBEM - Variational Bayesian Expectation Maximisation

VLMM - Variable-Length Markov Model

VST - Variance-Stabilising Transformation

Abstract

The importance of monocytes and monocyte-derived macrophages (MDMs) in both adaptive and innate immunity makes their study a topic of interest. Monocytes differentiate into macrophages through transcriptomic alterations, resulting in extensive changes in gene expression. Macrophage colony stimulating factor (M-CSF) and granulocyte-macrophage colony stimulating factor (GM-CSF) are the two primary cytokines that stimulate this differentiation, and are known to cause partial polarisation towards the M2 and M1 macrophage subtypes, respectively. However, the degree to which this polarisation takes place is not well-characterised. Therefore, this study aimed to use a computational approach to identify the differences and similarities in gene expression changes in macrophages induced with M-CSF and GM-CSF. RNA sequencing data for three human donors was obtained through EBI and used to quantify gene expression changes associated with M-CSF or GM-CSF treatment. Differential gene expression analysis was performed to identify the genes that were differentially expressed as a result of either treatment relative to the untreated monocytes. Over-representation analysis was used to determine the biological processes in which the differentially expressed genes (DEGs) were involved. Finally, transcription factors were identified within the lists of DEGs, as well as the genes encoding their known protein-protein interacting partners. Treatment with M-CSF and GM-CSF induced 4 072 and 4 399 DEGs, respectively, 2 734 of which were common. An examination of these DEGs revealed that the resultant macrophages lacked changes in expression of genes commonly associated with the M1 and M2 polarisation states. An investigation of the DEGs involved in myeloid cell differentiation and the regulation of inflammatory response revealed *CCR2*, *IGF1* and *INHBA* to be inversely regulated by the two treatments. Furthermore, nine uniquely differentially expressed transcription factors involved in these biological processes were identified, each of which may be contributing to the lack of complete polarisation following differentiation. These results revealed that M-CSF and GM-CSF-induced macrophages, in the absence of activation, experience highly similar gene expression changes and lack changes in the expression of key polarisation marker genes.

Chapter 1: Literature Review

1.1 The immune system

The human immune system is characterised by two distinct but interlinked responses, namely innate and adaptive immunity (Hamon and Quintin, 2016). The innate immune system serves as the first line of defence against pathogens, whilst adaptive immunity is slower to activate but provides long-term protection against the agents of prior infections. Innate immunity relies on a rapid but non-specific response, whilst simultaneously producing signals to direct a further immune response (Hamon and Quintin, 2016). The adaptive immune system builds on this response in a more specific manner whilst placing a greater emphasis on specific memory via the production of antibodies (Hamon and Quintin, 2016). Without this two-part defence, the body would be unable to successfully fend off microbes that seek to colonise it. The efficacy of these systems is dependent on the diversity of cells, signals and responses that are required in order to combat an ever-growing number of pathogens.

The key feature of the immune system is its ability to sense pathogens and mount an appropriate response (Hamon and Quintin, 2016). The immune system comprises a variety of immune cells, each of which performs a distinct role in combating pathogen spread and maintaining homeostasis. These immune cells originate from a self-renewing population of haematopoietic stem cells (HSCs), via the process of haematopoiesis (Basilico and Göttgens, 2017). HSC reservoirs are maintained within the bone marrow. Haematopoiesis is the differentiation of progenitor cells into the various myeloid and lymphoid cell types that are essential for innate and adaptive immune responses, respectively (Figure 1.1; Wahlster and Daley, 2016).

The varied cell types of the myeloid and lymphoid lineages differ significantly from one-another in their physiology and behaviour. This is a consequence not of underlying genetic differences, but of distinct patterns of gene expression (Andrews and Hemberg, 2018). This reprogramming of gene expression occurs during the process of differentiation, which involves large scale transcriptional changes (Basilico and Göttgens, 2017). Thus, transcriptional differences are the basis for phenotypic

variability (Andrews and Hemberg, 2018). Therefore, distinct cell types can not only be characterised by their phenotypic differences, but also by the differences in their transcriptomes.

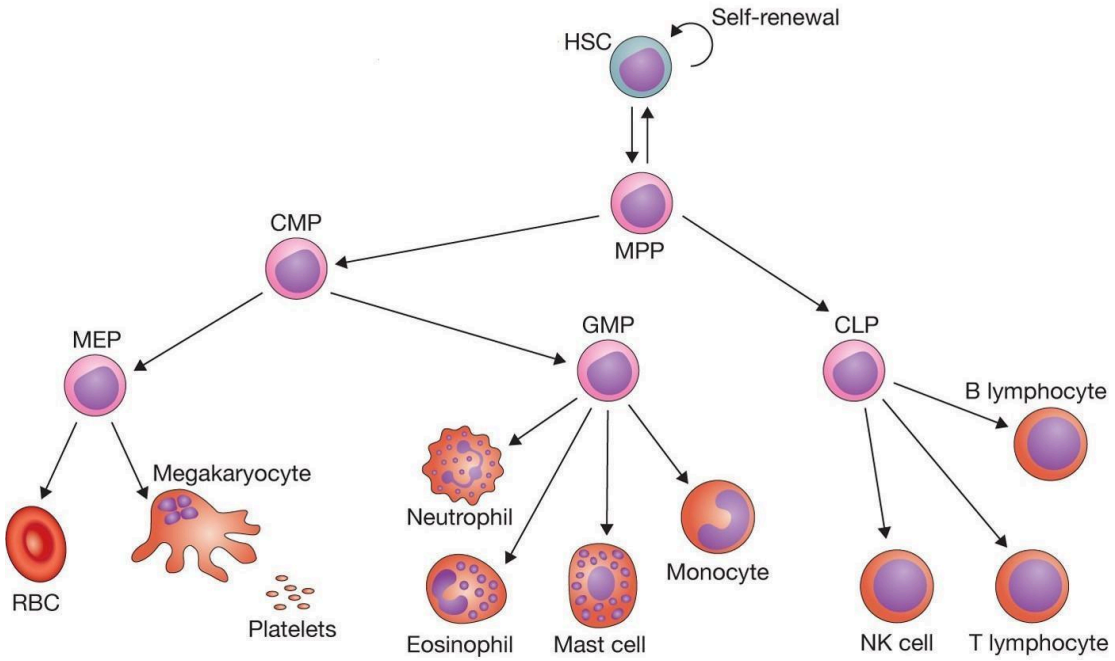

Figure 1.1 Haematopoiesis of various immune cells. An illustration of the various cell types that are able to develop from haematopoietic stem cells, and the relationships between each developmental cell lineage. Adapted from Wahlster & Daley (2016). Abbreviations: CLP – common lymphoid progenitor; CMP – common myeloid progenitor; GMP – granulocyte-macrophage progenitors; HSC – haematopoietic stem cell; MEP – megakaryocyte-erythrocyte progenitor; MPP – multipotent progenitor; NK – natural killer; RBC – red blood cell (Wahlster and Daley, 2016).

The peripheral blood mononuclear cells (PBMCs) that are produced during haematopoiesis are extremely diverse, ranging from the phagocytosing monocytes to the B lymphocytes that mass-produce antibodies (Zhang and Huang, 2012). This diversity is a consequence of alterations in the gene expression of the cell type which distinguishes it from other cell types and enables it to better perform its immune function (Roy, 2019). The concept of a defined 'cell type' is tentative as there is transcriptional diversity between cells that theoretically belong to the same cell type (Zeng, 2022). This heterogeneity results from each cell's unique molecular

environment. However, there are markers that one can use to profile a given cell type (Villani *et al*., 2018). For example, cells capable of antigen presentation (such as monocytes) require the major histocompatibility complex II (MHC II) protein as it is crucial for their immune function (Villani *et al*., 2018). Thus, a given cell type could be described by an aggregate of the genes it expresses that are crucial to its immune function. Furthermore, cells change their expression of genes in response to stimuli released from both invading pathogens and other immune cells. These changes can result in a transition from one cell type to another, such as monocyte-to-macrophage differentiation. Studying gene expression changes which occur within cell types during the differentiation process allows for a better understanding of the immune response. This in turn can inform improved diagnostics and more effective treatment strategies.

Thus, the distinction between innate and adaptive immunity lies in the distinct patterns of gene expression which inform the varied behaviours of immune cells (Chun and Heeger, 2016). Among these many cell types, monocytes and monocyte-derived macrophages (MDMs) are an important point of connection between the innate and adaptive immune systems. These mononuclear phagocytes are capable of important innate functions, such as clearing cellular debris via phagocytosis, while also enabling adaptive immunity via antigen presentation and cytokine signalling (Jakubzick *et al.*, 2017). This study will focus on monocyte differentiation into macrophages and the transcriptional changes associated with this process.

1.2 Monocytes

Monocytes are immune cells of the myeloid lineage that predominantly circulate within the blood (Swirski *et al*., 2009). These circulating monocyte populations are relatively short-lived and are constantly replenished via haematopoiesis of HSCs within the bone marrow (Yona *et al*., 2013). Monocytes participate in the innate immune response and may constitute up to ten percent of total leukocytes within the human body (Swirski *et al*., 2009).

Monocytes serve a key role as monitors of tissue health, with the ability to sense pathological conditions through detection of pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs; Kubes and Mehal, 2012). Upon sensing one or more of these signals, they are capable of extravasation into the affected tissue where they coordinate the early immune response by performing phagocytosis and antigen presentation (Chen *et al*., 2023). Phagocytosis involves the engulfment of large particles and disassembly through lysosome interactions, while antigen presentation involves cycling the constituents of this degraded material to the cell membrane where antigens are mounted on MHC II (Uribe-Querol and Rosales, 2020). However, these functions are far from the full extent of monocyte behaviour.

Monocytes are a diverse set of cells, taking on alternative phenotypes and behaviours in accordance with their molecular environment. Traditionally, monocytes were divided into three subpopulations based on cellular phenotype: classical, non-classical, and intermediate. These subtypes were distinguished by varying levels of both cluster of differentiation (CD) 14 and CD16 expression (Table 1.1; Ziegler-Heitbrock *et al*., 2010). In recent decades, understanding of monocyte heterogeneity has increased, revealing more diversity based on an expanded list of markers (Stansfield and Ingram, 2015).

Table 1.1 Cell surface markers that distinguish monocyte subpopulations. Classical monocytes can be distinguished from non-classical and intermediate monocytes by their lack of CD16 expression. Intermediate and non-classical monocytes both express CD14 and CD16, but with different levels. One plus sign indicates expression of the given gene. Two plus signs indicate a high level of expression of the given gene. A minus sign indicates a lack of expression of the gene.

Monocyte Subpopulation	CD14	CD16
Classical	+ +	-
Intermediate	+ +	+
Non-classical	+	+ +

Circulating classical monocytes survey tissues for areas requiring an immune response. Upon entering tissues, monocytes react in accordance with the combination of signals within their local environment (Kapellos *et al*., 2019). This often involves a transition to an intermediate state, allowing for the monocyte to engage in innate immunity through the production of tumour necrosis factor (TNF) and nitric oxide (NO) (Shi and Pamer, 2011). Other transcriptional changes may also occur such as increased interleukin (IL)-1β, MHC class II, CD206, CD11c and CCR7 expression (Jakubzick *et al*., 2017). Once the pathology is resolved, a further transition to non-classical behaviour may occur, allowing for a return to homeostasis (Narasimhan *et al*., 2019).

Although classical monocytes are known to alter their gene expression without transitioning into another cell type, differentiation provides a path to more extensive transcriptomic changes. Classical monocytes, or those expressing higher levels of CD14, more readily differentiate into macrophages or dendritic cells when exposed to specific molecular signals (González de la Aleja *et al*., 2022). Signals such as macrophage colony stimulating factor (M-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF) promote monocyte differentiation into macrophages (Hamilton, 2008).

1.3 Macrophages

Macrophages were originally described by Metchnikoff, who predicted their role in host-defence as phagocytosing cells (Metchnikoff, 1891). Macrophages have long been known to be important members of the mononuclear-phagocyte system, but their role as mediators of metabolism and homeostasis has become evident in recent years (Biswas and Mantovani, 2012). Macrophage cells arise from two distinct developmental pathways, being either embryonically-derived resident tissue macrophages (RTMs) or bone marrow-derived MDMs (Gomez Perdiguero *et al*., 2015; Park *et al*., 2022). MDMs develop from monocytes through the process of differentiation, becoming longer-lived and leaving the blood to become tissue-resident (Fujiwara and Kobayashi, 2005). The distinction between these groups lies in the lineage of the progenitor monocytes as either prenatal or postnatal, as well as their behaviour.

RTMs specialise for their local environment, where they play an important role in maintaining homeostasis (Davies *et al*., 2013). These macrophages are often crucial for regulating metabolism, maintaining the extracellular matrix (ECM), clearance of cellular debris, production of growth factors, angiogenesis, and immunomodulation (Davies *et al*., 2013). In addition, these macrophages are typically the first cells to sense pathologies and signal other innate and adaptive immune cells, recruiting them to the tissue and mounting an immune response (Chen *et al*., 2023). In this way, RTMs are often responsible for the recruitment of monocytes and MDMs into tissues experiencing pathological conditions.

Much like monocytes, MDMs display transcriptional and phenotypic diversity and are specifically recruited to resolve pathological conditions within tissues (Eligini *et al*., 2013). They achieve this resolution through phagocytosis, antigen presentation, regulation of inflammation, and cytokine signalling. Furthermore, their behaviour is dictated by the stimuli they receive in the local environment, allowing them to adapt and fulfil their required niche within tissues (Lambrecht, 2006). For example, in bacterial infections, macrophage activation and proinflammatory behaviours depend on the species of invading bacteria (Benoit *et al*., 2008).

Regulating the inflammatory response is one of the defining characteristics of macrophages (Chen *et al*., 2023). Inflammation is a state within a tissue characterised by heightened levels of immune activity in response to pathological conditions (Chen *et al*., 2017). RTMs survey their environment through cell surface receptors. Upon PAMP and/or DAMP detection, pattern recognition receptor (PRR) signalling upregulates the expression of genes involved in the inflammatory response, causing macrophages to release inflammatory cytokines and chemokines (Takeuchi and Akira, 2010). These conditions recruit other innate and adaptive immune cells to assist in resolving the pathology.

The activity of macrophages under pathological conditions typically follows a progression from early proinflammatory macrophages to more anti-inflammatory populations later in the process (Fujiwara and Kobayashi, 2005; Goerdt *et al*., 1999). The immune response can be characterised as a progression through the following phases: homeostasis and surveillance, pathological detection, initiation of inflammation,

and the restoration of homeostasis (Austermann *et al*., 2022). Thus, macrophages are said to initiate, regulate, and resolve the inflammatory response.

Macrophages have traditionally been broadly categorised into the M1 and M2 subtypes, where M1 macrophages exhibit more pro-inflammatory behaviour and M2 macrophages exhibit more anti-inflammatory behaviour (Mills *et al*., 2000). Over the course of the inflammatory response, there is typically a shift from M1 to M2 behaviour (Sica et al., 2015). The M1 and M2 classifications are based on a set of molecular markers expressed by macrophages, typically directly associated with the pro- or anti-inflammatory actions themself (Chen *et al*., 2023). The genes encoding a selection of key molecular markers are listed in Table 1.2. This polarisation and activation of pro- or anti-inflammatory behaviour occurs, in a similar manner to monocytes, through PRR detection of PAMPs and DAMPs (Medzhitov and Janeway, 2000).

According to the canonical understanding of M1 and M2 macrophages, activation occurs through specific signals. M1 macrophages are activated by lipopolysaccharide (LPS), Interferon-γ (IFN-γ), and tumour necrosis factor-alpha (TNF-α; Chen *et al*., 2023). Detection of these compounds causes macrophages to increase the production of proinflammatory markers such as IL-1β, IL-6, and inducible nitric oxide synthase (iNOS) via transcription factor activity such as nuclear factor of kappa light polypeptide gene enhancer in B‑cells (NF-κB) signalling (Chen *et al*., 2023). The expression of these subtype markers has been shown to be time-dependent, typically having optimal detection in the range of 4 - 72 hours following stimulation (Unuvar Purcu et al., 2022).

M2 macrophages are typically activated by IL-4, IL-13, and glucocorticoids (Doyle *et al*., 1994; Ehrchen *et al*., 2007; Stein *et al*., 1992). Detection of these signals induces STAT6 signalling, promoting the expression of genes associated with tissue repair and anti-inflammatory responses (Gong *et a*l., 2017).

Table 1.2 Cell surface markers that distinguish macrophage subpopulations.

Macrophage Subtype	Gene Symbol	Gene Name	Reference
M1	*CD80*	CD80 Molecule	(Bertani *et al*., 2017)
	CD86	CD86 Molecule	(Bertani *et al*., 2017)
	CSF2	Colony Stimulating Factor 2	(Orecchioni *et al.*, 2019)
	CXCL2	C-X-C Motif Chemokine Ligand 2	(Xuan *et al*., 2015)
	IL1B	Interleukin 1 Beta	(Shapouri-Moghaddam *et al*., 2018)
	IL1R1	Interleukin 1 Receptor Type 1	(Tugal *et al*., 2013)
	IL6	Interleukin 6	(Ross *et al.*, 2021)
	IL23A	Interleukin 23 Subunit Alpha	(Orecchioni *et al*., 2019)
	IRF5	Interferon regulatory factor 5	(Krausgruber *et al*., 2011)
	NFKB1	Nuclear Factor Kappa B Subunit 1	(Mussbacher *et al*., 2023)
	NFKB2	Nuclear Factor Kappa B Subunit 2	(Mussbacher *et al*., 2023)
	NOS2	Nitric Oxide Synthase 2	(Bertani *et al*., 2017)
	PELI1	Pellino E3 Ubiquitin Protein Ligase 1	(Donghyun Kim *et al.*, 2019)
	STAT1	Signal Transducer and Activator of Transcription 1	(Wang *et al*., 2014)
	TNF	Tumour Necrosis Factor	(Bertani *et al.*, 2017; Kratochvill *et al.*, 2015)
M2	*ARG1*	Arginase 1	(Lisi *et al*., 2017)
	CCL22	C-C Motif Chemokine Ligand 22	(Wang *et al.*, 2022)
	CD163	CD163 Molecule	(Ambarus *et al.*, 2012)
	CD209	CD209 Molecule	(Gudgeon *et al*., 2022)
	CLEC10A	C-Type Lectin Domain Containing 10A	(Tang *et al*., 2022)
	FOLR2	Folate Receptor Beta	(Puig-Kröger *et al*., 2009)
	IGF1	Insulin-like Growth Factor 1	(Spadaro *et al*., 2017)
	IL4R	Interleukin 4 Receptor	(Vannella *et al*., 2014)
	IL10	Interleukin 10	(Lurier *et al*., 2017)
	IRF4	Interferon regulatory factor 4	(Günthner and Anders, 2013)
	MSR1	Macrophage Scavenger Receptor 1	(Gudgeon *et al.*, 2022)
	PDCD1LG2	Programmed Cell Death 1 Ligand 2	(Li *et al*., 2022)
	PPARD	Peroxisome Proliferator Activated Receptor Delta	(Chawla, 2010)
	PPARG	Peroxisome Proliferator Activated Receptor Gamma	(Bouhlel *et al*., 2007; Chawla, 2010)
	STAT6	Signal Transducer and Activator of Transcription 6	(Gong *et al*., 2017)

Of the genes listed in Table 1.2, some are more crucial to pro- or anti-inflammatory activity than others. For example, *NOS2* encodes the iNOS enzyme, which allows for

the production of antimicrobial reactive oxygen species (ROS; Bertani *et al*., 2017). Conversely, expression of the macrophage scavenger receptor type 1 (*MSR1*) and arginase 1 (*ARG1*) genes is integral to anti-inflammatory activity of macrophages (Azad *et al.*, 2014; Pesce *et al*., 2009; Yi *et al*., 2009).

Although there are key subtype markers, there is diversity within both M1 and M2 macrophage gene expression. As a result, attempts have been made to refine the M1/M2 paradigm with sub-classifications. The M2a, M2b, M2c, and M2d sub-classifications have each been assigned their own particular molecular signatures (Austermann et al., 2022; Ross et al., 2021). However, transcriptomics, and more recently single cell ribonucleic acid (RNA) sequencing, have challenged this approach to macrophage classification (Strizova *et al*., 2023). It is now widely accepted that the strict M1/M2 pro/anti-inflammatory paradigm is an oversimplification of macrophage diversity. In addition, macrophages exhibit plasticity, which allows shifting between the M1 and M2 subtypes based on the local microenvironment (Yunna *et al*., 2020). Thus, macrophage behaviour has been revealed to be a spectrum based on the complex signalling of the local environment (Lambrecht, 2006). It is therefore clear that the complexities of macrophage behaviour cannot be so easily defined based on a small selection of genes, but rather the sum of many different genes which are involved in inflammation.

1.4 Monocyte-to-macrophage differentiation

Both monocytes and macrophages exhibit a high degree of diversity and plasticity. This is particularly evident in the pro- and anti-inflammatory behaviours of different monocyte and macrophage subsets, as well as their ability to switch between behaviours based on environmental signals (Arnold *et al*., 2007; Orekhov *et al*., 2019). This can affect the differentiation process, as classical monocytes undergo the transition into macrophages more readily than non-classical monocytes (Narasimhan *et al.*, 2019).

M-CSF and GM-CSF are the main cytokines responsible for monocyte-to-macrophage differentiation (Hamilton, 2008). M-CSF is encoded by the colony stimulating factor 1 (*CSF1*) gene. *CSF1* is constitutively expressed as a growth factor throughout the body,

enabling the proliferation of both monocytes and macrophages (Hume *et al*., 1988). M-CSF binds to the CSF1 receptor (CSF1R), also known as CD115, initiating a signalling cascade ending in the nuclear localisation of the Sp1 transcription factor (Pixley and Stanley, 2004). GM-CSF is encoded by the colony stimulating factor 2 (*CSF2*) gene, and binds CSF2R. This binding leads to the stimulation of STAT5 and the Erk/Akt-dependent pathway, and nuclear translocation of the transcription factors IRF5 and NF-κB (Krausgruber *et al*., 2011)

Macrophages induced by M-CSF and GM-CSF exhibit distinct behaviours, resembling M2 and M1 macrophages, respectively (Fleetwood *et al*., 2007; Jaguin *et al*., 2013). This behaviour is exhibited without activation by polarising cytokines such as IFN-γ or IL-4. For example, IRF5 and NF-κB transcription factor activity leads to increased expression of IL-1β, IL-6, and TNF-α, and thus an M1-like phenotype (Lehtonen *et al*., 2007).

Despite the distinct ways in which M-CSF and GM-CSF shape the macrophage transcriptome, there are common genes which are expressed by macrophages irrespective of the stimuli that caused them to differentiate (Table 1.3). Expression changes in the genes encoding CD16, CD163 (Ambarus *et al*., 2012), CD68 (Chistiakov *et al*., 2017), CD115 (Bonifer and Hume, 2008; Stanley and Chitu, 2014), CD169 (York *et al*., 2007), CD206 (Stahl and Ezekowitz, 1998), and Osteopontin (Zhang *et al*., 2017) occur as a result of differentiation. These markers will therefore be more prevalent in macrophages relative to monocytes, making them signatures of differentiated macrophages.

Table 1.3 Common genetic markers of differentiated macrophages.

Gene	Gene product	Reference
CD68	CD68	(Chistiakov *et al.*, 2017)
CD163	CD163	(Ambarus *et al*., 2012)
CSF1R	CD115	(Bonifer and Hume, 2008; Stanley and Chitu, 2014)
FCGR3A	CD16	(Ambarus *et al*., 2012)
MRC1	CD206	(Stahl and Ezekowitz, 1998)
SIGLEC1	CD169	(York *et al.*, 2007)
SPP1	Osteopontin	(Zhang *et al*., 2017)

Thus, it is unclear where the boundary between differentiation and polarisation lies. This is complicated by the natural plasticity of macrophages, which can be seen when macrophages differentiated using M-CSF readily polarise into both M1 and M2 states when exposed to appropriate activating compounds. One proposed model which deals with this is the 'M0' paradigm, which suggests that macrophages generated through M-CSF display an inactive state and an incomplete M2-like polarisation (Chistiakov *et al*., 2015). Although both M-CSF and GM-CSF progress polarisation to a certain degree, the resultant macrophages do not represent the final endpoints of polarisation (Ushach and Zlotnik, 2016).

The complexities described above illustrate why monocyte-to-macrophage differentiation is an interesting point of study, especially in the context of the ways in which M-CSF and GM-CSF affect macrophage behaviour. Cell lines, murine models, and primary monocytes are the three primary means by which this topic is researched. Murine models offer a close *in vivo* analogue to the human immune system, but those differences that exist limit the value of the conclusions which are drawn. Multiple cell lines exist as models, including HL-60, U937, and THP-1. However, recent transcriptomic analyses have shown the inconsistencies in macrophage polarisation using these models (Rynikova *et al*., 2023). Thus, primary monocytes, being those

extracted from PBMC populations, represent the most promising vehicle for the research of monocyte-to-macrophage transition.

1.5 RNA sequencing and transcriptomics

The field of transcriptomics has developed rapidly in recent years, with advancements in RNA sequencing platforms and approaches (Deshpande *et al*., 2023). The development of ribonucleic acid sequencing (RNA-seq), in conjunction with next generation sequencing platforms, enabled the streamlined study of whole transcriptomes (Stark *et al*., 2019). RNA-seq assays require library preparation, referring to the process by which RNA molecules can be transformed into read data. Library preparation involves the extraction of RNA from cells, removal of non-mRNA RNAs, fragmentation of the mRNA molecules, reverse transcription to produce cDNA, adapter ligation, and the sequencing of these molecules (Deshpande *et al*., 2023). Gene expression estimates can then be generated from this read data, allowing for downstream analyses such as differential gene expression (DGE) analysis (Oshlack *et al*., 2010). DGE analysis is useful in examining differences in gene expression levels between two conditions, such as monocytes and MDMs. Given the scale of the data produced by RNA-seq experiments, a variety of packages and techniques have been developed to address the unique challenges posed by the data (Conesa *et al*., 2016; Stark *et al*., 2019).

1.6 Study aim and objectives

The abundance of monocytes and MDMs and their important role in both adaptive and innate immunity makes their study a topic of great interest. In particular, the process of differentiation, whereby large-scale transcriptional alterations occur, provides an important point of study. Studying gene expression changes that occur within cell types during the differentiation process allows for a better understanding of the immune response and consequently, better diagnostics, which in turn can inform more effective

treatment strategies. M-CSF and GM-CSF are known to cause differentiation in monocytes, as well as a certain degree of polarisation towards the M1 and M2 macrophage subtypes, respectively. However, the degree to which this polarisation occurs is not fully understood.

Therefore, the aim of this study was to use a computational approach to identify differences and similarities within changes in macrophage gene expression resulting from treatment with M-CSF and GM-CSF.

This aim was achieved through the following objectives:

1. To perform transcript quantification on RNA-seq data obtained from untreated, M-CSF-treated, and GM-CSF-treated primary monocytes using pseudoalignment.

2. To perform differential gene expression analysis between untreated and M-CSF-treated monocytes and between untreated and GM-CSF-treated monocytes.

3. To identify and characterise genes uniquely differentially expressed and commonly differentially expressed in monocytes treated with M-CSF and GM-CSF.

Chapter 2: Methodology

2.1 Study Design

In order to identify markers of differentiation in monocyte-derived-macrophages, a bioinformatic data analysis pipeline was employed (Figure 2.1). This data pipeline utilised bulk RNA-seq data in the form of fastq files to construct a count matrix and perform DGE analysis. The means by which this data was generated and acquired are discussed further in section 2.2. Genes that were significantly differentially expressed were subjected to functional enrichment analysis to infer biological relevance. Quality control checks were incorporated into the data pipeline to ensure the accuracy of analyses. All code used to generate the findings in this study can be found at https://github.com/MatthewGibs/gibson_masters.

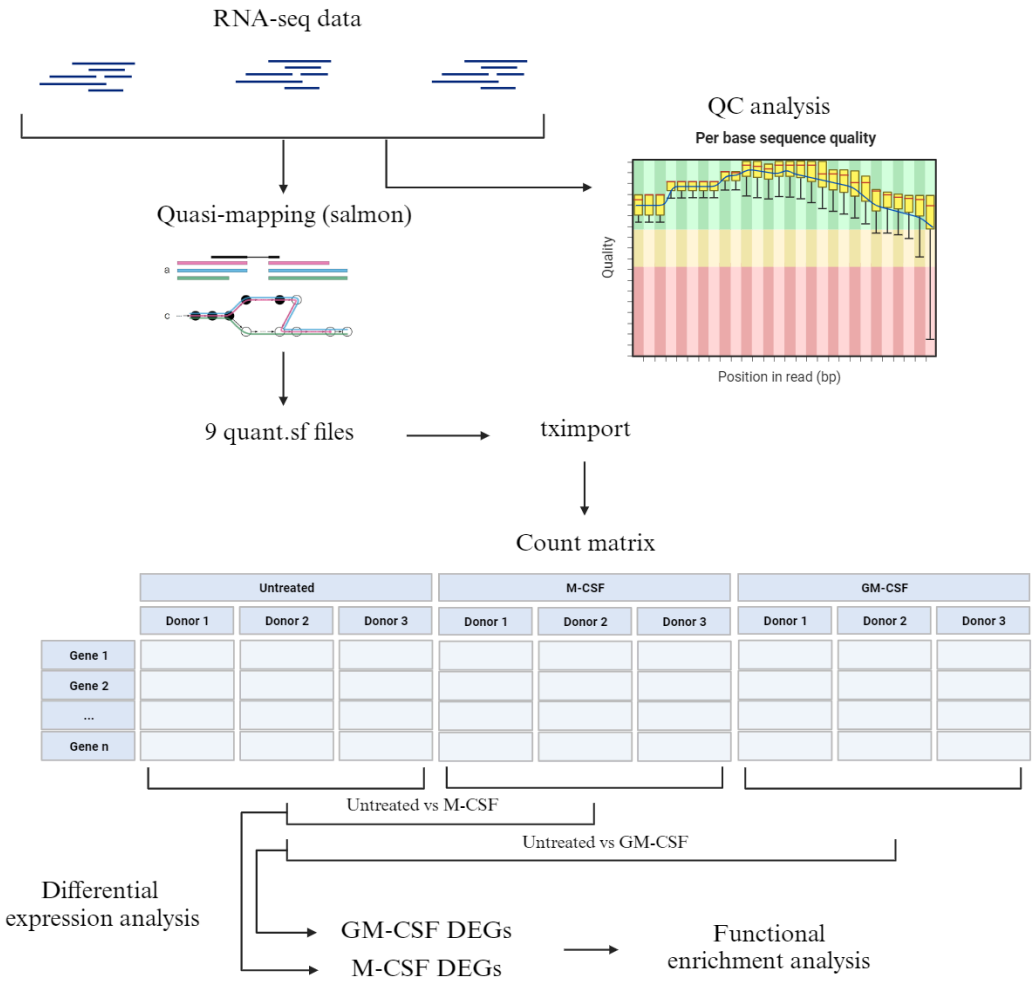

Figure 2.1 Diagrammatic representation of the data processing pipeline used in this study. The RNA sequencing data was subjected to quality control (QC) analysis using FastQC v0.11.9 (Andrews, 2010). Pseudoalignment was performed using salmon v1.9.0 (Patro *et al.*, 2017) was performed, after which a count matrix was constructed for each treatment condition with each row representing a gene and each column representing a unique donor (n = 3). This count matrix was then filtered and used to perform differential gene expression analysis. Image obtained from www.biorender.com.

2.2 Data Acquisition

The fastq files upon which all subsequent analyses were conducted are publicly accessible through EBI (EBI accession PRJNA778191). A study conducted by González de la Aleja *et al.* (2022) produced the aforementioned files as part of their research into the activation of liver X receptors and their role in monocyte-to-macrophage differentiation and inflammation. Briefly, PBMCs were isolated from buffy coats from three healthy human donors (Figure 2.2) and monocytes were subsequently purified from the PBMCs via anti-CD14 magnetic cell microbeads (Miltenyi Biotec), with > 95% $CD14^+$ cells. These cells were then cultured at 0.5 x 10^6 cells/ml in RPMI 1640 medium (Gibco) supplemented with 10% fetal bovine serum (Biowest) in the presence of either M-CSF (10 ng/ml) or GM-CSF (1000 U/ml; ImmunoTools) for seven days.

Following the seven-day cell culture, RNA extraction, library preparation and sequencing were performed. Total RNA extraction was performed via the NucleoSpin RNA kit (Macherey-Nagel, Düren, Germany). A qualitative analysis of total RNA was performed using a Bioanalyzer Agilent 2100, after which RNA libraries were prepared for sequencing using the Takara Bio SMARTer Stranded Total RNA-Seq - Pico Input Mammalian protocol. Finally, bulk sequencing was performed using an Illumina HiSeq 4000 platform, which produced paired-end reads of 151 base pairs (bp) in length.

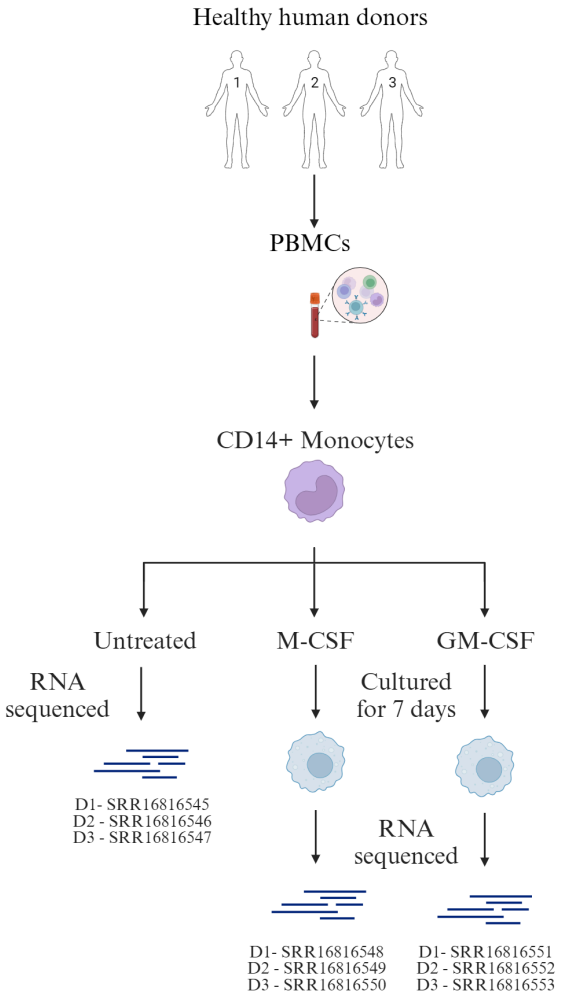

Figure 2.2 Diagrammatic representation of cell culture procedure used by González de la Aleja *et al.* (2022). Image obtained from www.biorender.com.

2.3 Sequencing Quality Control

FastQC v0.11.9 (Andrews, 2010) was used to evaluate key quality control metrics of the raw RNA-seq read data. FastQC is a widely used tool that conducts various quality control tests by leveraging metrics stored in fastq files. Specifically, FastQC evaluated each of the metrics in Table 2.1, assigning either a pass, warning, or fail for each file for each metric. MultiQC v1.10.1 (Ewels *et al.*, 2016) was used to compile all FastQC reports and visualise trends across the nine pairs of fastq files.

Table 2.1 Quality metrics assessed by FastQC.

FastQC metric	Description	Pass threshold
Per base quality score	Examines the Phred scores for each base averaged across every read in a file.	Median Phred score ≥ 25 (99.68% accuracy).
Per tile sequence quality	Measures each tile's average read quality in relation to the average quality scores across all tiles.	No base with a mean Phred score lower than that base across all tiles.
Per sequence quality	Examines the average Phred scores for each read, indicating the overall read quality within a file.	Mean Phred score ≥ 27 (99.80% accuracy).
Per base sequence content	Examines the average ratio of nucleotide bases present at each base across all reads.	Percentage of A and T or C and G differ by < 10% for any given base.
Per sequence GC content	Models the expected GC content across reads along a normal distribution and detects deviations from the model.	Less than 15% of reads deviate from the distribution.
Per base N content	Indicates the percentage of ambiguous base calls (N) by the sequencing platform, averaged across each read for each base.	Less than 5% N across all bases.
Sequence length distribution	Indicates the distribution of sequence length within each file, allowing one to detect deviations from expected read lengths.	All sequences are the same length.
Sequence duplication	Indicates the percentage of sequences which appear multiple times, as well as the distribution of the number of duplicates relative to the percentage and distribution of unduplicated sequences.	> 80% unique sequences.
Overrepresented sequences	Detects and reports specific sequences within the first one million reads of each file that appear in a significant percentage of reads.	No single sequence represents > 0.1% of total sequences.
Adapter content	Determines the percentage of reads containing specific adapter sequences by following a similar approach to the overrepresented sequences metric.	No adapter sequence found in > 5% of reads.

2.4 Transcript quantification

Transcript quantification is the process by which reads or fragments present in fastq files are assigned to their transcript of origin and counted. Numerous approaches exist for performing quantification and these can be broadly grouped into alignment-based and pseudoalignment-based approaches. Alignment-based approaches were developed first and involve a base-by-base alignment step, followed by a separate quantification based on the alignment (Munster *et al.*, 2023). Tools such as Bowtie2 (Langmead and Salzberg, 2012) , HISAT2 (Daehwan Kim *et al*., 2019), and STAR (Dobin *et al*., 2013) perform the alignment, while RSEM (Li and Dewey, 2011), HTSeq-count (Anders *et al*., 2014), or featureCounts (Liao *et al*., 2014) enable quantification based on the alignments. Pseudoalignment packages such as salmon (Patro *et al*., 2017) and kallisto (Bray *et al*., 2016) streamline this process by performing mapping of reads and quantification simultaneously.

Pseudoalignment improves computational efficiency by forgoing individual base-base alignments used by alignment-based approaches (Patro *et al*., 2017). As quantification does not require the degree of detail produced by traditional alignment, the only concern is accuracy (Patro *et al*., 2017). Recent benchmarking of the aforementioned tools showed pseudoalignment approaches to vastly outperform traditional alignment approaches in speed while achieving similar, if not greater accuracy in downstream DGE testing (Munster *et al*., 2023). Therefore, quantification of fragments was performed using salmon v1.9.0 (Patro *et al*., 2017), due to its balance of speed, accuracy, and lightweight data footprint.

Salmon utilises three stages to facilitate its quasi-mapping approach to estimate relative transcript abundance, namely; lightweight-mapping model (index) assembly, online estimation of expression levels and model parameters, and offline refinement of the model and final estimates (Patro *et al*., 2017). Salmon requires an index of all known transcripts against which fragments from the read files may be compared. An index was generated for the human reference genome (GRCh38.p14) using the GENCODE v44 annotations of the transcriptome. The index took the form of a transcriptome de Bruijn

graph, constructed from 31 base-pair long k-mers. K-mer length has an effect on the selectivity of the quantification, as smaller k-mers are more likely to produce mappings due to reduced selectivity (Patro *et al*., 2017). A k-mer length of 31 was selected for this study, as reads longer than 75 bp (as is the case here) typically map well under those conditions.

Once the index was constructed, salmon was used to perform quantification. Firstly, an online phase allows salmon to build equivalence classes for each transcript, which are a collapsed estimate of expected abundance for each fragment produced by the sequencing experiment, generated using a variational Bayesian expectation maximisation (VBEM) algorithm (Bishop, 2006).

However, generation of the equivalence classes depends on the nucleotide fraction for each fragment in the sample. The nucleotide fraction is the proportion of bases for a given fragment relative to the total number of bases in the sample. Thus, the nucleotide fraction represents an estimate of each fragment's relative abundance assuming random sampling during sequencing. As random sampling in a realistic sequencing experiment is unlikely, the nucleotide fractions are transformed using 'effective transcript lengths'. The effective transcript length parameter incorporates differences in transcript length across and within features (isoforms), as well as various bias models aimed at correcting for differences in the likelihood of sampling a given fragment given its unique properties. These properties include numerous biases to which RNA-seq data is prone ('t Hoen *et al.*, 2013). Once the bias models are incorporated, they are applied to transform the nucleotide fraction, producing 'rich' equivalence classes.

The final abundance estimates for each feature are refined in a subsequent offline stage of salmon. Each rich equivalence class represents a grouping of similarly mapped fragments and an average conditional probability across said fragments. Certain fragments may be ambiguous, potentially belonging to multiple transcripts. Hence, this probability represents the likelihood that the fragments belonging to the equivalence class have been correctly assigned to their transcript of origin. Salmon utilises the VBEM algorithm to optimise the abundance estimates, thereby producing a final abundance estimate for each feature of the indexed transcriptome. The process of quantification is complicated by biases inherent in the RNA-seq library preparation

procedure and the sequencing process itself. Examples include positional biases in coverage, sequence-specific biases at the 5′ and 3′ end of sequenced fragments, fragment-level GC bias, and the fragment length distribution (’t Hoen *et al.*, 2013).

Fragment GC content has been shown to influence fragment capture during sequencing (Love *et al*., 2016). FastQC analysis showed potential GC bias in some of the samples, therefore the ‘gcBias’ flag was employed. Specifically, salmon learns a foreground (online) and background (offline) model of fragment GC bias by observing the distribution of GC content across fragments. Importantly, the ‘seqBias’ flag was also used. GC bias and sequence-specific bias have been shown to not be independent of one-another, and therefore supplying both flags alters salmon's approach for correcting for such biases (Conesa *et al.*, 2016). Instead of a single model, salmon learns three models, namely the sequence start, sequence end, and overall fragment GC composition. These bias models are then applied to the effective fragment length parameter, which in turn influences final abundance estimates.

The ‘seqBias’ flag models and corrects for potential sampling bias introduced during library preparation, while the ‘posBias’ flag aims to correct for biases within the coverage of sequencing due to transcript length. Certain nucleotide motifs can undergo preferential sequencing through the use of random hexamers (Roberts *et al.*, 2011). Salmon models this bias through a variable-length Markov Model (VLMM), sampling the first 1 000 000 reads in the input. This approach is similar to that employed by Roberts *et al*., (2011) and allows for correction of sequence bias at the 5’ and 3’ ends of fragments.

Lastly, the ‘validateMappings’ flag was used, indicating that salmon should perform selective alignment. Selective alignment increases the sensitivity of mapping by employing a similar chaining algorithm to minimap2 (Li, 2018). This can reduce speed, but can noticeably increase accuracy, as potential mapping loci are more heavily scrutinised. Therefore, the following command was used for each of the nine file pairs, with file_x_1 and file_x_2 representing the pairs of files for a given donor and treatment combination. Library type was set to ‘A’, allowing salmon to automatically determine the properties of the read files. Four optional flags were used, namely ‘gcBias’, ‘seqBias’, ‘posBias’, and ‘validateMappings’.

```
salmon quant
        --libType A
        -1 fastq_x_1.fastq.gz
        -2 fastq_x_2.fastq.gz
        -i index_file
        -o file_x_output
        --gcBias
        --seqBias
        --posBias
        --validateMappings
```

2.5 Quantification of gene expression

Following transcript quantification, a gene-level count matrix was constructed using tximport v1.26.1 (Soneson *et al.*, 2016). Tximport was selected due to its robust integration with downstream elements of the pipeline, implemented using R v4.2.3. The transcript-level count estimates produced by salmon were converted to gene-level estimates via a transcript-to-gene file. This transcript-to-gene file was constructed using the GENCODE v44 annotations of the human reference genome (GRCh38.p14). Soneson *et al.*, (2016) previously demonstrated gene-level estimates to be more accurate for use in downstream DGE analysis, as collapsing reads to the gene-level reduces the number of features which must be tested. In the case of this study, differential transcript usage (DTU) is not of interest, hence the use of gene-level estimates is preferable. This had the technical benefit of reducing strain on the DESeq2 algorithm while potentially lowering the multiple testing penalty which may have resulted from increased features (Soneson *et al.*, 2016).

One challenge tximport is designed to address is providing the necessary information for downstream normalisation of count estimates. Transcript length has been shown to influence the ability for said transcript to be captured during sequencing (Oshlack and Wakefield, 2009). Therefore, raw count estimates are an inaccurate measure of underlying expression due to variations in transcript length across features (Trapnell *et al.*, 2013). Tximport retains information regarding average transcript length for each feature within the count matrix, thereby enabling effective normalisation downstream.

The collapsing of transcript-level to gene-level estimates introduces an issue regarding DTU across samples. Just as the treatments may affect transcript number, they may also affect the balance of isoform usage. Different isoforms of a single gene may vary transcript length, resulting in the effective average lengths of certain transcripts being substantially different across conditions (Wang *et al.*, 2010). However, another advantage of tximport is its ability to capture average transcript length on a per-sample basis and provide this information alongside the count matrix. This 'gene-level offset' is then automatically used to correct for changes in average transcript length across samples.

The resultant count matrix contained nine columns, one for each donor-treatment combination. This count matrix was filtered to remove genes that exhibited consistently low counts, defined as having fewer than ten counts across the nine samples. These genes are highly unlikely to be identified as differentially expressed, and their removal reduced the computational burden on subsequent analyses. Importantly, this filtering was performed on unnormalised counts in order to prevent potential skewing of downstream normalisation processes. Alongside this matrix was another, containing the gene-level offset required for effective normalisation.

2.6 Annotation

Annotation was conducted using the biomaRt v2.54.1 package (Durinck *et al*., 2005). This enabled the conversion of ID types for the purposes of certain downstream analyses, as well as the acquisition of biologically meaningful attributes. Furthermore, this annotation process was modularised to allow ID conversion at any stage of the data pipeline. In practice, this meant that the Ensembl Stable ID format (Martin *et al*., 2023) was predominantly used throughout the pipeline and was converted to other ID types when necessary. A crucial detail of this conversion process is that different ID types occasionally lack a perfect one-to-one match. Therefore, an important implication of this is that not all transcripts detected during the initial sequencing experiment have a corresponding ID type required for certain downstream processes. Thus, the process of

converting between IDs can result in a loss of power due to the lack of unified ID type between various databases.

For example, in section 2.8 the pre-existing Ensembl gene IDs from the count matrix were matched with their corresponding Entrez gene IDs (Maglott *et al*., 2007). This conversion resulted in a loss of genes due to the lack of matching IDs. Special consideration was given to gene IDs with no matching attribute. Specifically, the original ID type was retained and the total number of such instances were reported. In cases where an initial gene ID had multiple matching final IDs, the first of these matches were used, as subsequent matches typically represented pseudogenes.

2.7 Differential gene expression analysis

DGE analysis is a widely used technique for the identification of statistically significant gene-level expression changes across one or more experimental conditions (Das *et al*., 2022). DESeq2 v1.38.3 (Love *et al*., 2014) was selected to perform DGE analysis. DESeq2 performs numerous steps prior to DGE analysis to increase accuracy. These include corrections for library size across samples, outlier detection, shrinkage of dispersion, and hypothesis test with multiple testing correction.

In order to determine which genes are differentially expressed, DESeq2 fits a generalised linear model (GLM; McCullagh and Nedler, 1989) to the expression estimates of each gene, following a negative binomial (NB) distribution. The mean of this distribution is scaled by a normalisation factor, referred to as a size factor. Size factors for each sample are estimated using the median-of-ratios method, which avoids the use of total read counts as this value can be strongly influenced by a small number of highly and differentially expressed genes (Anders and Huber, 2010). Thus, size factors were used to transform each gene's mean to correct for differences in sequencing depth across samples.

DESeq2 subsequently estimated the gene-wise dispersion across samples. This, in conjunction with the normalised means described previously, was used to model the

variance of counts within the matrix for a given gene. A challenge with the estimation of dispersion (and hence variance) is that small sample sizes tend to lead to noisy dispersion estimates, which could, in turn, affect which genes are determined to be differentially expressed. DESeq2 accounts for this issue using a three-stage process. Firstly, an initial estimate of dispersion is calculated for each gene in the count matrix using maximum likelihood. Secondly, each gene's initial dispersion estimate is plotted against its mean of normalised counts. A curve is fitted to this plot, and it is assumed that genes with a similar mean are likely to have a similar dispersion. Lastly, an empirical Bayes approach is used to transform the initial dispersion estimates closer to the curve, with shrinkage strength dependent on the estimated accuracy of the curve and the sample size. If an initial gene dispersion estimate is more than two residual standard deviations above the curve, it is assumed to be an outlier and is not transformed. These transformed dispersion estimates are then taken to be true measures of variability across samples.

Following both estimation of size factors and gene dispersion estimates, DESeq2 performed hypothesis testing. A Wald test was conducted, dividing the log2 fold change (LFC) between two groups by its standard error. DESeq2's default operation tests against the null hypothesis of zero LFC. However, in biological systems many genes are not truly independent of one another. Treatment with M-CSF or GM-CSF results in numerous cascades of transcriptional changes, hence a fold change of zero would not be restrictive enough for isolating truly differentially expressed genes (Krausgruber *et al*., 2011; Pixley and Stanley, 2004). The magnitude of the LFC threshold heavily impacts the number of genes which are determined to be differentially expressed. Therefore, multiple LFC thresholds were tested in relation to known markers of differentiation in monocyte-derived macrophages. In this way, an LFC threshold |1| was selected as appropriate based on the change in expression of known macrophage markers. In practice, this meant that the null hypothesis for the Wald test was that the average fold change across groups did not differ significantly from two. The resulting z-statistic was compared to a standard normal distribution, thereby assigning p-values.

Importantly, DESeq2 automatically performed independent filtering of results prior to a Benjamini–Hochberg correction (Benjamini and Hochberg, 1995). Multiple testing

correction can potentially reduce the power of the analysis by falsely correcting accurately called differentially expressed genes (DEGs). DESeq2 therefore removes genes that are unlikely to be differentially expressed following correction. Firstly, DESeq2 detects outliers via Cook's distance (Cook, 1977), specifically using a cutoff of 0.99. In the case of this study, fewer than six replicates are present for each condition. The outlier genes were therefore removed from further analyses (Love *et al*., 2014). Secondly, normalised mean expression across all samples is used in conjunction with a filtering threshold to reduce the number of genes parsed to the following correction step. DESeq2 evaluates the mean count of each gene and excludes the maximum number of genes based on a specified adjusted *p*-value. In the case of this study, an adjusted *p*-value threshold of 0.05 was selected.

Data quality was further examined at this stage. A variance stabilising transformation (VST) was applied to assist in visualising the dataset using the apeglm approach (Zhu *et al.*, 2019). This transformation removes the dependance of variance on the mean of counts, thereby preventing particularly low or high counts from interfering with visualisations and clustering. These transformed counts were used to conduct a principal component analysis (PCA) to examine clustering of replicates (Pearson, 1901). In addition, other plots were constructed to examine the dataset.

In summation, the unnormalised count matrix was filtered for genes with low expression. The filtered count matrix was used to perform DGE analysis with DESeq2, which involved scaling the counts and estimating a corrected dispersion measure for each gene. A Wald test was conducted comparing the LFC between the three untreated and both sets of three treated samples. In this way, a *p*-value was assigned for each gene when comparing the fold change in expression between untreated to M-CSF-treated and untreated to GM-CSF-treated samples. The *p*-values were corrected for multiple testing via a Benjamini–Hochberg correction and an adjusted *p*-value threshold of 0.05 was selected. Genes with an adjusted p-value lower than 0.05 were therefore determined to be differentially expressed. In this way, two lists of DEGs were identified based on treatment with either M-CSF or GM-CSF.

2.8 Over-representation analysis

While the DEGs provide a wealth of information regarding how the gene expression of samples is correlated with treatment, they lack the specific context of biological function. Functional enrichment analysis (FEA) is the process by which this biological context may be assigned to a list of genes, in this case, those whose expression change was significantly correlated with M-CSF or GM-CSF treatment. The clusterProfiler v4.7.1.003 (Yu *et al.*, 2012) package was selected for this process due to its vast array of up-to-date tools for FEA and for visualising enrichment results (Wu *et al*., 2021).

ClusterProfiler makes use of biological ontologies through a number of databases including Gene Ontology (GO; Ashburner *et al.*, 2000), Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa and Goto, 2000), and Disease Ontology (DO; Osborne *et al.*, 2009). Ontologies refer to a set of well-defined terms and their relationships to one-another. The GO database provides annotations for the following ontologies: molecular functions (MF), cellular components (CC), and biological processes (BP; Ashburner *et al*., 2000). The three ontologies provided by the GO Consortium house different kinds of information. MF refers to the biochemical activity of a gene product, whereas CC is related to the cellular location at which these MFs typically occur. The BP ontology refers to the biological objective to which a gene or gene product contributes (Ashburner *et al.*, 2000). The primary interest of this study is the identification of markers of differentiation, which is itself a biological process. Thus, the BP ontology was the primary lens through which the FEA was performed.

Each ontology type (MF, CC, and BP) within the GO database has a defined set of ‘GO terms’ for (Ashburner *et al*., 2000). These GO terms comprise a group of genes/gene products which represent a biologically meaningful class. The relationship between GO terms can be described as nodes in a network, referred to as a directed acyclic graph. Linked nodes in the graph represent ontologies which share common genes. This graph is directed, as some nodes are ‘parents’ to others, meaning that they encompass all of the genes in the ‘child’ term. This system allows a great deal of experimentally-validated information about genes and their products to be represented as part of this network (Ashburner *et al*., 2000). ClusterProfiler gathers human ontology

information from the relevant databases through the use of the org.Hs.eg.db v3.16.0 package, available through bioconductor (Gentleman *et al*., 2004).

FEA was performed through the use of over-representation analysis (ORA), which identifies statistically significant differences in the observed number of differentially expressed genes when compared to the expected number if they were to be randomly sampled. Specifically, *p*-values are calculated for each GO term using a hypergeometric distribution. As the goal is to identify over-represented GO terms, statistically under-represented terms are ignored. This corresponds to a one-sided version of Fisher's exact test. Essentially, the number of genes belonging to each GO term are counted and related to the total number of supplied genes. This ratio is then also determined for the 'universe', which is the set of all genes supplied to prior statistical testing, i.e. all genes whose expression was quantified in this study. In this case, the universe was the set of genes that were previously supplied to DESeq2. Lastly, Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) is employed to adjust *p*-values, along with *q*-value estimation to control for the false discovery rate (Storey, 2002).

As outlined in section 2.6, the GO database makes use of Entrez gene IDs (Maglott *et al*., 2007), necessitating the conversion from Ensembl gene IDs (Martin *et al*., 2023). The discrepancy in the conversion between these IDs had implications for FEA due to the means by which clusterProfiler determines the statistical significance of the over-representation of genes for a given GO term.

ORA was performed on the lists of DEGs using clusterProfiler's 'enrichGO' function, which utilised the most up to date annotations from the GO database (The Gene Ontology Consortium *et al.*, 2023). This process was then repeated for partitioned sets of DEGs, namely those that were found to be differentially expressed in only one treatment, or common to both treatments. In each case, a strict adjusted *p*-value threshold of $p < 0.05$ was implemented, as well as a *q*-value threshold of $q < 0.05$. Thus, five separate sets of significantly enriched GO terms were identified, each corresponding to a specific set of DEGs.

GO enrichment results often have high levels of redundancy due to the interconnectedness of similar terms (Wu *et al.*, 2021). Therefore, clusterProfiler implements a means of simplifying the resultant GO terms using the GOSemSim

v2.24.0 (Yu, 2020) package. The graph-based Wang method (Wang *et al.*, 2007) was used to calculate the semantic similarity scores between GO terms. Groups of GO terms sharing a semantic similarity higher than a specified threshold were excluded, retaining only the single GO term from the group with the smallest adjusted *p*-value. A similarity threshold of 0.7 was implemented to remove redundancy in the enrichment results.

2.9 Construction of transcription factor protein-protein interaction networks

In order to explain the distinct patterns of gene expression across the two treatments, potential transcription factors were identified within the lists of DEGs, as well as the genes encoding their known protein-protein interacting partners. This was achieved by identifying genes belonging to the 'sequence-specific DNA binding' molecular function GO term (Tripathi *et al.*, 2013). BiomaRt was used to gather known GO term names for each DEG associated with M-CSF and GM-CSF. The lists of DEGs were then filtered for those belonging to the sequence-specific DNA binding GO term, allowing two lists of potential transcription factors to be generated. These lists were then compared with key GO enrichment results to identify transcription factors that were differentially expressed and involved in key biological processes such as myeloid cell differentiation and regulation of inflammatory response.

Known protein-protein interactions for the transcription factors were gathered via the STRING v12.0 database (Szklarczyk *et al.*, 2015). STRING aggregates protein-protein interactions based from high-throughput experimental data, through the mining of databases and literature, and from predictions based on genomic context analysis (von Mering *et al.*, 2005). Probability scores are assigned to each protein-protein interaction from a given source, and these scores are then integrated to give a final probability score (von Mering *et al*., 2005). The known protein-protein interactions for each transcription factor were downloaded in the form of a .tsv file. Interaction parameters were set to a maximum of 50 interactions with a high confidence threshold of 0.7 (von Mering *et al.*, 2005). This allowed for a large number of highly likely protein-protein interactions to be identified for each transcription factor.

Lastly, the full lists of interaction partners for each transcription factor were checked against the lists of DEGs to identify those that were significantly differentially expressed in association with one or both of the treatments. Interacting partners that were not differentially expressed in at least one treatment were removed. The differentially expressed transcription factors, along with their differentially expressed interacting partners, were then used to construct networks using Cytoscape v3.10.1 (Shannon *et al*., 2003) in conjunction with stringApp v2.0.3 (Doncheva *et al*., 2019) and Omics Visualiser v1.3.1 (Legeay *et al*., 2020). The clusterMaker2 v2.3.4 (Morris *et al*., 2011) app was used to deconvolute networks via a Markov clustering algorithm (MCL). A selection of the most interconnected and uniquely differentially expressed transcription factors were retained and shown within the final network diagrams.

Chapter 3: Results

3.1 Sequencing data quality

FastQC, and subsequently MultiQC, were used to aggregate ten metrics of sequence data quality (Figure 3.1). Interpretation of these metrics was conducted as outlined by Andrews (2010), with each file being marked as either successful (light blue), abnormal (intermediate blue), or failing (dark blue) for each metric. As each pair of files represents an individual donor and condition combination, file pairs (files sharing the same code such as SRR16816545.1 and SRR16816545.2) are referred to as samples henceforth.

Most importantly, the per base sequence quality was successful across all nine samples. This metric is determined using the per-base Phred scores present within the fastq files as reported by the sequencing instrument. For each base, FastQC examines the median and quartiles for Phred scores. A file is considered to be of passing quality if the lower quartile Phred score for every base is ten or higher, and if the median score is 25 or higher. A Phred score of 25 corresponds to a 99.68% accuracy in the base call, meaning that the median quality score across all bases in all files was at least 99.68% accurate.

The per tile sequence quality for 14 of the 18 files was marked as a failure, with one file being flagged as abnormal, and the final three files passing. Failures of this metric were expected due to the stringent requirements for per tile sequence quality and the nature of this particular dataset. If any tile shows a mean Phred score two or more below the mean for that base across all tiles then this metric is marked as abnormal for the entire sample. If any tile shows a mean Phred score five or more lower than the mean for that base across all tiles then it is marked as a failure. However, an examination of the individual plots reveals no large-scale errors within tiles (data not shown). Wherever a tile was flagged, it was typically only for a small section of compromised bases, indicating that reads produced from said file were still overall of high quality. As a result, no tiles were filtered from any of the files.

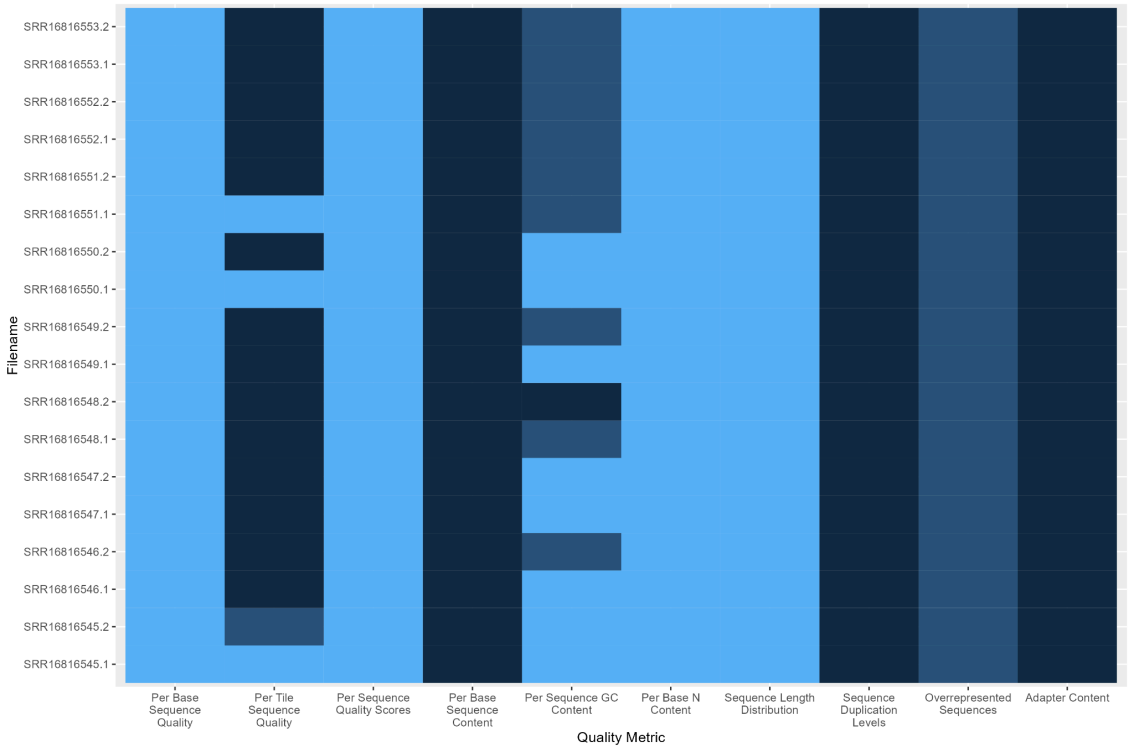

Figure 3.1 MultiQC heatmap of aggregate statistics. For each of the nine pairs of fastq files the per base sequence quality, per tile sequence quality, per sequence quality, per base sequence content, per sequence GC content, per base N content, sequence length distribution, sequence duplication, overrepresented sequences, and adapter content were determined. A light blue tile indicates the metric passed for the sample. Intermediate blue indicates an abnormality, whereas dark blue indicates a failure.

The per sequence quality scores were acceptable for all nine samples. Unlike the per base quality score, which examined each base averaged over all sequences, this metric examined each individual sequence's average Phred score, with a threshold of 27 (99.8% accuracy). As each of the nine samples met this quality metric, it can be firmly stated that accurate reads were present across all samples.

The per base sequence content was marked as highly unusual, however this is to be expected for RNA-seq data given the library preparation involved. Specifically, up to the first twelve base pairs at the 5' end of an RNA-seq read are likely to have a biassed composition due to the use of random hexamers and fragmentation using transposases (Roberts *et al*., 2011). The presence of this bias across all sequences can alter the proportions of nucleotide bases beyond the 20% margin of error measured by FastQC.

Hence, whilst there is an underlying technical bias present, it is the result of normal RNA-seq library preparation. Furthermore, trimming these sections does not markedly improve downstream analyses as salmon (Patro *et al*., 2017) accounts for such biases through its bias models. Hence, no trimming was performed on reads.

Per sequence GC content was the most inconsistent metric across all nine samples. Eight files passed, nine files were abnormal, and one file failed outright. Abnormalities and failures of this metric are defined by summing the deviations from the modelled normal distribution automatically determined by FastQC. Specifically, if more than 15% or 30% of the reads are detected as deviating from the distribution then the sample is marked as abnormal or a failure, respectively. This is a more complex metric to explain, however it is likely that the highly varied nature of RNA-seq reads, in combination with certain overrepresented sequences or highly expressed genes, could be enough to perturb average GC content within a sample.

Both per base N content and sequence length distribution were successful metrics for all samples. A base call of 'N' represents an ambiguous call by the sequencer and is typically not systemic. FastQC raises a warning if any base position shows an N content of > 5%, where all nine samples were closer to 0%. Sequence length distribution was also unlikely to fail as all reads were exactly 151bp across all samples.

Sequence duplication levels were failed for all samples, indicating that more than 50% of the total sequences within the first 100 000 were non-unique. RNA-seq typically produces this phenomenon when utilising FastQC, as sequencing depth is prioritised in order to capture lowly-expressed genes (Andrews, 2010). Hence, more common transcripts are greatly over-represented, as is further indicated by the overrepresented sequences metric being marked as abnormal for all samples.

Lastly, there was a high degree of adapter content within each sample. In particular, 'Illumina Universal Adapters' were detected as having a presence in more than 10% of all reads. While this could affect mapping rates, correction for such bias would involve trimming of adapter sequences. The degree to which trimming is performed has been shown to affect RNA-seq expression estimates (Williams *et al*., 2016). Furthermore,

trimming can vastly increase the computational burden. As a result, the decision was made to perform no trimming in order to maintain a streamlined data pipeline.

3.2 Pseudoalignment mapping rates

Mapping performed by salmon produced low mapping rates relative to what would be expected for RNA-seq data of substantive quality (Figure 3.2). In particular, the untreated group achieved mapping rates of 28.31%, 39.16%, and 33.49% for donors one, two, and three, respectively. This indicates that less than half of fragments within the files mapped to a known feature on the human reference transcriptome. The six treated samples had mapping rates ranging from 51.10% - 59.83%, substantially higher than the untreated samples. One would typically expect rates between 60 and 70% although this can be slightly lower when mapping to a reference transcriptome (Conesa *et al*., 2016).

The discrepancy in mapping rates between the untreated and treated datasets could be caused by underlying biological and technical differences. Biologically, monocytes and macrophages have distinct patterns of gene expression. Distinctions in which transcripts were present within the cells may thus be contributing to lower mapping rates for the untreated samples. The FastQC analysis of the untreated files revealed a higher degree of sequence duplication, which could be further contributing to the mapping rate discrepancies. This presents an important shortcoming of these analyses, as unmapped transcripts could represent meaningful underlying biological differences. However, without the presence of technical replicates the nature of these unmapped fragments is difficult to determine. As a result, despite the statistical rigour present throughout these analyses, care should be taken when drawing biological conclusions in the absence of further validation.

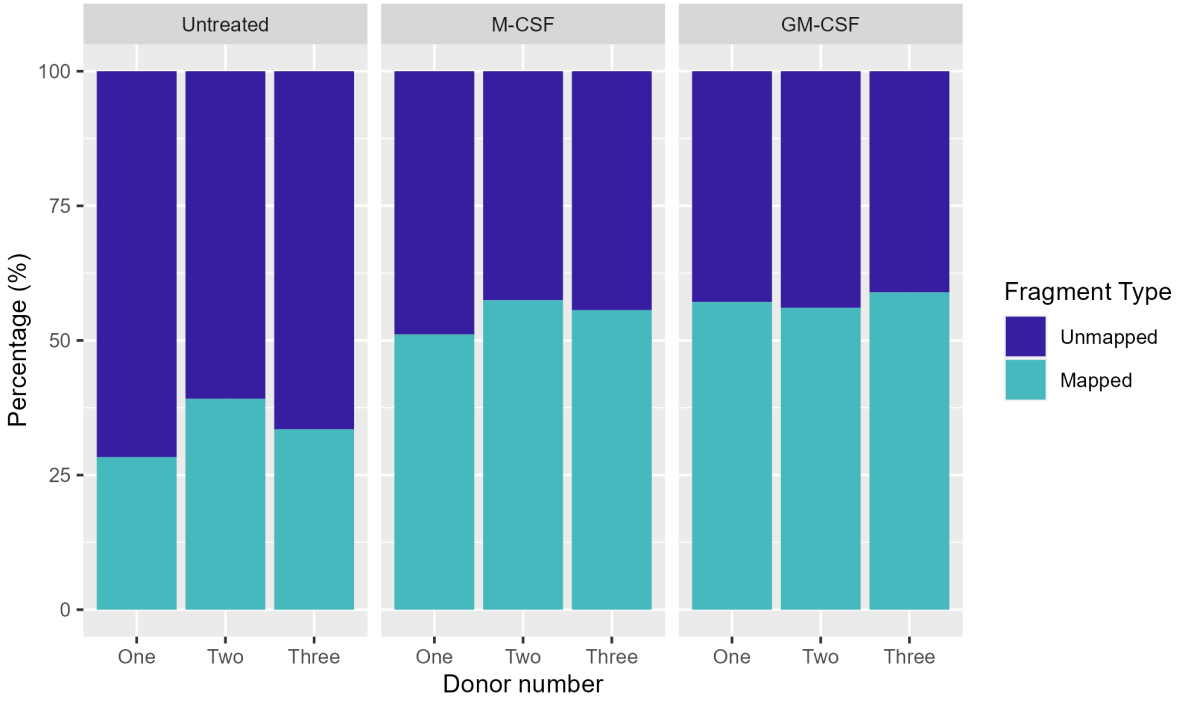

Figure 3.2 Percentage of mapped reads per sample. The turquoise portion of the bars represent the percentage of reads successfully mapped to the reference transcriptome during salmon quantification, while the dark blue portion shows unmapped reads. Samples were grouped according to treatment type, specifically untreated (left), M-CSF-treated (middle), and GM-CSF-treated (right).

3.3 Filtering and quality control

The count matrix constructed using tximport retained 28 340 features for subsequent analyses. The total normalised and unnormalised counts for each sample were examined to determine how DESeq2's normalisation via size factors altered the number of counts per sample (Figure 3.3). Donor two's untreated sample displayed a smaller number of total counts (and hence lower sequencing depth) when compared to all other samples. Therefore, normalisation vastly increased this sample's total counts. The three M-CSF-treated samples were varied in their total counts prior to normalisation. Following normalisation this variation was removed, generally reducing overall count

levels. A similar reduction in total counts was observed for the three GM-CSF-treated samples, although they lacked initial variation in total counts.

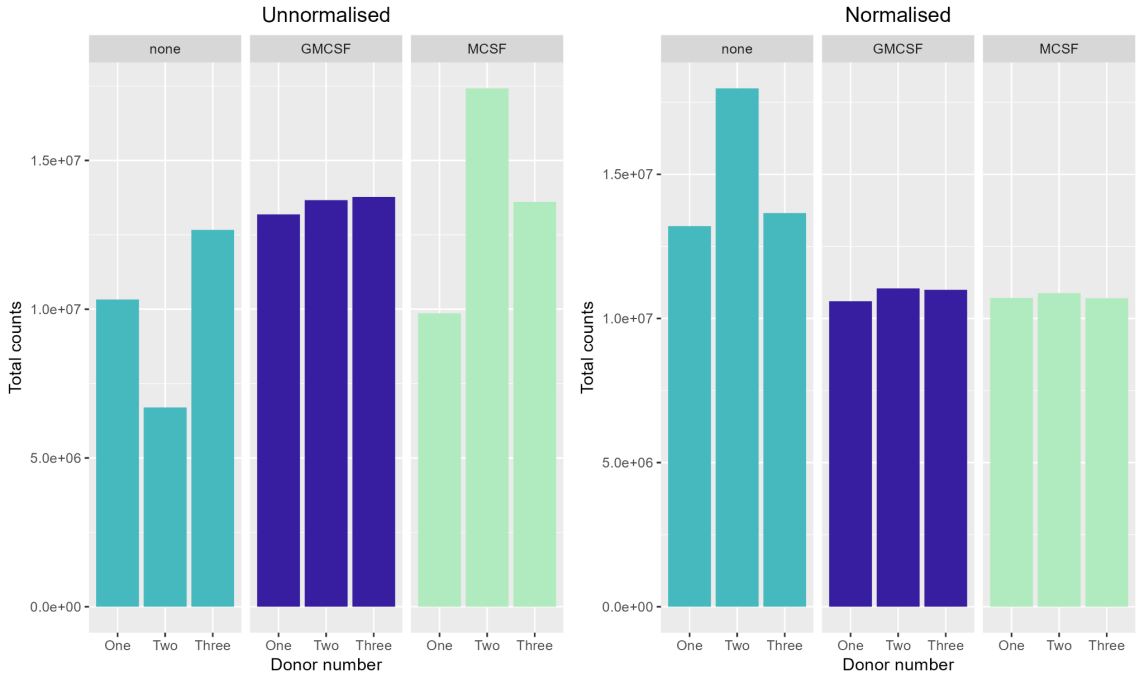

Figure 3.3 Barplots of total counts before and after normalisation. Total counts per column (sample) of the count matrix were determined pre and post normalisation. The total unnormalised counts for the untreated samples (turquoise) were substantially increased via DESeq2's normalisation, whereas the opposite was true for GM-CSF (dark blue) and M-CSF (light green).

Given the unusual distribution of total counts with respect to donor two's untreated sample, as well as the overall count variations, the normalised counts produced by DESeq2's algorithm were subjected to a PCA to examine the clustering of the nine samples (Figure 3.4). A tight clustering was observed for each sample group based on treatment type, with PC1 splitting untreated samples from treated samples and accounting for 70% of variance. PC2 accounted for 22% of variance and distinctly separated the M-CSF and GM-CSF samples. This distinctive clustering indicates that each biological replicate was highly similar to other members of its treatment group following normalisation.

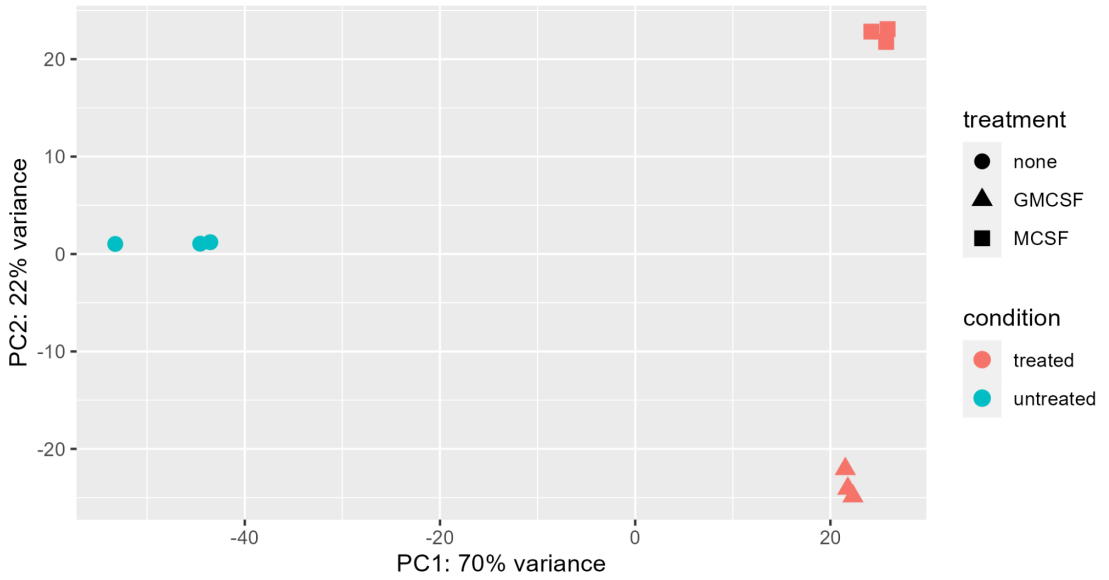

Figure 3.4 PCA of normalised counts. Untreated (blue) and treated (red) groups were heavily split by PC1, with M-CSF and GM-CSF treatments being split by PC2. Clusters were distinct, with replicates from a given group converging tightly for M-CSF and GM-CSF. There was slightly more dispersion in the untreated cluster, with one replicate further distancing itself along the first principal component.

3.4 Macrophage marker gene expression changes

Four known markers of differentiation in monocyte-derived-macrophages (MDMs), namely *SPP1*, *CSF1R*, *FCGR3A*, and *MRC1,* were examined in order to validate that differentiation had successfully occurred within the treated cells. All of these selected genes were significantly differentially expressed and upregulated by both M-CSF and GM-CSF (Figure 3.5). *MRC1* showed the highest level of counts, followed by *SPP1*, *CSF1R*, and finally *FCGR3A*. For each of these genes, the untreated samples displayed low counts with a small degree of variance. The expression of genes such as *FCGR3A* and *SPP1* was varied within treatment groups, being especially pronounced for GM-CSF. Given the importance of these markers in distinguishing monocytes from macrophages, these results clearly indicate the successful differentiation of these monocytes into macrophages.

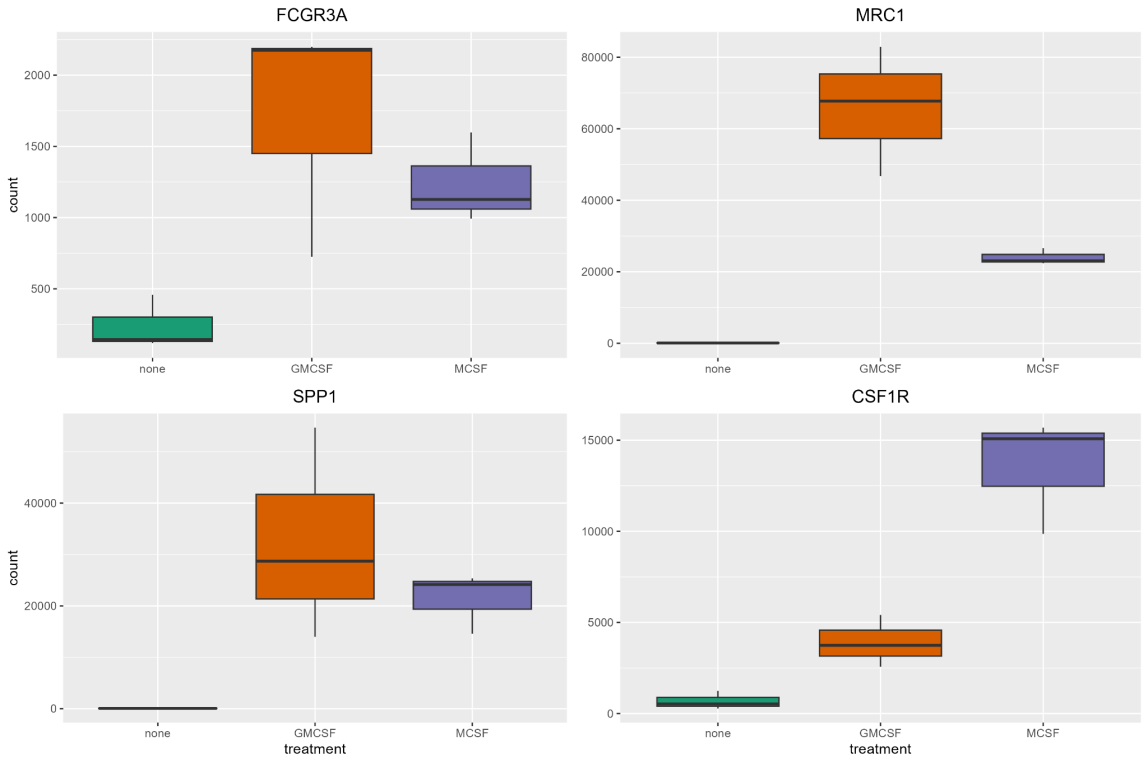

Figure 3.5 Boxplots of known marker genes of MDMs. Four differentially expressed known markers of differentiated monocyte-derived-macrophages. Each of the boxplots are titled by gene name and show normalised counts for the three replicates based on condition. Untreated are green (left), GM-CSF are red (orange), and M-CSF is blue (right) for each of the plots. The thresholds for significance were an adjusted *p*-value < 0.05 and a LFC > |1|.

Given that M-CSF and GM-CSF are known to polarise macrophages towards the M2 and M1 subtypes, respectively, the relative expression change of 15 M1 and 15 M2 macrophage subtype marker genes was examined (Figure 3.6). Six M1 and nine M2 markers were significantly differentially expressed in at least one treatment. Of these significantly differentially expressed markers, all M1 genes were downregulated in both treatments. Conversely, seven of the nine differentially expressed M2 genes were upregulated in M-CSF, while five were significantly upregulated for GM-CSF. *FOLR2* was not differentially expressed in association with GM-CSF treatment, while *IGF1* and *IL10* were significantly downregulated. Finally, *IL10* and *PPARD* were slightly downregulated for M-CSF, although this was not significant enough to represent differential expression. The direction of expression change was largely consistent across

both treatments, with the notable exception of *IGF1*. *IGF1* was the only differentially expressed marker to have an inverse direction of change between treatments, as it was upregulated in M-CSF and downregulated in GM-CSF.

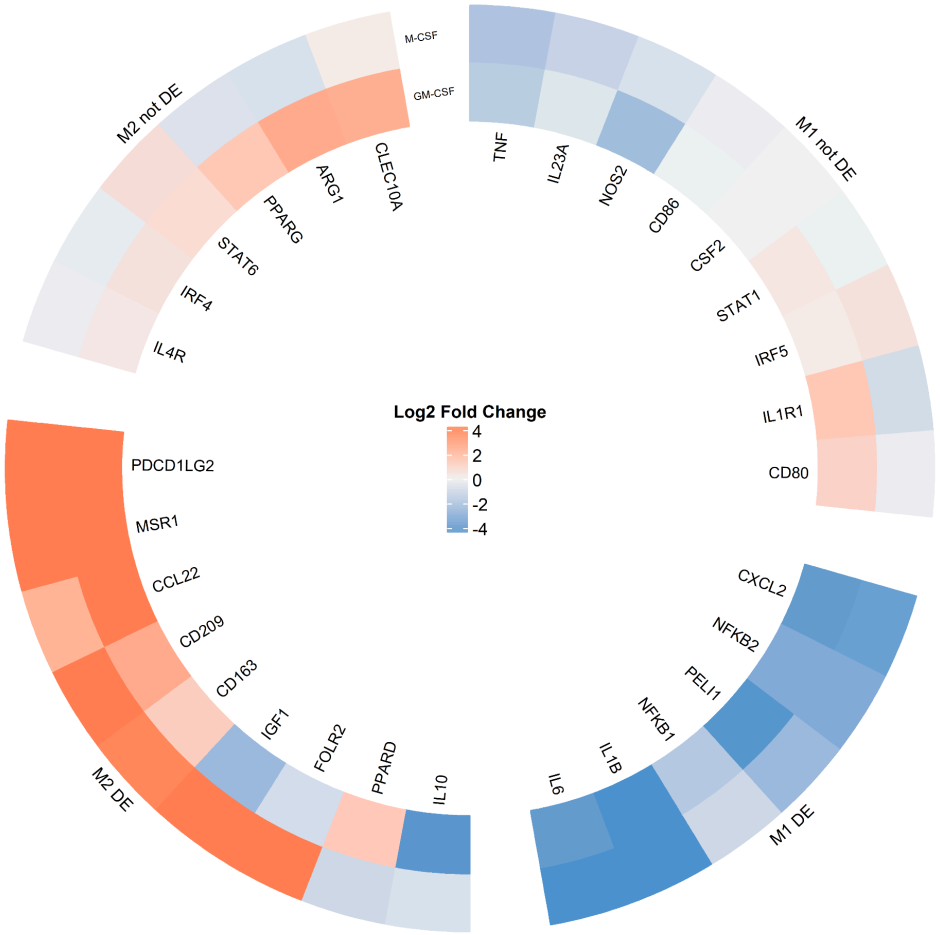

Figure 3.6 Heatmap of known macrophage subtype markers for M-CSF and GM-CSF. 30 known markers of differentiated monocyte-derived-macrophages were examined, with M1 markers on the right and M2 markers on the left. Tiles are coloured according to the LFC of a given treatment's normalised counts relative to the untreated samples. The outer ring of the heatmap represents the M-CSF treated samples relative to the untreated samples. The inner ring of the heatmap represents the GM-CSF treated samples relative to the untreated samples. Entirely grey bars indicate genes that were not present within the count matrix. The thresholds for significance were an adjusted *p*-value < 0.05 and a LFC > |1|.

3.5 Differential gene expression analysis

Following the evaluation of specific marker genes, a broader examination of the differential gene expression results was conducted. Of the 28 340 features present within the count matrix a total of 5 737 (~20%) DEGs were identified across both treatments (Figure 3.7A). M-CSF treatment produced 4 072 DEGs while GM-CSF treatment produced 4 399 DEGs. The full lists of DEGs can be found in the appendices (Tables S1 and S2). A substantial number of DEGs (2 734) were differentially expressed for both treatments, meaning that ~47% of the total number of DEGs across both lists were common.

These DEGs were further separated by their LFC, with a positive LFC indicating an increase in expression and a negative LFC indicating a reduction in expression in relation to the expression levels in the untreated monocytes. A higher number of downregulated DEGs (Figure 3.7C) were observed relative to the number of upregulated DEGs (Figure 3.7B). However, the degree of commonality among upregulated and downregulated DEGs across the two treatments was very consistent, at ~46% and ~47%, respectively.

The total set of DEGs across both treatments skewed towards being downregulated, as 3 677 (~64%) of the total 5 737 DEGs were downregulated, while only 2 085 (~36%) of all DEGs were upregulated. Upon inspection there appears to be a discrepancy in the total number of DEGs between figures. However this is due to a set of 25 genes that were inversely regulated, being upregulated in one treatment and downregulated in another. Given the high number of DEGs, even when splitting based on LFC, reduced sets of known marker genes were examined to elucidate the potential phenotypic characteristics of the macrophages following treatment.

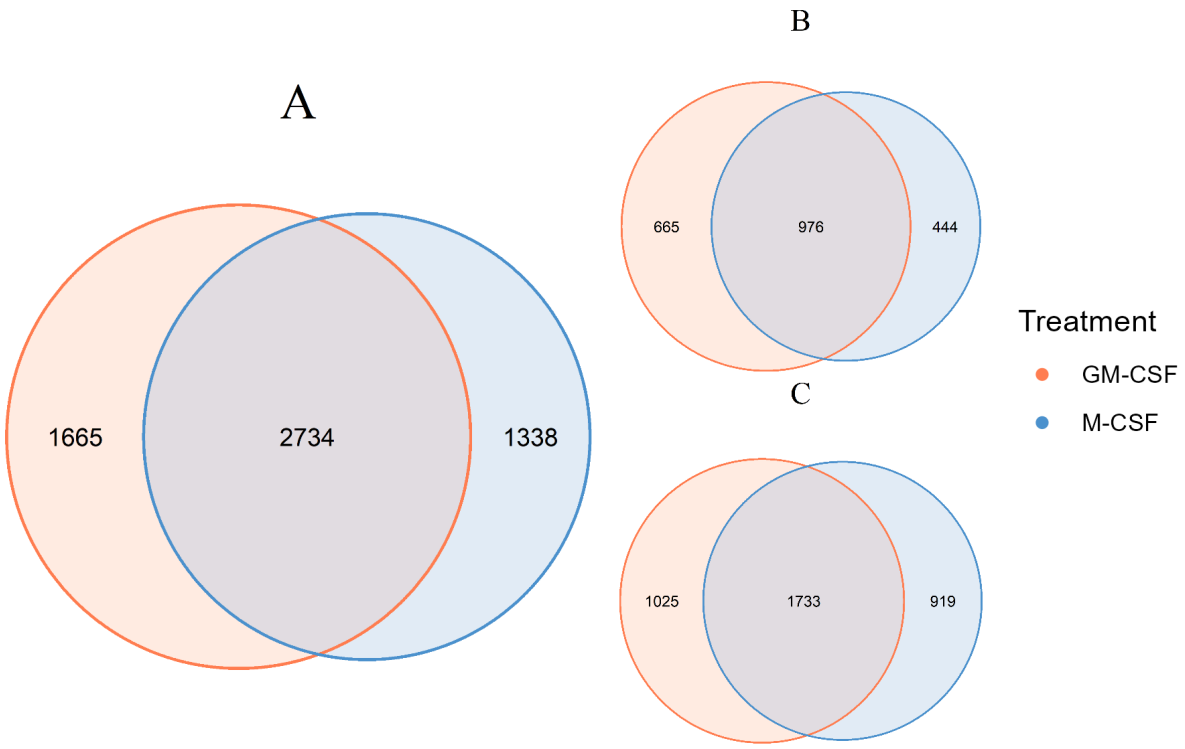

Figure 3.7 Venn diagrams of differentially expressed genes. Three Venn diagrams are shown, namely; (A) all genes (left), (B) upregulated genes (top right), and (C) downregulated genes (bottom right). Red circles represent DEGs from GM-CSF treatment, whereas blue circles show DEGs from M-CSF treatment. Overlaps of the two circles show the number of genes shared between the two given lists. The proportion of DEGs unique to a given list are represented by the non-overlapping regions of the circles. The thresholds for significance were an adjusted *p*-value < 0.05 and a LFC > |1|.

3.6 GO Enrichment Analysis

GO (Ashburner *et al*., 2000) enrichment analysis was performed on the full set of 4 072 and 4 399 DEGs resulting from each treatment. This resulted in 67 (Table S3) and 76 (Table S4) enriched GO terms for M-CSF and GM-CSF, respectively. A selection of the GO terms most statistically overrepresented within the lists of DEGs was visualised as an enrichment map to illustrate the relationships between the top GO terms for both M-CSF and GM-CSF (Figure 3.8). The network comprises 48 nodes, 20 of which are shared GO terms that were enriched for both treatments. One third of the network is

made up of unconnected GO terms (16 nodes), indicating a range of independent biological processes associated with the sets of DEGs.

Three main modules can be identified within the network, being those that have four or more connections between terms. The smallest of these modules, located in the bottom right of the network, contains five nodes exclusive to GM-CSF. The GO terms of this module are involved in the process and regulation of DNA replication. Towards the top of the network is the largest module, containing enriched GO terms unique to and shared across the two treatments. A highly connected core of this module involves biological processes related to chromosome and nucleosome organisation. However, the largest node of this module is distant from this core and is shared by M-CSF and GM-CSF, namely ‘myeloid cell differentiation’. The last of the three modules, located on the left side of the network, contained seven nodes associated with leukocyte functions. Four of these nodes were shared by M-CSF and GM-CSF, namely ‘leukocyte migration’, ‘leukocyte chemotaxis’, ‘response to lipopolysaccharide’, and ‘regulation of inflammatory response’. The myeloid cell differentiation and regulation of inflammatory response GO terms were of particular interest in this study.

A potential avenue of further investigation was the use of partitioned DEG sets in ORA. Using a partitioned set of DEGs (rather than the full set) in ORA provides a more focused GO enrichment outcome, although it lacks the holistic context provided by the full set. Partitioning of the full DEG set for each treatment was performed on the basis of being unique to one of the treatments, or common across both. Neither the 1 338 nor the 1 665 DEGs unique to M-CSF and GM-CSF, respectively, produced an enrichment result, meaning that no biological processes were overrepresented within these lists of DEGs. This leads one to expect that a high number of the commonly differentially expressed genes would be involved in the biological processes enriched for the full sets of DEGs. This was demonstrated in the enrichment result produced by this common set of 2 734 DEGs, which comprised 47 GO terms. Included in this enrichment were the GO terms ‘regulation of inflammatory response’ and ‘myeloid cell differentiation’, which had already been deemed to be of interest.

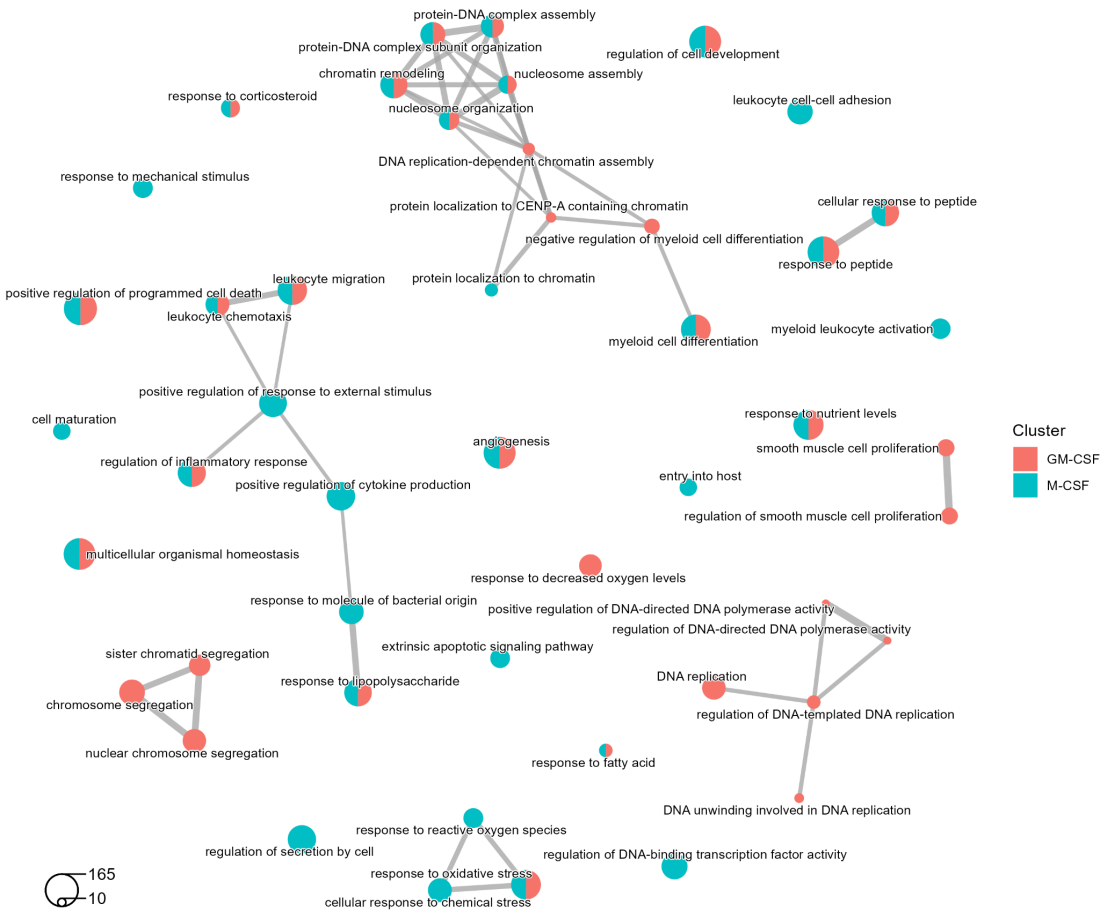

Figure 3.8 Enrichment map of GO terms for GM-CSF and M-CSF. A subset of GO terms for GM-CSF (red) and M-CSF (blue) are displayed. Nodes (coloured circles) represent GO terms and are sized according to the number of DEGs they contain. The two circles in the bottom left of the figure show scale for node size according to the smallest node (ten genes) and the largest node (165 genes). The grey lines connecting two nodes indicate an overlap in the gene sets comprising the nodes, with thicker lines showing greater overlap. GO terms that were enriched for both GM-CSF and M-CSF are a mixture of red and blue.

In order to identify differences in gene expression within the aforementioned biological processes across treatments, the degree of expression change in the corresponding genes was examined. The 'myeloid cell differentiation' biological process contained 85 DEGs

common to both treatments, 32 DEGs unique to GM-CSF, and 22 DEGs unique to M-CSF (Figure 3.9). Of the 85 common DEGs, inhibin subunit beta-A (*INHBA*) was the only one to display an inverse direction of change across treatments. A number of other common DEGs displayed different magnitudes of change in their expression across the two treatments.

Figure 3.9 Heatmap of DEGs found within the myeloid cell differentiation GO term. A total of 139 differentially expressed genes were associated with the myeloid cell differentiation GO term across both treatments. The common genes (right) and the uniquely differentially expressed genes (left) are split between the two arcs of the heatmap. Tiles are coloured according to the LFC of a given treatment's normalised counts relative to the untreated samples, with orange showing increased expression and

blue representing decreased expression relative to untreated cells. Gene symbols are displayed along the interior of the ring. The inner ring of the heatmap represents the GM-CSF treated samples relative to the untreated samples. The outer ring of the heatmap represents the M-CSF treated samples relative to the untreated samples. Grey tiles indicate a lack of differential expression based on thresholds of an adjusted *p*-value < 0.05 and a LFC > |1|.

The 'regulation of inflammatory response' biological process was closely examined given its particular relevance to the distinct way in which M-CSF and GM-CSF are expected to prime differentiated macrophages. A high degree of commonality was once again observed, with 70 DEGs common to both treatments, 24 DEGs unique to GM-CSF, and 30 DEGs unique to M-CSF (Figure 3.10). Two genes displayed inverse expression changes, namely CC chemokine receptor 2 (*CCR2*) and *IGF1*, being upregulated in GM-CSF and M-CSF, respectively.

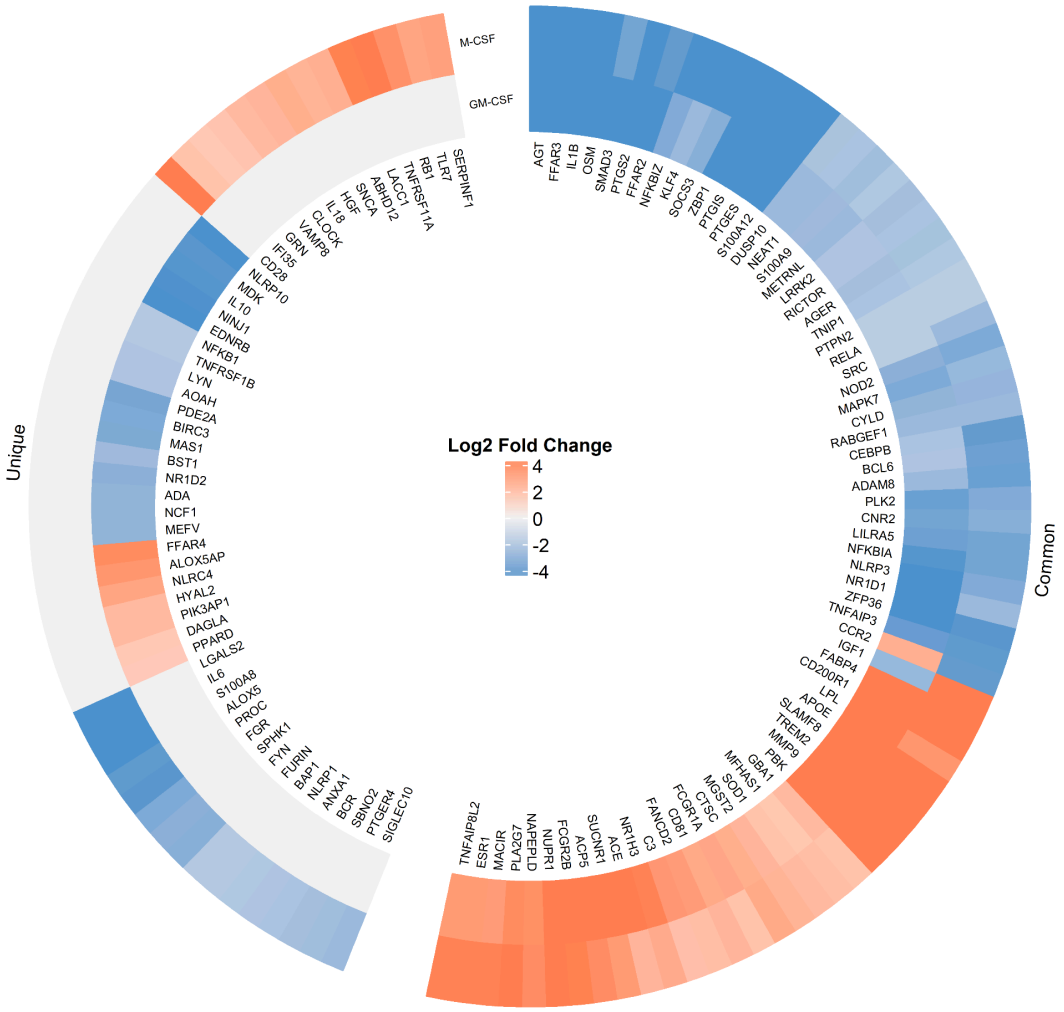

Figure 3.10 Heatmap of DEGs found within the regulation of inflammatory response GO term. A total of 124 differentially expressed genes were associated with the regulation of inflammatory response GO term across both treatments. The common genes (right) and the uniquely differentially expressed genes (left) are split between the two arcs of the heatmap. Tiles are coloured according to the LFC of a given treatment's normalised counts relative to the untreated samples, with orange showing increased expression and blue representing decreased expression relative to untreated cells. Gene symbols are displayed along the interior of the ring. The inner ring of the heatmap represents the GM-CSF treated samples relative to the untreated samples. The outer ring of the heatmap represents the M-CSF treated samples relative to the untreated samples. Grey tiles indicate a lack of differential expression based on thresholds of an adjusted p-value < 0.05 and a LFC > |1|.

3.7 Transcription factor network construction

Given the large number of DEGs involved in both the 'myeloid cell differentiation' and 'regulation of inflammatory response' GO terms, a targeted examination of the transcription factors was conducted. Of the 4 072 and 4 399 DEGs for M-CSF and GM-CSF, respectively, 233 and 222 were identified as possible transcription factors due to having sequence specific DNA-binding activity (Tables S5 and S6). These full lists of transcription factors were cross-referenced with the DEGs associated with the two GO terms of interest. The DEGs associated with myeloid cell differentiation contained 27 potential transcription factors, while 15 were associated with regulation of inflammatory response (Table S7). The known protein-protein interactions (PPIs) for each of these differentially expressed transcription factors were obtained from the STRING database as described in section 2.9.

Of the 27 differentially expressed transcription factors associated with myeloid cell differentiation, four were selected due to their high number of PPIs, relevance in the literature, and their treatment-specific differential expression. *MAFB*, *MYC*, *RUNX1*, and *SPI1* are all known to play important roles in coordinating the differentiation of monocytes into macrophages. A brief description of each of these transcription factors and their roles is provided in Table 3.1. A network was constructed based on these transcription factors and their differentially expressed interacting partners (Figure 3.11). These networks were constructed to display the LFC of both M-CSF and GM-CSF in a ring around each node (gene), with the left hand half representing M-CSF and the right hand half representing GM-CSF.

Table 3.1 Uniquely differentially expressed transcription factors involved in myeloid cell differentiation.

Treatment	Transcription Factor	Log2 Fold Change	Function	Reference
GM-CSF	MAFB	-5.96341	Regulation of lineage-specific hematopoiesis via repression of ETS1-mediated transcription of erythroid-specific genes in myeloid cells.	(Kelly *et al*., 2000)
	MYC	1.997188	Forms a heterodimer with the related transcription factor MAX, which in turn regulates the transcription of genes involved in cell cycle progression, apoptosis and cellular transformation.	(Lüscher, 2001)
M-CSF	RUNX1	-2.74228	Binds to the core element of many enhancers and promoters for genes involved in the development of normal hematopoiesis.	(de Bruijn and Dzierzak, 2017)
	SPI1	-1.68897	Activates gene expression during myeloid and B-lymphoid cell development via promoters of target genes, and regulates their expression in coordination with other transcription factors and cofactors.	(Li *et al.*, 2020)

The network resulting from the four myeloid cell differentiation transcription factors comprised four main modules, one for each transcription factor. The *MYC* gene, being uniquely differentially expressed in M-CSF, had the highest degree of connectivity involving 33 other DEGs. The other modules were smaller than *MYC*, with *SPI1* having 14 other genes in its network, *MAFB* having five, and *RUNX1* having three. Lastly, the

SRC and *NCF1* genes, although having known interactions with one or more of the query transcription factors, clustered separately.

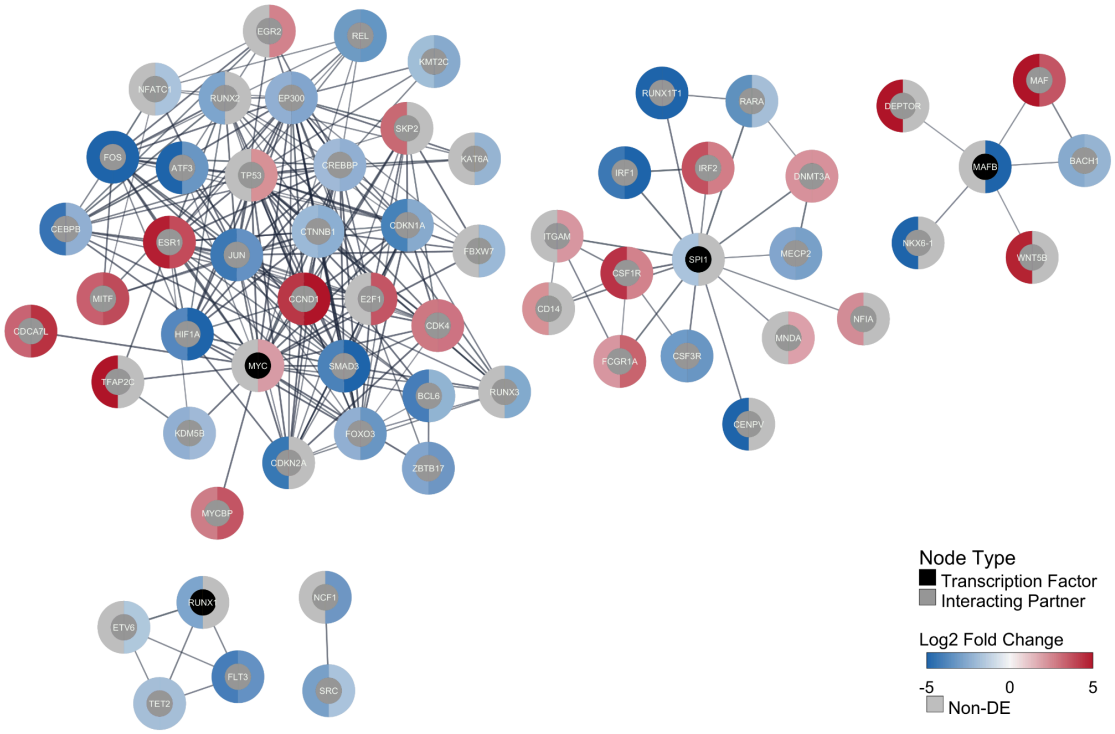

Figure 3.11 STRING network of select transcription factors involved in myeloid cell differentiation. Nodes represent differentially expressed genes. Nodes coloured black represent the selected transcription factors, while grey nodes represent interacting partners. Nodes are connected via edges, the thickness of which is proportional to confidence of the protein-protein functional interaction. Each node is surrounded by a ring representing the Log2 fold change the gene experienced under each treatment, with the left hand half representing M-CSF and the right hand half representing GM-CSF. In cases where one of these halves are grey, it indicates that the gene was not differentially expressed in the corresponding treatment (significance thresholds being an adjusted p-value < 0.05 and a LFC > |1|).

Of the 15 differentially expressed transcription factors associated with the regulation of inflammatory response, five were selected following a similar justification to the previous network (Table 3.2). The network for these five genes was likewise constructed following the same approach, but now involved *NFKB1*, *NR1D2*, *PPARD*, *CLOCK*, and *RB1* (Figure 3.12).

Table 3.2 Uniquely differentially expressed transcription factors involved in the regulation of inflammatory response.

Treatment	Transcription Factor	Log2 Fold Change	Function	Reference
GM-CSF	NFKB1	-1.99603	DNA binding subunit of the NF-κB protein complex, which directly controls inflammatory gene expression.	(Cartwright *et al*., 2016; Somma *et al*., 2021)
	NR1D2	-3.31761	Transcriptional repressor involved in circadian rhythms and metabolism.	(Chen *et al*., 2020; Ikeda *et al*., 2019)
	PPARD	1.718916	Integrator of transcriptional repression and nuclear receptor signalling involved in lipid metabolism.	(Chawla, 2010)
M-CSF	CLOCK	2.432183	Regulation of circadian rhythms.	(Chen *et al*., 2020)
	RB1	4.10156	Negative regulator of the cell cycle and inflammation.	(Hutcheson *et al*., 2015)

Four main modules were once again observed, due to *CLOCK* and *NR1D2* clustering together. *NFKB1* was found within the largest module, being associated with 36 other DEGs. The smaller modules comprised twelve, twelve, and four interacting partners in addition to the transcription factors *CLOCK/NR1D2*, *RB1*, and *PPARD*, respectively. Lastly, the *BTRC*, *CRABP2*, and *RARA* genes did not cluster together with any of the query transcription factors.

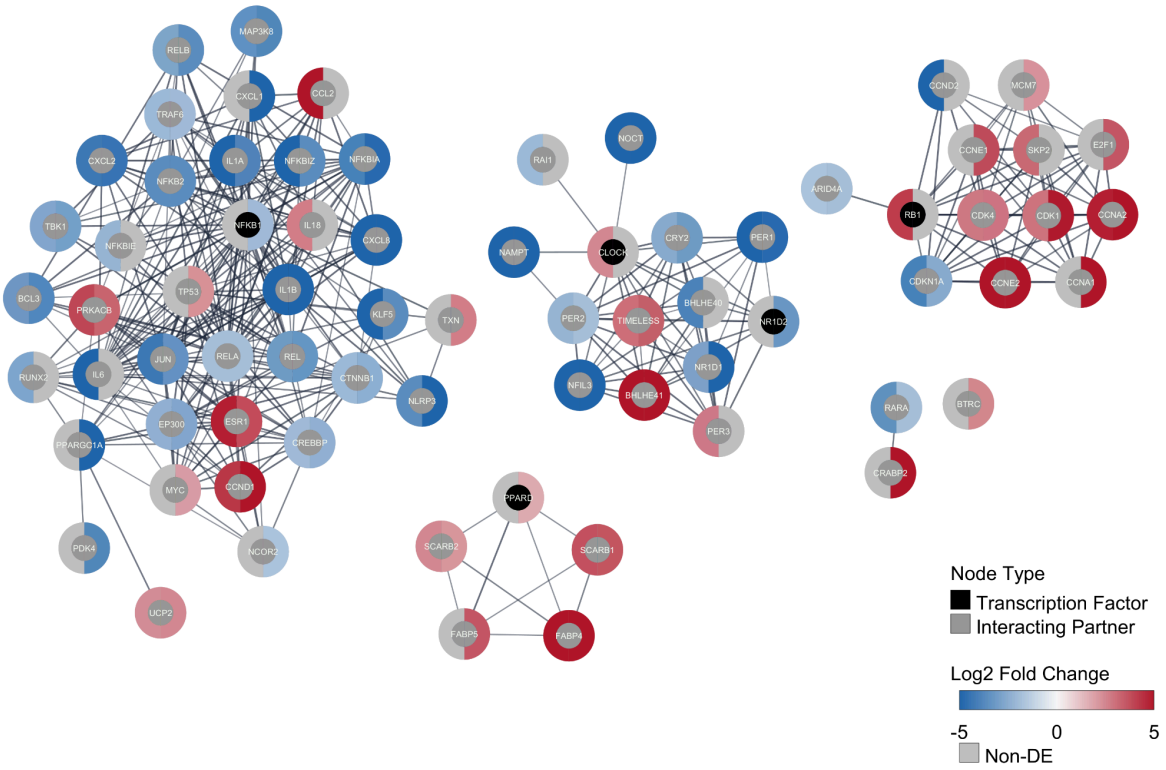

Figure 3.12 STRING network of select transcription factors involved in the regulation of inflammatory response. Nodes represent differentially expressed genes. Nodes coloured black represent the selected transcription factors, while grey nodes represent interacting partners. Nodes are connected via edges, the thickness of which is proportional to confidence of the protein-protein functional interaction. Each node is surrounded by a ring representing the Log2 fold change the gene experienced under each treatment, with the left hand half representing M-CSF and the right hand half representing GM-CSF. In cases where one of these halves are grey, it indicates that the gene was not differentially expressed in the corresponding treatment (significance thresholds being an adjusted p-value < 0.05 and a LFC > |1|).

Chapter 4: Discussion

M-CSF and GM-CSF are known to induce differentiation in monocytes, as well as distinctly priming the resultant macrophages. This study aimed to identify the distinct changes in gene expression which occur as a result of these two treatments. Differential expression analysis identified the genes which experienced significant change in expression correlated with the two treatments. Over-representation analysis was then used to determine the GO terms which were statistically overrepresented within the two lists of DEGs. Of the large number of biological processes which were overrepresented, myeloid cell differentiation and the regulation of inflammatory response were of particular interest in addressing the aforementioned aim. The genes involved in these processes were examined, in particular, those whose expression changed in different ways across the two treatments. Finally, potential transcription factors were identified, along with their associated interacting partners, in an attempt to further characterise distinct expression changes.

4.1 Marker gene expression changes

The examination of 30 known M1 and M2 macrophage markers revealed that M-CSF-induced macrophages did not exhibit a definitive M2-like gene expression profile and GM-CSF-induced macrophages did not exhibit a definitive M1-like gene expression profile. The M1 marker genes displayed a distinct lack of significant upregulation, while significant downregulation occurred for *CXCL2*, *IL1B*, *IL6*, *NFKB1*, *NFKB2*, and *PELI1*. This indicates an avoidance of the M1-like gene expression profile in macrophages produced by both M-CSF and GM-CSF. Conversely, the M2 marker genes *CCL22*, *CD163*, *CD209*, *FOLR2*, *IGF1*, *MSR1*, and *PDCD1LG2* experienced upregulation, particularly associated with M-CSF. Where M-CSF-induced macrophages experienced downregulation of M2 markers, such was the case with *IL10* and *PPARD*, this downregulation was not significant. Three M2 markers experienced downregulation in association with GM-CSF. *IGF1* and *IL10* were significantly downregulated and *FOLR2* was not.

The lack of definitive polarisation towards either M1 or M2 was further evident in those marker genes that were not differentially expressed. For example, essential M1 subtype genes encoding enzymes and cell surface markers such as *NOS2*, *CD80*, *CD86*, and *TNF* were not differentially expressed. A similar lack of commitment to M2-like behaviour was observed in the lack of significant change in *ARG1*, *IL4R*, and *CLEC10A*. Although limited conclusions can be drawn from the lack of differential expression of known marker genes, this lack of differential expression was in itself significant.

Certain M1 and M2 marker genes demonstrated a unique expression change for one or more treatments. *PPARD* was the only M2 marker gene to be significantly upregulated in GM-CSF and not differentially expressed for M-CSF. Loss of *PPARD* expression has been associated with a progression towards an M1 phenotype and increased expression of inflammatory genes such as *IL1B* (Chawla, 2010). Thus, increased expression of *PPARD* in GM-CSF-induced macrophages is incongruous with the expected progression towards M1 polarisation. This may indicate a tempering of inflammatory behaviour in unstimulated GM-CSF-induced macrophages, and will be discussed further in section 4.4. *IGF1* experienced an inverse change in expression, being significantly upregulated in M-CSF and significantly downregulated in GM-CSF. *IGF1* expression is strongly upregulated in IL-4-treated murine macrophages and is crucial in tissue repair mediated by M2 macrophages (Spadaro *et al*., 2017). These results demonstrate that, even in the absence of macrophage stimulation, *IGF1* expression may be distinct between M-CSF and GM-CSF-induced macrophages. Finally, *NFKB1* was significantly downregulated in GM-CSF but not differentially expressed in M-CSF, the implications of which are discussed in section 4.4.

While numerous gene expression changes were observed that did not correspond with the expected M1/M2 paradigm, this was not universal. Potentially the most M1-like behaviour exhibited by GM-CSF macrophages was the downregulation of *IL10*, which is well documented within inflammatory macrophages (Iyer and Cheng, 2012).

Although there were both expected and unexpected patterns of marker gene expression, these results indicate that neither M-CSF nor GM-CSF, in isolation of activating molecules, caused commitment towards their respective M2 and M1 polarisation states.

It should be noted that the lack of polarisation may belie a discrepancy in the ability of macrophages to undergo polarisation depending on the means by which they were stimulated. M-CSF-treated macrophages may more readily polarise towards an M2 state, while the same may be true of GM-CSF-treated macrophages and an M1 state. This was further investigated through an examination of the genes associated with the myeloid cell differentiation and regulation of inflammatory response GO terms.

4.2 Inversely regulated differentially expressed genes

139 and 124 differentially expressed genes were identified in association with myeloid cell differentiation and the regulation of inflammatory response, respectively. It was noted, however, that among these differentially expressed genes, only three were inversely differentially expressed. These were *INHBA*, involved in myeloid cell differentiation, and *CCR2* and *IGF1*, involved in the regulation of inflammatory response. The expression changes of these genes represent clear points to distinguish between M-CSF and GM-CSF induced macrophages.

INHBA expression was downregulated in M-CSF upregulated in GM-CSF. The *INHBA* gene encodes the βA-subunit of the activin/inhibin protein complexes. These complexes fall under the transforming growth factor-beta (TGF-β) superfamily of proteins, which represent important inflammatory cytokines in a wide range of cells. Certain tissue-specific macrophages have been shown to utilise TGF-β as a means of downregulating pro-inflammatory cytokine production (Smythies *et al*., 2005). The observed expression changes in *INHBA* therefore appear to contradict the expected inflammatory paradigm of M-CSF and GM-CSF macrophages. However, TGF-β activity is known to be context-dependent, with blood monocytes and macrophages increasing inflammatory activity in its presence (Maekawa *et al*., 2022). Thus, the observed *INHBA* expression changes may indicate the context-dependent potential for macrophages to be pro- and anti-inflammatory in association with GM-CSF and M-CSF, respectively.

CCR2 expression was downregulated in M-CSF and upregulated in GM-CSF. CCR2 enables chemotaxis in monocytes and macrophages through binding of monocyte chemoattractant protein-1 (MCP-1), which is encoded by the *CCL2* gene (Tsou *et al.*, 2007). The resultant signalling cascade enables infiltration of sites of infection, where MCP-1 is prevalent. CCR2 deficiency has been shown to impair wound healing through disruption of inflammation initiation in murine models (Boniakowski *et al*., 2018). The *CCR2*/*CCL2* axis of control has been shown to be affected uniquely by both GM-CSF and M-CSF and enable the polarisation towards their respective M1 and M2 states (Sierra-Filardi *et al*., 2014). Increased *CCR2* expression associated with GM-CSF therefore indicates a proclivity towards macrophage recruitment and initiation of inflammation. The inverse is likely true, as M-CSF macrophages are potentially less sensitive to MCP-1, preventing them from being recruited in early infection and enabling the initial inflammatory response to develop.

IGF1 expression was upregulated in M-CSF and downregulated in GM-CSF. As previously mentioned, IGF1 is a traditional M2 macrophage subtype marker. IGF1 has a broad range of effects including cell growth, proliferation, metabolism, and tissue repair and regeneration (Spadaro *et al*., 2017). Early transcriptomic examinations of macrophage polarisation revealed a ~6 fold difference in the expression of IGF1 between M1 and M2 macrophages (Martinez *et al*., 2006). This difference has been determined to be less extreme between M0 and M2 macrophages (Spadaro *et al.*, 2017). These findings therefore suggest IGF1 as a potential early indicator of polarisation in M-CSF and GM-CSF-induced macrophages.

This study has highlighted that *INHBA, CCR2* and *IGF1* may be useful as markers of differentiation in MDMs, with the potential to distinguish macrophages induced by M-CSF and GM-CSF. Further research is required to validate the expression levels of these genes and the resultant proteins in M-CSF and GM-CSF-induced MDMs.

4.3 Transcription factors involved in myeloid cell differentiation

Four uniquely differentially expressed transcription factors belonging to the myeloid cell differentiation GO term were identified and deemed to be of interest.

The *MYC* gene encodes the c-MYC transcription factor which binds specific DNA sequences through a basic-helix-loop-helix motif (Blackwell *et al.*, 1990). Dysregulation of c-MYC is typically associated with tumorigenesis due to its important role in cellular metabolism and proliferation (Dang, 2013). In macrophages, c-MYC has been shown to enable M2 polarisation through regulating the expression of key anti-inflammatory genes such as *STAT6* and *PPARG* (Pello *et al*., 2012). However, neither *STAT6* nor *PPARG* were significantly differentially expressed. The upregulation of *MYC* by GM-CSF may therefore present an explanation for the lack of M1-like gene expression. The lack of differential expression of *MYC* in M-CSF may be the result of expression levels returning to baseline following differentiation.

MAFB expression was significantly downregulated in association with GM-CSF treatment. MAFB, in conjunction with MAF, plays an extensive role in regulating differentiation of numerous cell lineages (Eychène *et al*., 2008). In the context of macrophage biology, both MAFB and MAF drive M2-like gene expression (Cuevas *et al*., 2017; Kang *et al*., 2017). MAFB upregulation in M-CSF and GM-CSF polarised macrophages has been shown to increase expression of genes such as *IL10* and *CCL2*, correlated with severe COVID-19 pathogenesis in the lung (Simón-Fuentes *et al*., 2023). Therefore, the downregulation of *MAFB* by GM-CSF would indicate a suppression of these M2-like characteristics. However, MAF expression experienced significant upregulation in M-CSF and GM-CSF. Given that these transcription factors regulate similar gene sets, this could present a mechanism by which cell-fate commitment is being modulated in these inactivated macrophages.

Both *SPI1* and *RUNX1* were downregulated in association with M-CSF treatment, but not GM-CSF. The *SPI1* gene encodes the PU.1 transcription factor, an essential regulator of haematopoiesis (Gupta *et al*., 2009). The *RUNX1* gene is similarly essential, with normal haematopoiesis being disrupted in deficient murine models (Okuda *et al*., 1996). High levels of PU.1 are essential for myeloid cell differentiation, including

monocyte-to-macrophage differentiation (Li *et al*., 2020). In macrophages, PU.1 has been shown to upregulate *CSF1R*, thereby controlling macrophage behaviour and sensitivity to M-CSF (Celada *et al*., 1996). *CSF1R* was highly upregulated in association with M-CSF, and to a lesser extent in GM-CSF. When taken in combination, this may imply a return to baseline expression levels of *SPI1* and *RUNX1* in M-CSF-induced macrophages, with the transcription factors having completed their activity in those cells. Conversely, GM-CSF-induced macrophages did not significantly downregulate *SPI1* or *RUNX1* expression. Murine models lacking PU.1 activity exhibited impaired inflammatory gene expression (Karpurapu *et al.*, 2011). The lack of change in *SPI1* expression may therefore indicate an increased capability to rapidly alter expression of inflammatory genes.

4.4 Transcription factors involved in the regulation of inflammatory response

Five uniquely differentially expressed transcription factors belonging to the regulation of inflammatory response GO term were identified and deemed to be of interest.

The *NFKB1* gene was uniquely downregulated in GM-CSF induced macrophages. *NFKB1* encodes one of the subunits for the NF-κB transcription factor (Sen and Baltimore, 1986). The NF-κB complex acts as a master regulator of many inflammation-related genes, and its activity is tightly regulated by a number of other proteins (Miraghazadeh and Cook, 2018; Mussbacher *et al.*, 2023). However, *NFKB1* is merely one of a family of five genes which contribute to the NF-κB dimer, the others being *NFKB2*, *REL*, *RELA*, and *RELB* (Pasparakis *et al*., 2006). Furthermore, numerous inhibitors of NF-κB exist, which bind dimers within the cytoplasm and prevent translocation to the nucleus, termed the inhibitors of kappa B kinases (IKKs; Miraghazadeh and Cook, 2018).

An examination of the protein-protein interaction network reveals that many of the genes responsible for regulating NF-κB activity were differentially expressed, including the IKK genes *NFKBIA*, *NFKBIZ*, and *NFKBIE*. These genes were almost exclusively

downregulated across both treatments, with the exception of *NFKBIE*, which was only downregulated in M-CSF-induced macrophages. However, it must be noted that the complexities of NF-κB regulation and activity are not fully captured by transcriptomics alone, as regulation of NF-κB-target genes does not directly correlate with the level of expression of *NFKB1* (Mussbacher *et al*., 2023). Nevertheless, these results suggest large-scale changes in the activity of NF-κB which may be halting the progression towards an M1-like phenotype in GM-CSF-induced macrophages.

The CLOCK and REV-ERBβ transcription factors are encoded by the *CLOCK* and *NR1D2* genes, respectively, and regulate gene expression in accordance with the circadian rhythm (Ikeda *et al*., 2019). *CLOCK* upregulation was only observed in M-CSF-induced macrophages, while *NR1D2* was downregulated in GM-CSF-induced macrophages. *CLOCK* expression naturally fluctuates over 24-hour cycles through negative feedback loops with PER and CRY. In macrophages, CLOCK regulates the expression of genes involved in antigen presentation, phagocytosis, and general immune function (Keller *et al.*, 2009). The deletion of *PER1* and *PER2* in murine models has been shown to promote an M1-like phenotype in macrophages (Xu *et al*., 2014). It is difficult to say whether or not the downregulation of *CRY2*, *NR1D2*, *PER1*, and *PER2* and the upregulation of *CLOCK* is due to the natural rhythm or if it may be the result of differentiation. These results point to *CLOCK* and other circadian rhythm genes as potentially playing a role in macrophage polarisation. Differences in the way this system is regulated may contribute to the differences in polarisation between M-CSF and GM-CSF-induced macrophages.

RB1, *E2F1*, and their associated genes are all involved in cell-cycle regulation. *RB1* has long been known as a tumour suppressor gene given its ability to halt cell-cycle progression. In recent years, RB1 has also been implicated in innate immune activity (Hutcheson *et al*., 2015). Conversely, E2F1 controls the transition from the growth to synthesis phases of the cell cycle (Weinberg, 1995). *RB1* expression was upregulated for M-CSF, while *E2F1* expression was upregulated for GM-CSF. These results suggest that GM-CSF-induced macrophages were engaged in cell cycle progression, while M-CSF-induced macrophages were not. This is at odds with the expected results of these treatments, as M-CSF is known to induce a higher degree of proliferation than

GM-CSF (Draijer *et al*., 2019). This may be explained by the time at which cells were sequenced and the increased expression of *CSF1* by GM-CSF-induced macrophages. Cells were sequenced after seven days, allowing for M-CSF-treated macrophages to complete their differentiation. However, the increased expression of *CSF1* in GM-CSF-induced macrophages may have caused an increase in proliferation closer to the time at which they were sequenced. Nonetheless, these differences in cell cycle control further distinguish the macrophages induced by the two treatments.

GM-CSF-induced macrophages showed an increase in *PPARD* expression. The *PPARD* gene encodes the PPARδ transcription factor, a member of the broader PPAR transcription factor family which is known to affect macrophage polarisation and inflammatory behaviour (Toobian *et al*., 2021). Unlike PPARγ, PPARδ's role in contributing towards inflammation is poorly understood (Toobian *et al*., 2021). However, murine models have demonstrated the ability of PPARδ to decrease M1-related cytokines such as IL-6, IL-1β, TNFα, and NF-κB. The *SCARB1*/*2* and *FABP4*/*5* genes with which PPARδ interacts are all involved in lipid metabolism, which is differently regulated by M1 and M2 macrophages (Remmerie and Scott, 2018). There is evidence to suggest that PPARδ enables a switch from the M1 to the M2 phenotype through changes in lipid metabolism (Schug and Li, 2009). Therefore, the upregulation of *PPARD* may be further evidence of a move away from the M1 polarisation state in GM-CSF-induced macrophages.

It is clear that all nine of the transcription factors discussed above play a role, to one extent or another, in macrophage polarisation. In particular, the unexpected expression changes observed in *MYC*, *NFKB1*, and *PPARD* may underlie the lack of M1-like polarisation in GM-CSF-induced macrophages. Furthermore, the lack of differential expression in certain transcription factors in M-CSF-induced macrophages may also underlie the expected lack of polarisation which is characteristic of the treatment.

4.5 Study limitations

No true biological or technical replicates were present for any of the nine samples, further necessitating the pooling of similar samples across donors. The introduction of technical replicates is typically not feasible due to the expense in producing RNA-seq datasets, and is thus an inherent limitation within the field. However, the tools used in differential gene expression analysis take these limitations into account.

An important consideration is the loss of resolution when performing ORA due to the need to convert between gene identifier types. For example, of the 28 340 genes that formed the background set for ORA, only 15 025 had matching annotation data within the GO database. The number of DEGs for M-CSF and GM-CSF were 4 072 and 4 399, respectively. However these numbers were reduced to 3 425 and 3 634 during the ID conversion process. This reduction in features is unavoidable, as this is an inherent limitation of the databases. However, the impact of this limitation is potentially mitigated by the fact that those genes for which ID conversion was unsuccessful are likely poorly characterised as a result of less functional relevance. Therefore, they may be less important for this study and the characterisation of monocyte-to-macrophage differentiation.

When examining DGE analysis results, it is important to consider that a lack of differential expression does not equate with a lack of expression. The corollary of this can also be true, where differential expression does not directly equate with the magnitude of gene expression. Furthermore, this change is measured only at two given time points, and does not capture nuanced fluctuations in gene expression between time points. These factors limit the breadth of applicability of the conclusions of studies such as this.

4.6 Concluding remarks

This study aimed to identify markers of differentiation in MDMs induced with M-CSF or GM-CSF. A large number of differentially expressed genes were identified, many of which were shared across both treatments. An examination of the enriched biological processes based on the sets of differentially expressed genes, in conjunction with known macrophage marker genes revealed that the resultant macrophages lacked changes in expression of genes commonly associated with M1 and M2 polarisation states. Interrogation of important biological processes revealed three of the commonly differentially expressed genes to be inversely regulated by the two treatments, namely *CCR2*, *IGF1* and *INHBA*. These genes may be of interest in the characterisation of unstimulated MDMs following their differentiation. Furthermore, nine uniquely differentially expressed transcription factors involved in key biological processes were identified, each of which may be contributing to the lack of complete polarisation following differentiation. Specifically, the downregulation of *NFKB1* and upregulation of *MYC* and *PPARD* may be contributing to an absence of inflammatory markers in GM-CSF-induced macrophages. Ultimately, these results reveal that M-CSF and GM-CSF-induced macrophages, in the absence of activation, experience highly similar gene expression changes and a distinct lack of change in key polarisation marker genes.

4.7 Future work

This study identified *CCR2*, *IGF1*, and *INHBA* as genes whose expression changes distinguished M-CSF and GM-CSF-induced macrophages relative to monocytes. Experimental validation of the expression levels of these genes through techniques such as RT-qPCR could strengthen these conclusions. Furthermore, the findings of this study could benefit from the integration of other data types, including proteomic and metabolomic data. Integrating RNA-seq data from intermediary time points (between zero and seven days) could also improve understanding of the way in which MDMs develop from their monocyte precursors in early differentiation.

Supplementary Tables

Table S1 M-CSF differentially expressed genes. S1_MCSF_DEGs

Table S2 GM-CSF differentially expressed genes. S2_GMCSF_DEGs

Table S3 M-CSF enriched GO terms. S3_MCSF_GO. Within tables S3 and S4, ID indicates the identifier for the GO term. Description shows the description for the GO term of that row. Gene ratio shows the number of DEGs that belong to that GO terms, divided by the total number of DEGs. Bg ratio represents the total number of genes from the background set annotated to that GO term, divided by the total number of genes supplied as the background. P value and p adjusted are metrics for significance.

Table S4 GM-CSF enriched GO terms. S4_GMCSF_GO

Table S5 Potential transcription factors differentially expressed for M-CSF. S5_MCSF_DETFs

Table S6 Potential transcription factors differentially expressed for GM-CSF. S6_GMCSF_DETFs

Table S7 Transcription factors associated with GO terms of interest. S7_GO_DETFs

References

Ambarus, C.A., Krausz, S., van Eijk, M., Hamann, J., Radstake, T.R.D.J., Reedquist, K.A., Tak, P.P., Baeten, D.L.P., 2012. Systematic validation of specific phenotypic markers for in vitro polarized human macrophages. Journal of Immunological Methods 375, 196–206. https://doi.org/10.1016/j.jim.2011.10.013

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biology 11, R106. https://doi.org/10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P.T., Huber, W., 2014. HTSeq – A Python framework to work with high-throughput sequencing data. https://doi.org/10.1101/002824

Andrews, S., 2010. FastQC: A Quality Control tool for High Throughput Sequence Data [WWW Document]. URL https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 3.24.20).

Andrews, T.S., Hemberg, M., 2018. Identifying cell populations with scRNASeq. Mol. Aspects Med., The emerging field of single-cell analysis 59, 114–122. https://doi.org/10.1016/j.mam.2017.07.002

Arnold, L., Henry, A., Poron, F., Baba-Amer, Y., van Rooijen, N., Plonquet, A., Gherardi, R.K., Chazaud, B., 2007. Inflammatory monocytes recruited after skeletal muscle injury switch into antiinflammatory macrophages to support myogenesis. J Exp Med 204, 1057–1069. https://doi.org/10.1084/jem.20070075

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29. https://doi.org/10.1038/75556

Austermann, J., Roth, J., Barczyk-Kahlert, K., 2022. The Good and the Bad: Monocytes' and Macrophages' Diverse Functions in Inflammation. Cells 11, 1979. https://doi.org/10.3390/cells11121979

Azad, A.K., Rajaram, M.V.S., Schlesinger, L.S., 2014. Exploitation of the Macrophage Mannose Receptor (CD206) in Infectious Disease Diagnostics and Therapeutics. J Cytol Mol Biol 1, 1000003. https://doi.org/10.13188/2325-4653.1000003

Basilico, S., Göttgens, B., 2017. Dysregulation of haematopoietic stem cell regulatory programs in acute myeloid leukaemia. J. Mol. Med. 95, 719–727. https://doi.org/10.1007/s00109-017-1535-3

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 57, 289–300.

Benoit, M., Desnues, B., Mege, J.-L., 2008. Macrophage Polarization in Bacterial Infections. The Journal of Immunology 181, 3733–3739. https://doi.org/10.4049/jimmunol.181.6.3733

Bertani, F.R., Mozetic, P., Fioramonti, M., Iuliani, M., Ribelli, G., Pantano, F., Santini, D., Tonini, G., Trombetta, M., Businaro, L., Selci, S., Rainer, A., 2017. Classification of M1/M2-polarized human macrophages by label-free hyperspectral reflectance confocal microscopy and multivariate analysis. Sci Rep 7, 8965. https://doi.org/10.1038/s41598-017-08121-8

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer New York.

Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N., Weintraub, H., 1990. Sequence-Specific DNA Binding by the c-Myc Protein. Science 250, 1149–1151. https://doi.org/10.1126/science.2251503

Boniakowski, A.E., Kimball, A.S., Joshi, A., Schaller, M., Davis, F.M., denDekker, A., Obi, A.T., Moore, B.B., Kunkel, S.L., Gallagher, K.A., 2018. Macrophage chemokine receptor CCR2 plays a crucial role in macrophage recruitment and regulated inflammation in wound healing. Eur J Immunol 48, 1445–1455. https://doi.org/10.1002/eji.201747400

Bonifer, C., Hume, D.A., 2008. The transcriptional regulation of the Colony-Stimulating Factor 1 Receptor (csf1r) gene during hematopoiesis. Front Biosci 13, 549–560. https://doi.org/10.2741/2700

Bouhlel, M.A., Derudas, B., Rigamonti, E., Dièvart, R., Brozek, J., Haulon, S., Zawadzki, C., Jude, B., Torpier, G., Marx, N., Staels, B., Chinetti-Gbaguidi, G., 2007. PPARgamma activation primes human monocytes into alternative M2 macrophages with anti-inflammatory properties. Cell Metab 6, 137–143. https://doi.org/10.1016/j.cmet.2007.06.010

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 34, 525–527. https://doi.org/10.1038/nbt.3519

Cartwright, T., Perkins, N.D., L. Wilson, C., 2016. NFKB1: a suppressor of inflammation, ageing and cancer. The FEBS Journal 283, 1812–1822. https://doi.org/10.1111/febs.13627

Celada, A., Borràs, F.E., Soler, C., Lloberas, J., Klemsz, M., van Beveren, C., McKercher, S., Maki, R.A., 1996. The transcription factor PU.1 is involved in macrophage proliferation. Journal of Experimental Medicine 184, 61–69. https://doi.org/10.1084/jem.184.1.61

Chawla, A., 2010. Control of macrophage activation and function by PPARs. Circ Res 106, 1559–1569. https://doi.org/10.1161/CIRCRESAHA.110.216523

Chen, L., Deng, H., Cui, H., Fang, J., Zuo, Z., Deng, J., Li, Y., Wang, X., Zhao, L., 2017. Inflammatory responses and inflammation-associated diseases in organs. Oncotarget 9, 7204–7218. https://doi.org/10.18632/oncotarget.23208

Chen, S., Fuller, K.K., Dunlap, J.C., Loros, J.J., 2020. A Pro- and Anti-inflammatory Axis Modulates the Macrophage Circadian Clock. Frontiers in Immunology 11.

Chen, S., Saeed, A.F.U.H., Liu, Q., Jiang, Q., Xu, H., Xiao, G.G., Rao, L., Duo, Y., 2023. Macrophages in immunoregulation and therapeutics. Signal Transduct Target Ther 8, 207. https://doi.org/10.1038/s41392-023-01452-1

Chistiakov, D.A., Bobryshev, Y.V., Orekhov, A.N., 2015. Changes in transcriptome of macrophages in atherosclerosis. J Cell Mol Med 19, 1163–1173. https://doi.org/10.1111/jcmm.12591

Chistiakov, D.A., Killingsworth, M.C., Myasoedova, V.A., Orekhov, A.N., Bobryshev, Y.V., 2017. CD68/macrosialin: not just a histochemical marker. Lab Invest 97, 4–13. https://doi.org/10.1038/labinvest.2016.116

Chun, N., Heeger, P.S., 2016. Blurring the Lines between Innate and Adaptive Immunity. Transplantation 100, 1789–1790. https://doi.org/10.1097/TP.0000000000001429

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. Genome Biology 17, 13.

https://doi.org/10.1186/s13059-016-0881-8
Cook, R.D., 1977. Detection of Influential Observation in Linear Regression. Technometrics.
Cuevas, V.D., Anta, L., Samaniego, R., Orta-Zavalza, E., Vladimir de la Rosa, J., Baujat, G., Domínguez-Soto, Á., Sánchez-Mateos, P., Escribese, M.M., Castrillo, A., Cormier-Daire, V., Vega, M.A., Corbí, Á.L., 2017. MAFB Determines Human Macrophage Anti-Inflammatory Polarization: Relevance for the Pathogenic Mechanisms Operating in Multicentric Carpotarsal Osteolysis. The Journal of Immunology 198, 2070–2081. https://doi.org/10.4049/jimmunol.1601667
Dang, C.V., 2013. MYC, Metabolism, Cell Growth, and Tumorigenesis. Cold Spring Harb Perspect Med 3, a014217. https://doi.org/10.1101/cshperspect.a014217
Das, S., Rai, A., Rai, S.N., 2022. Differential Expression Analysis of Single-Cell RNA-Seq Data: Current Statistical Approaches and Outstanding Challenges. Entropy (Basel) 24, 995. https://doi.org/10.3390/e24070995
Davies, L.C., Jenkins, S.J., Allen, J.E., Taylor, P.R., 2013. Tissue-resident macrophages. Nat Immunol 14, 986–995. https://doi.org/10.1038/ni.2705
de Bruijn, M., Dzierzak, E., 2017. Runx transcription factors in the development and function of the definitive hematopoietic system. Blood 129, 2061–2069. https://doi.org/10.1182/blood-2016-12-689109
Deshpande, D., Muszyńska, A., Zhao, F., P. Łabaj, P., Mangul, S., 2023. RNA-seq data science: From raw data to effective interpretation. Front. Genet. 14. https://doi.org/10.3389/fgene.2023.997383
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635
Doncheva, N.T., Morris, J.H., Gorodkin, J., Jensen, L.J., 2019. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J. Proteome Res. 18, 623–632. https://doi.org/10.1021/acs.jproteome.8b00702
Doyle, A.G., Herbein, G., Montaner, L.J., Minty, A.J., Caput, D., Ferrara, P., Gordon, S., 1994. Interleukin-13 alters the activation state of murine macrophages in vitro: comparison with interleukin-4 and interferon-gamma. Eur J Immunol 24, 1441–1445. https://doi.org/10.1002/eji.1830240630
Draijer, C., Penke, L.R.K., Peters-Golden, M., 2019. Distinctive Effects of GM-CSF and M-CSF on Proliferation and Polarization of Two Major Pulmonary Macrophage Populations. J Immunol 202, 2700–2709. https://doi.org/10.4049/jimmunol.1801387
Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., Huber, W., 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21, 3439–3440. https://doi.org/10.1093/bioinformatics/bti525
Ehrchen, J., Steinmüller, L., Barczyk, K., Tenbrock, K., Nacken, W., Eisenacher, M., Nordhues, U., Sorg, C., Sunderkötter, C., Roth, J., 2007. Glucocorticoids induce differentiation of a specifically activated, anti-inflammatory subtype of human monocytes. Blood 109, 1265–1274. https://doi.org/10.1182/blood-2006-02-001115
Eligini, S., Crisci, M., Bono, E., Songia, P., Tremoli, E., Colombo, G.I., Colli, S., 2013. Human monocyte-derived macrophages spontaneously differentiated in vitro

show distinct phenotypes. Journal of Cellular Physiology 228, 1464–1472. https://doi.org/10.1002/jcp.24301
Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354
Eychène, A., Rocques, N., Pouponnot, C., 2008. A new MAFia in cancer. Nat Rev Cancer 8, 683–693. https://doi.org/10.1038/nrc2460
Fleetwood, A.J., Lawrence, T., Hamilton, J.A., Cook, A.D., 2007. Granulocyte-macrophage colony-stimulating factor (CSF) and macrophage CSF-dependent macrophage phenotypes display differences in cytokine profiles and transcription factor activities: implications for CSF blockade in inflammation. J Immunol 178, 5245–5252. https://doi.org/10.4049/jimmunol.178.8.5245
Fujiwara, N., Kobayashi, K., 2005. Macrophages in inflammation. Curr Drug Targets Inflamm Allergy 4, 281–286. https://doi.org/10.2174/1568010054022024
Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 5, R80. https://doi.org/10.1186/gb-2004-5-10-r80
Goerdt, S., Politz, O., Schledzewski, K., Birk, R., Gratchev, A., Guillot, P., Hakiy, N., Klemke, C.D., Dippel, E., Kodelja, V., Orfanos, C.E., 1999. Alternative versus classical activation of macrophages. Pathobiology 67, 222–226. https://doi.org/10.1159/000028096
Gomez Perdiguero, E., Klapproth, K., Schulz, C., Busch, K., Azzoni, E., Crozet, L., Garner, H., Trouillet, C., de Bruijn, M.F., Geissmann, F., Rodewald, H.-R., 2015. Tissue-resident macrophages originate from yolk sac-derived erythro-myeloid progenitors. Nature 518, 547–551. https://doi.org/10.1038/nature13989
Gong, M., Zhuo, X., Ma, A., 2017. STAT6 Upregulation Promotes M2 Macrophage Polarization to Suppress Atherosclerosis. Med Sci Monit Basic Res 23, 240–249. https://doi.org/10.12659/MSMBR.904014
González de la Aleja, A., Herrero, C., Torres-Torresano, M., de la Rosa, J.V., Alonso, B., Capa-Sardón, E., Muller, I.B., Jansen, G., Puig-Kröger, A., Vega, M.A., Castrillo, A., Corbí, Á.L., 2022. Activation of LXR Nuclear Receptors Impairs the Anti-Inflammatory Gene and Functional Profile of M-CSF-Dependent Human Monocyte-Derived Macrophages. Frontiers in Immunology 13.
Gudgeon, J., Marín-Rubio, J.L., Trost, M., 2022. The role of macrophage scavenger receptor 1 (MSR1) in inflammatory disorders and cancer. Front. Immunol. 13. https://doi.org/10.3389/fimmu.2022.1012002
Günthner, R., Anders, H.-J., 2013. Interferon-Regulatory Factors Determine Macrophage Phenotype Polarization. Mediators Inflamm 2013, 731023. https://doi.org/10.1155/2013/731023
Gupta, P., Gurudutta, G.U., Saluja, D., Tripathi, R.P., 2009. PU.1 and partners: regulation of haematopoietic stem cell fate in normal and malignant haematopoiesis. Journal of Cellular and Molecular Medicine 13, 4349–4363. https://doi.org/10.1111/j.1582-4934.2009.00757.x
Hamilton, J.A., 2008. Colony-stimulating factors in inflammation and autoimmunity.

Nat Rev Immunol 8, 533–544. https://doi.org/10.1038/nri2356
Hamon, M.A., Quintin, J., 2016. Innate immune memory in mammals. Semin. Immunol., SI: Innate Immune Memory 28, 351–358. https://doi.org/10.1016/j.smim.2016.05.003
Hume, D.A., Pavli, P., Donahue, R.E., Fidler, I.J., 1988. The effect of human recombinant macrophage colony-stimulating factor (CSF-1) on the murine mononuclear phagocyte system in vivo. J Immunol 141, 3405–3409.
Hutcheson, J., Witkiewicz, A.K., Knudsen, E.S., 2015. The RB tumor suppressor at the intersection of proliferation and immunity: relevance to disease immune evasion and immunotherapy. Cell Cycle 14, 3812–3819. https://doi.org/10.1080/15384101.2015.1010922
Ikeda, R., Tsuchiya, Y., Koike, N., Umemura, Y., Inokawa, H., Ono, R., Inoue, M., Sasawaki, Y., Grieten, T., Okubo, N., Ikoma, K., Fujiwara, H., Kubo, T., Yagita, K., 2019. REV-ERBα and REV-ERBβ function as key factors regulating Mammalian Circadian Output. Sci Rep 9, 10171. https://doi.org/10.1038/s41598-019-46656-0
Iyer, S.S., Cheng, G., 2012. Role of Interleukin 10 Transcriptional Regulation in Inflammation and Autoimmune Disease. Crit Rev Immunol 32, 23–63.
Jaguin, M., Houlbert, N., Fardel, O., Lecureur, V., 2013. Polarization profiles of human M-CSF-generated macrophages and comparison of M1-markers in classically activated macrophages from GM-CSF and M-CSF origin. Cell Immunol 281, 51–61. https://doi.org/10.1016/j.cellimm.2013.01.010
Jakubzick, C.V., Randolph, G.J., Henson, P.M., 2017. Monocyte differentiation and antigen-presenting functions. Nat. Rev. Immunol. 17, 349–362. https://doi.org/10.1038/nri.2017.28
Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28, 27–30. https://doi.org/10.1093/nar/28.1.27
Kang, Kyuho, Park, S.H., Chen, J., Qiao, Y., Giannopoulou, E., Berg, K., Hanidu, A., Li, J., Nabozny, G., Kang, Keunsoo, Park-Min, K.-H., Ivashkiv, L.B., 2017. Interferon-γ Represses M2 Gene Expression in Human Macrophages by Disassembling Enhancers Bound by the Transcription Factor MAF. Immunity 47, 235-250.e4. https://doi.org/10.1016/j.immuni.2017.07.017
Kapellos, T.S., Bonaguro, L., Gemünd, I., Reusch, N., Saglam, A., Hinkley, E.R., Schultze, J.L., 2019. Human Monocyte Subsets and Phenotypes in Major Chronic Inflammatory Diseases. Front. Immunol. 10. https://doi.org/10.3389/fimmu.2019.02035
Karpurapu, M., Wang, X., Deng, J., Park, H., Xiao, L., Sadikot, R.T., Frey, R.S., Maus, U.A., Park, G.Y., Scott, E.W., Christman, J.W., 2011. Functional PU.1 in macrophages has a pivotal role in NF-κB activation and neutrophilic lung inflammation during endotoxemia. Blood 118, 5255–5266. https://doi.org/10.1182/blood-2011-03-341123
Keller, M., Mazuch, J., Abraham, U., Eom, G.D., Herzog, E.D., Volk, H.-D., Kramer, A., Maier, B., 2009. A circadian clock in macrophages controls inflammatory immune responses. Proceedings of the National Academy of Sciences 106, 21407–21412. https://doi.org/10.1073/pnas.0906361106
Kelly, L.M., Englmeier, U., Lafon, I., Sieweke, M.H., Graf, T., 2000. MafB is an inducer of monocytic differentiation. The EMBO Journal 19, 1987. https://doi.org/10.1093/emboj/19.9.1987

Kim, Donghyun, Koh, J., Ko, J.S., Kim, H.Y., Lee, H., Chung, D.H., 2019. Ubiquitin E3 Ligase Pellino-1 Inhibits IL-10-mediated M2c Polarization of Macrophages, Thereby Suppressing Tumor Growth. Immune Netw 19, e32. https://doi.org/10.4110/in.2019.19.e32

Kim, Daehwan, Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kratochvill, F., Neale, G., Haverkamp, J.M., Van de Velde, L.-A., Smith, A.M., Kawauchi, D., McEvoy, J., Roussel, M.F., Dyer, M.A., Qualls, J.E., Murray, P.J., 2015. TNF Counterbalances the Emergence of M2 Tumor Macrophages. Cell Reports 12, 1902–1914. https://doi.org/10.1016/j.celrep.2015.08.033

Krausgruber, T., Blazek, K., Smallie, T., Alzabin, S., Lockstone, H., Sahgal, N., Hussell, T., Feldmann, M., Udalova, I.A., 2011. IRF5 promotes inflammatory macrophage polarization and TH1-TH17 responses. Nat Immunol 12, 231–238. https://doi.org/10.1038/ni.1990

Kubes, P., Mehal, W.Z., 2012. Sterile inflammation in the liver. Gastroenterology 143, 1158–1172. https://doi.org/10.1053/j.gastro.2012.09.008

Lambrecht, B.N., 2006. Alveolar macrophage in the driver's seat. Immunity 24, 366–368. https://doi.org/10.1016/j.immuni.2006.03.008

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Legeay, M., Doncheva, N.T., Morris, J.H., Jensen, L.J., 2020. Visualize omics data on networks with Omics Visualizer, a Cytoscape App. F1000Res 9, 157. https://doi.org/10.12688/f1000research.22280.2

Lehtonen, A., Ahlfors, H., Veckman, V., Miettinen, M., Lahesmaa, R., Julkunen, I., 2007. Gene expression profiling during differentiation of human monocytes to macrophages or dendritic cells. J Leukoc Biol 82, 710–720. https://doi.org/10.1189/jlb.0307194

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323. https://doi.org/10.1186/1471-2105-12-323

Li, G., Hao, W., Hu, W., 2020. Transcription factor PU.1 and immune cell differentiation (Review). International Journal of Molecular Medicine 46, 1943–1950. https://doi.org/10.3892/ijmm.2020.4763

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, W., Mei, M., Liu, T., Zhang, S., Wang, Z., Suo, Y., Wang, S., Liu, Y., Zhang, N., Lu, W., 2022. Identification of PDCD1 and PDCD1LG2 as Prognostic Biomarkers and Associated with Immune Infiltration in Hepatocellular Carcinoma. Int J Gen Med 15, 437–449. https://doi.org/10.2147/IJGM.S347163

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lisi, L., Ciotti, G.M.P., Braun, D., Kalinin, S., Currò, D., Dello Russo, C., Coli, A., Mangiola, A., Anile, C., Feinstein, D.L., Navarra, P., 2017. Expression of iNOS, CD163 and ARG-1 taken as M1 and M2 markers of microglial polarization in human glioblastoma and the surrounding normal parenchyma. Neurosci Lett 645, 106–112. https://doi.org/10.1016/j.neulet.2017.02.076

Love, M.I., Hogenesch, J.B., Irizarry, R.A., 2016. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. Nat Biotechnol 34, 1287–1291. https://doi.org/10.1038/nbt.3682
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. https://doi.org/10.1186/s13059-014-0550-8
Lurier, E.B., Dalton, D., Dampier, W., Raman, P., Nassiri, S., Ferraro, N.M., Rajagopalan, R., Sarmady, M., Spiller, K.L., 2017. Transcriptome analysis of IL-10-stimulated (M2c) macrophages by next-generation sequencing. Immunobiology 222, 847–856. https://doi.org/10.1016/j.imbio.2017.02.006
Lüscher, B., 2001. Function and regulation of the transcription factors of the Myc/Max/Mad network. Gene 277, 1–14. https://doi.org/10.1016/S0378-1119(01)00697-7
Maekawa, H., Jin, Y., Nishio, M., Kawai, S., Nagata, S., Kamakura, T., Yoshitomi, H., Niwa, A., Saito, M.K., Matsuda, S., Toguchida, J., 2022. Recapitulation of pro-inflammatory signature of monocytes with ACVR1A mutation using FOP patient-derived iPSCs. Orphanet Journal of Rare Diseases 17, 364. https://doi.org/10.1186/s13023-022-02506-3
Maglott, D., Ostell, J., Pruitt, K.D., Tatusova, T., 2007. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 35, D26–D31. https://doi.org/10.1093/nar/gkl993
Martin, F.J., Amode, M.R., Aneja, A., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S.K., Bignell, A., Boddu, S., Branco Lins, P.R., Brooks, L., Ramaraju, S.B., Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., Davidson, C., Dodiya, K., Donaldson, S., El Houdaigui, B., El Naboulsi, T., Fatima, R., Giron, C.G., Genez, T., Ghattaoraya, G.S., Martinez, J.G., Guijarro, C., Hardy, M., Hollis, Z., Hourlier, T., Hunt, T., Kay, M., Kaykala, V., Le, T., Lemos, D., Marques-Coelho, D., Marugán, J.C., Merino, G.A., Mirabueno, L.P., Mushtaq, A., Hossain, S.N., Ogeh, D.N., Sakthivel, M.P., Parker, A., Perry, M., Piližota, I., Prosovetskaia, I., Pérez-Silva, J.G., Salam, A.I.A., Saraiva-Agostinho, N., Schuilenburg, H., Sheppard, D., Sinha, S., Sipos, B., Stark, W., Steed, E., Sukumaran, R., Sumathipala, D., Suner, M.-M., Surapaneni, L., Sutinen, K., Szpak, M., Tricomi, F.F., Urbina-Gómez, D., Veidenberg, A., Walsh, T.A., Walts, B., Wass, E., Willhoft, N., Allen, J., Alvarez-Jarreta, J., Chakiachvili, M., Flint, B., Giorgetti, S., Haggerty, L., Ilsley, G.R., Loveland, J.E., Moore, B., Mudge, J.M., Tate, J., Thybert, D., Trevanion, S.J., Winterbottom, A., Frankish, A., Hunt, S.E., Ruffier, M., Cunningham, F., Dyer, S., Finn, R.D., Howe, K.L., Harrison, P.W., Yates, A.D., Flicek, P., 2023. Ensembl 2023. Nucleic Acids Research 51, D933–D941. https://doi.org/10.1093/nar/gkac958
Martinez, F.O., Gordon, S., Locati, M., Mantovani, A., 2006. Transcriptional Profiling of the Human Monocyte-to-Macrophage Differentiation and Polarization: New Molecules and Patterns of Gene Expression1. The Journal of Immunology 177, 7303–7311. https://doi.org/10.4049/jimmunol.177.10.7303
McCullagh, P., Nedler, JA., 1989. Monographs on Statistics & Applied Probability. Routledge.
Medzhitov, R., Janeway, C., 2000. Innate immune recognition: mechanisms and pathways. Immunol Rev 173, 89–97.

https://doi.org/10.1034/j.1600-065x.2000.917309.x
Metchnikoff, E., 1891. Lecture on Phagocytosis and Immunity. Br Med J 1, 213–217. https://doi.org/10.1136/bmj.1.1570.213
Mills, C.D., Kincaid, K., Alt, J.M., Heilman, M.J., Hill, A.M., 2000. M-1/M-2 Macrophages and the Th1/Th2 Paradigm1. The Journal of Immunology 164, 6166–6173. https://doi.org/10.4049/jimmunol.164.12.6166
Miraghazadeh, B., Cook, M.C., 2018. Nuclear Factor-kappaB in Autoimmunity: Man and Mouse. Front Immunol 9, 613. https://doi.org/10.3389/fimmu.2018.00613
Morris, J.H., Apeltsin, L., Newman, A.M., Baumbach, J., Wittkop, T., Su, G., Bader, G.D., Ferrin, T.E., 2011. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics 12, 436. https://doi.org/10.1186/1471-2105-12-436
Munster, S.K., Nicholson, S.J., Uyhelji, H.A., United States. Department of Transportation. Federal Aviation Administration. Office of Aviation. Civil Aerospace Medical Institute, 2023. RNA-Seq Alignment and Differential Expression Software Comparison (No. DOT/FAA/AM-23/02). https://doi.org/10.21949/1524443
Mussbacher, M., Derler, M., Basílio, J., Schmid, J.A., 2023. NF-κB in monocytes and macrophages – an inflammatory master regulator in multitalented immune cells. Front Immunol 14, 1134661. https://doi.org/10.3389/fimmu.2023.1134661
Narasimhan, P.B., Marcovecchio, P., Hamers, A.A.J., Hedrick, C.C., 2019. Nonclassical Monocytes in Health and Disease. Annu. Rev. Immunol. 37, 439–456. https://doi.org/10.1146/annurev-immunol-042617-053119
Okuda, T., Deursen, J. van, Hiebert, S.W., Grosveld, G., Downing, J.R., 1996. AML1, the Target of Multiple Chromosomal Translocations in Human Leukemia, Is Essential for Normal Fetal Liver Hematopoiesis. Cell 84, 321–330. https://doi.org/10.1016/S0092-8674(00)80986-1
Orecchioni, M., Ghosheh, Y., Pramod, A.B., Ley, K., 2019. Macrophage Polarization: Different Gene Signatures in M1(LPS+) vs. Classically and M2(LPS–) vs. Alternatively Activated Macrophages. Front. Immunol. 10. https://doi.org/10.3389/fimmu.2019.01084
Orekhov, A.N., Orekhova, V.A., Nikiforov, N.G., Myasoedova, V.A., Grechko, A.V., Romanenko, E.B., Zhang, D., Chistiakov, D.A., 2019. Monocyte differentiation and macrophage polarization. Vessel Plus 3, 10. https://doi.org/10.20517/2574-1209.2019.04
Osborne, J.D., Flatow, J., Holko, M., Lin, S.M., Kibbe, W.A., Zhu, L. (Julie), Danila, M.I., Feng, G., Chisholm, R.L., 2009. Annotating the human genome with Disease Ontology. BMC Genomics 10, S6. https://doi.org/10.1186/1471-2164-10-S1-S6
Oshlack, A., Robinson, M.D., Young, M.D., 2010. From RNA-seq reads to differential expression results. Genome Biology 11, 220. https://doi.org/10.1186/gb-2010-11-12-220
Oshlack, A., Wakefield, M.J., 2009. Transcript length bias in RNA-seq data confounds systems biology. Biology Direct 4, 14. https://doi.org/10.1186/1745-6150-4-14
Park, M.D., Silvin, A., Ginhoux, F., Merad, M., 2022. Macrophages in health and disease. Cell 185, 4259–4279. https://doi.org/10.1016/j.cell.2022.10.007
Pasparakis, M., Luedde, T., Schmidt-Supprian, M., 2006. Dissection of the NF-κB signalling cascade in transgenic and knockout mice. Cell Death Differ 13,

861–872. https://doi.org/10.1038/sj.cdd.4401870
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419. https://doi.org/10.1038/nmeth.4197
Pearson, K., 1901. LIII.On lines and planes of closest fit to systems of points in space. Philosophical Magazine Series 6. https://doi.org/10.1080/14786440109462720
Pello, O.M., De Pizzol, M., Mirolo, M., Soucek, L., Zammataro, L., Amabile, A., Doni, A., Nebuloni, M., Swigart, L.B., Evan, G.I., Mantovani, A., Locati, M., 2012. Role of c-MYC in alternative activation of human macrophages and tumor-associated macrophage biology. Blood 119, 411–421. https://doi.org/10.1182/blood-2011-02-339911
Pesce, J.T., Ramalingam, T.R., Mentink-Kane, M.M., Wilson, M.S., El Kasmi, K.C., Smith, A.M., Thompson, R.W., Cheever, A.W., Murray, P.J., Wynn, T.A., 2009. Arginase-1-expressing macrophages suppress Th2 cytokine-driven inflammation and fibrosis. PLoS Pathog 5, e1000371. https://doi.org/10.1371/journal.ppat.1000371
Pixley, F.J., Stanley, E.R., 2004. CSF-1 regulation of the wandering macrophage: complexity in action. Trends Cell Biol 14, 628–638. https://doi.org/10.1016/j.tcb.2004.09.016
Puig-Kröger, A., Sierra-Filardi, E., Domínguez-Soto, A., Samaniego, R., Corcuera, M.T., Gómez-Aguado, F., Ratnam, M., Sánchez-Mateos, P., Corbí, A.L., 2009. Folate Receptor β Is Expressed by Tumor-Associated Macrophages and Constitutes a Marker for M2 Anti-inflammatory/Regulatory Macrophages. Cancer Research 69, 9395–9403. https://doi.org/10.1158/0008-5472.CAN-09-2050
Remmerie, A., Scott, C.L., 2018. Macrophages and lipid metabolism. Cell Immunol 330, 27–42. https://doi.org/10.1016/j.cellimm.2018.01.020
Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., Pachter, L., 2011. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biology 12, R22. https://doi.org/10.1186/gb-2011-12-3-r22
Ross, E.A., Devitt, A., Johnson, J.R., 2021. Macrophages: The Good, the Bad, and the Gluttony. Front. Immunol. 12. https://doi.org/10.3389/fimmu.2021.708186
Roy, A.L., 2019. Transcriptional Regulation in the Immune System: One Cell at a Time. Front Immunol 10, 1355. https://doi.org/10.3389/fimmu.2019.01355
Rynikova, M., Adamkova, P., Hradicka, P., Stofilova, J., Harvanova, D., Matejova, J., Demeckova, V., 2023. Transcriptomic Analysis of Macrophage Polarization Protocols: Vitamin D3 or IL-4 and IL-13 Do Not Polarize THP-1 Monocytes into Reliable M2 Macrophages. Biomedicines 11, 608. https://doi.org/10.3390/biomedicines11020608
Schug, T.T., Li, X., 2009. PPARδ-mediated macrophage activation: a matter of fat. Disease Models & Mechanisms 2, 421–422. https://doi.org/10.1242/dmm.003913
Sen, R., Baltimore, D., 1986. Inducibility of kappa immunoglobulin enhancer-binding protein Nf-kappa B by a posttranslational mechanism. Cell 47, 921–928. https://doi.org/10.1016/0092-8674(86)90807-x
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13,

2498–2504. https://doi.org/10.1101/gr.1239303
Shapouri-Moghaddam, A., Mohammadian, S., Vazini, H., Taghadosi, M., Esmaeili, S.-A., Mardani, F., Seifi, B., Mohammadi, A., Afshari, J.T., Sahebkar, A., 2018. Macrophage plasticity, polarization, and function in health and disease. Journal of Cellular Physiology 233, 6425–6440. https://doi.org/10.1002/jcp.26429
Shi, C., Pamer, E.G., 2011. Monocyte recruitment during infection and inflammation. Nat Rev Immunol 11, 762–774. https://doi.org/10.1038/nri3070
Sica, A., Erreni, M., Allavena, P., Porta, C., 2015. Macrophage polarization in pathology. Cell Mol Life Sci 72, 4111–4126. https://doi.org/10.1007/s00018-015-1995-y
Sierra-Filardi, E., Nieto, C., Domínguez-Soto, Á., Barroso, R., Sánchez-Mateos, P., Puig-Kroger, A., López-Bravo, M., Joven, J., Ardavín, C., Rodríguez-Fernández, J.L., Sánchez-Torres, C., Mellado, M., Corbí, Á.L., 2014. CCL2 Shapes Macrophage Polarization by GM-CSF and M-CSF: Identification of CCL2/CCR2-Dependent Gene Expression Profile. J Immunol 192, 3858–3867. https://doi.org/10.4049/jimmunol.1302821
Simón-Fuentes, M., Ríos, I., Herrero, C., Lasala, F., Labiod, N., Luczkowiak, J., Roy-Vallejo, E., Córdoba-Oñate, S.F. de, Delgado-Wicke, P., Bustos, M., Fernández-Ruiz, E., Colmenares, M., Puig-Kröger, A., Delgado, R., Vega, M.A., Corbí, Á.L., Domínguez-Soto, Á., 2023. MAFB shapes human monocyte–derived macrophage response to SARS-CoV-2 and controls severe COVID-19 biomarker expression. JCI Insight 8. https://doi.org/10.1172/jci.insight.172862
Smythies, L.E., Sellers, M., Clements, R.H., Mosteller-Barnum, M., Meng, G., Benjamin, W.H., Orenstein, J.M., Smith, P.D., 2005. Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. J Clin Invest 115, 66–75. https://doi.org/10.1172/JCI200519229
Somma, D., Kok, F.O., Kerrigan, D., Wells, C.A., Carmody, R.J., 2021. Defining the Role of Nuclear Factor (NF)-κB p105 Subunit in Human Macrophage by Transcriptomic Analysis of NFKB1 Knockout THP1 Cells. Frontiers in Immunology 12.
Soneson, C., Love, M.I., Robinson, M.D., 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 4, 1521. https://doi.org/10.12688/f1000research.7563.2
Spadaro, O., Camell, C., Bosurgi, L., Nguyen, K., Youm, Y.-H., Rothlin, C.V., Dixit, V.D., 2017. IGF1 shapes the macrophage activation in response to immunometabolic challenge. Cell Rep 19, 225–234. https://doi.org/10.1016/j.celrep.2017.03.046
Stahl, P.D., Ezekowitz, R.A.B., 1998. The mannose receptor is a pattern recognition receptor involved in host defense. Current Opinion in Immunology 10, 50–55. https://doi.org/10.1016/S0952-7915(98)80031-9
Stanley, E.R., Chitu, V., 2014. CSF-1 Receptor Signaling in Myeloid Cells. Cold Spring Harb Perspect Biol 6, a021857. https://doi.org/10.1101/cshperspect.a021857
Stansfield, B.K., Ingram, D.A., 2015. Clinical significance of monocyte heterogeneity. Clin Transl Med 4, 5. https://doi.org/10.1186/s40169-014-0040-3
Stark, R., Grzelak, M., Hadfield, J., 2019. RNA sequencing: the teenage years. Nat Rev Genet 20, 631–656. https://doi.org/10.1038/s41576-019-0150-2

Stein, M., Keshav, S., Harris, N., Gordon, S., 1992. Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. J Exp Med 176, 287–292. https://doi.org/10.1084/jem.176.1.287

Storey, J.D., 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64, 479–498. https://doi.org/10.1111/1467-9868.00346

Strizova, Z., Benesova, I., Bartolini, R., Novysedlak, R., Cecrdlova, E., Foley, L.K., Striz, I., 2023. M1/M2 macrophages and their overlaps – myth or reality? Clin Sci (Lond) 137, 1067–1093. https://doi.org/10.1042/CS20220531

Swirski, F.K., Nahrendorf, M., Etzrodt, M., Wildgruber, M., Cortez-Retamozo, V., Panizzi, P., Figueiredo, J.-L., Kohler, R.H., Chudnovskiy, A., Waterman, P., Aikawa, E., Mempel, T.R., Libby, P., Weissleder, R., Pittet, M.J., 2009. Identification of Splenic Reservoir Monocytes and Their Deployment to Inflammatory Sites. Science 325, 612–616. https://doi.org/10.1126/science.1175202

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., von Mering, C., 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43, D447-452. https://doi.org/10.1093/nar/gku1003

’t Hoen, P.A.C., Friedländer, M.R., Almlöf, J., Sammeth, M., Pulyakhina, I., Anvar, S.Y., Laros, J.F.J., Buermans, H.P.J., Karlberg, O., Brännvall, M., GEUVADIS Consortium, den Dunnen, J.T., van Ommen, G.-J.B., Gut, I.G., Guigó, R., Estivill, X., Syvänen, A.-C., Dermitzakis, E.T., Lappalainen, T., 2013. Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. Nat Biotechnol 31, 1015–1022. https://doi.org/10.1038/nbt.2702

Takeuchi, O., Akira, S., 2010. Pattern Recognition Receptors and Inflammation. Cell 140, 805–820. https://doi.org/10.1016/j.cell.2010.01.022

Tang, S., Zhang, Y., Lin, X., Wang, H., Yong, L., Zhang, H., Cai, F., 2022. CLEC10A can serve as a potential therapeutic target and its level correlates with immune infiltration in breast cancer. Oncol Lett 24, 285. https://doi.org/10.3892/ol.2022.13405

The Gene Ontology Consortium, Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., Hill, D.P., Lee, R., Mi, H., Moxon, S., Mungall, C.J., Muruganugan, A., Mushayahama, T., Sternberg, P.W., Thomas, P.D., Van Auken, K., Ramsey, J., Siegele, D.A., Chisholm, R.L., Fey, P., Aspromonte, M.C., Nugnes, M.V., Quaglia, F., Tosatto, S., Giglio, M., Nadendla, S., Antonazzo, G., Attrill, H., dos Santos, G., Marygold, S., Strelets, V., Tabone, C.J., Thurmond, J., Zhou, P., Ahmed, S.H., Asanitthong, P., Luna Buitrago, D., Erdol, M.N., Gage, M.C., Ali Kadhum, M., Li, K.Y.C., Long, M., Michalak, A., Pesala, A., Pritazahra, A., Saverimuttu, S.C.C., Su, R., Thurlow, K.E., Lovering, R.C., Logie, C., Oliferenko, S., Blake, J., Christie, K., Corbani, L., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Smith, C., Cuzick, A., Seager, J., Cooper, L., Elser, J., Jaiswal, P., Gupta, P., Jaiswal, P., Naithani, S., Lera-Ramirez, M., Rutherford, K., Wood, V., De Pons, J.L., Dwinell, M.R., Hayman, G.T., Kaldunski, M.L., Kwitek, A.E., Laulederkind, S.J.F., Tutaj, M.A., Vedi, M., Wang, S.-J., D’Eustachio, P., Aimo,

L., Axelsen, K., Bridge, A., Hyka-Nouspikel, N., Morgat, A., Aleksander, S.A., Cherry, J.M., Engel, S.R., Karra, K., Miyasato, S.R., Nash, R.S., Skrzypek, M.S., Weng, S., Wong, E.D., Bakker, E., Berardini, T.Z., Reiser, L., Auchincloss, A., Axelsen, K., Argoud-Puy, G., Blatter, M.-C., Boutet, E., Breuza, L., Bridge, A., Casals-Casas, C., Coudert, E., Estreicher, A., Livia Famiglietti, M., Feuermann, M., Gos, A., Gruaz-Gumowski, N., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Le Mercier, P., Lieberherr, D., Masson, P., Morgat, A., Pedruzzi, I., Pourcel, L., Poux, S., Rivoire, C., Sundaram, S., Bateman, A., Bowler-Barnett, E., Bye-A-Jee, H., Denny, P., Ignatchenko, A., Ishtiaq, R., Lock, A., Lussi, Y., Magrane, M., Martin, M.J., Orchard, S., Raposo, P., Speretta, E., Tyagi, N., Warner, K., Zaru, R., Diehl, A.D., Lee, R., Chan, J., Diamantakis, S., Raciti, D., Zarowiecki, M., Fisher, M., James-Zorn, C., Ponferrada, V., Zorn, A., Ramachandran, S., Ruzicka, L., Westerfield, M., 2023. The Gene Ontology knowledgebase in 2023. Genetics 224, iyad031. https://doi.org/10.1093/genetics/iyad031

Toobian, D., Ghosh, P., Katkar, G.D., 2021. Parsing the Role of PPARs in Macrophage Processes. Front. Immunol. 12. https://doi.org/10.3389/fimmu.2021.783780

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31, 46–53. https://doi.org/10.1038/nbt.2450

Tripathi, S., Christie, K.R., Balakrishnan, R., Huntley, R., Hill, D.P., Thommesen, L., Blake, J.A., Kuiper, M., Lægreid, A., 2013. Gene Ontology annotation of sequence-specific DNA binding transcription factors: setting the stage for a large-scale curation effort. Database (Oxford) 2013, bat062. https://doi.org/10.1093/database/bat062

Tsou, C.-L., Peters, W., Si, Y., Slaymaker, S., Aslanian, A.M., Weisberg, S.P., Mack, M., Charo, I.F., 2007. Critical roles for CCR2 and MCP-3 in monocyte mobilization from bone marrow and recruitment to inflammatory sites. J Clin Invest 117, 902–909. https://doi.org/10.1172/JCI29919

Tugal, D., Liao, X., Jain, M.K., 2013. Transcriptional Control of Macrophage Polarization. Arteriosclerosis, Thrombosis, and Vascular Biology 33, 1135–1144. https://doi.org/10.1161/ATVBAHA.113.301453

Unuvar Purcu, D., Korkmaz, A., Gunalp, S., Helvaci, D.G., Erdal, Y., Dogan, Y., Suner, A., Wingender, G., Sag, D., 2022. Effect of stimulation time on the expression of human macrophage polarization markers. PLoS One 17, e0265196. https://doi.org/10.1371/journal.pone.0265196

Uribe-Querol, E., Rosales, C., 2020. Phagocytosis: Our Current Understanding of a Universal Biological Process. Front. Immunol. 11. https://doi.org/10.3389/fimmu.2020.01066

Ushach, I., Zlotnik, A., 2016. Biological role of granulocyte macrophage colony-stimulating factor (GM-CSF) and macrophage colony-stimulating factor (M-CSF) on cells of the myeloid lineage. Journal of Leukocyte Biology 100, 481–489. https://doi.org/10.1189/jlb.3RU0316-144R

Vannella, K.M., Barron, L., Borthwick, L.A., Kindrachuk, K.N., Narasimhan, P.B., Hart, K.M., Thompson, R.W., White, S., Cheever, A.W., Ramalingam, T.R., Wynn, T.A., 2014. Incomplete Deletion of IL-4Rα by LysMCre Reveals Distinct Subsets of M2 Macrophages Controlling Inflammation and Fibrosis in Chronic Schistosomiasis. PLOS Pathogens 10, e1004372.

https://doi.org/10.1371/journal.ppat.1004372
Villani, A.-C., Sarkizova, S., Hacohen, N., 2018. Systems Immunology: Learning the Rules of the Immune System. Annu. Rev. Immunol. 36, 813–842. https://doi.org/10.1146/annurev-immunol-042617-053035
von Mering, C., Jensen, L.J., Snel, B., Hooper, S.D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A., Bork, P., 2005. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res 33, D433-437. https://doi.org/10.1093/nar/gki005
Wahlster, L., Daley, G.Q., 2016. Progress towards generation of human haematopoietic stem cells. Nat. Cell Biol. 18, 1111–1117. https://doi.org/10.1038/ncb3419
Wang, J.Z., Du, Z., Payattakool, R., Yu, P.S., Chen, C.-F., 2007. A new method to measure the semantic similarity of GO terms. Bioinformatics 23, 1274–1281. https://doi.org/10.1093/bioinformatics/btm087
Wang, N., Liang, H., Zen, K., 2014. Molecular Mechanisms That Influence the Macrophage M1–M2 Polarization Balance. Front Immunol 5, 614. https://doi.org/10.3389/fimmu.2014.00614
Wang, Q., Sudan, K., Schmoeckel, E., Kost, B.P., Kuhn, C., Vattai, A., Vilsmaier, T., Mahner, S., Jeschke, U., Heidegger, H.H., 2022. CCL22-Polarized TAMs to M2a Macrophages in Cervical Cancer In Vitro Model. Cells 11, 2027. https://doi.org/10.3390/cells11132027
Wang, X., Wu, Z., Zhang, X., 2010. Isoform abundance inference provides a more accurate estimation of gene expression levels in RNA-seq. J Bioinform Comput Biol 8 Suppl 1, 177–192. https://doi.org/10.1142/s0219720010005178
Weinberg, R.A., 1995. The retinoblastoma protein and cell cycle control. Cell 81, 323–330. https://doi.org/10.1016/0092-8674(95)90385-2
Williams, C.R., Baccarella, A., Parrish, J.Z., Kim, C.C., 2016. Trimming of sequence reads alters RNA-Seq gene expression estimates. BMC Bioinformatics 17, 103. https://doi.org/10.1186/s12859-016-0956-2
Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G., 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation 2. https://doi.org/10.1016/j.xinn.2021.100141
Xu, H., Li, H., Woo, S.-L., Kim, S.-M., Shende, V.R., Neuendorff, N., Guo, X., Guo, T., Qi, T., Pei, Y., Zhao, Y., Hu, X., Zhao, J., Chen, Lili, Chen, Lulu, Ji, J.-Y., Alaniz, R.C., Earnest, D.J., Wu, C., 2014. Myeloid Cell-specific Disruption of Period1 and Period2 Exacerbates Diet-induced Inflammation and Insulin Resistance *. Journal of Biological Chemistry 289, 16374–16388. https://doi.org/10.1074/jbc.M113.539601
Xuan, W., Qu, Q., Zheng, B., Xiong, S., Fan, G.-H., 2015. The chemotaxis of M1 and M2 macrophages is regulated by different chemokines. Journal of Leukocyte Biology 97, 61–69. https://doi.org/10.1189/jlb.1A0314-170R
Yi, H., Yu, X., Gao, P., Wang, Y., Baek, S.-H., Chen, X., Kim, H.L., Subjeck, J.R., Wang, X.-Y., 2009. Pattern recognition scavenger receptor SRA/CD204 down-regulates Toll-like receptor 4 signaling-dependent CD8 T-cell activation. Blood 113, 5819–5828. https://doi.org/10.1182/blood-2008-11-190033
Yona, S., Kim, K.-W., Wolf, Y., Mildner, A., Varol, D., Breker, M., Strauss-Ayali, D., Viukov, S., Guilliams, M., Misharin, A., Hume, D.A., Perlman, H., Malissen, B., Zelzer, E., Jung, S., 2013. Fate mapping reveals origins and dynamics of

monocytes and tissue macrophages under homeostasis. Immunity 38, 79–91. https://doi.org/10.1016/j.immuni.2012.12.001
York, M.R., Nagai, T., Mangini, A.J., Lemaire, R., van Seventer, J.M., Lafyatis, R., 2007. A macrophage marker, siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by type i interferons and toll-like receptor agonists. Arthritis & Rheumatism 56, 1010–1020. https://doi.org/10.1002/art.22382
Yu, G., 2020. Gene Ontology Semantic Similarity Analysis Using GOSemSim, in: Kidder, B.L. (Ed.), Stem Cell Transcriptional Networks: Methods and Protocols, Methods in Molecular Biology. Springer US, New York, NY, pp. 207–215. https://doi.org/10.1007/978-1-0716-0301-7_11
Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS 16, 284–287. https://doi.org/10.1089/omi.2011.0118
Yunna, C., Mengru, H., Lei, W., Weidong, C., 2020. Macrophage M1/M2 polarization. European Journal of Pharmacology 877, 173090. https://doi.org/10.1016/j.ejphar.2020.173090
Zeng, H., 2022. What is a cell type and how to define it? Cell 185, 2739–2755. https://doi.org/10.1016/j.cell.2022.06.031
Zhang, M., Huang, B., 2012. The multi-differentiation potential of peripheral blood mononuclear cells. Stem Cell Res Ther 3, 48. https://doi.org/10.1186/scrt139
Zhang, Y., Du, W., Chen, Z., Xiang, C., 2017. Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma. Experimental Cell Research 359, 449–457. https://doi.org/10.1016/j.yexcr.2017.08.028
Zhu, A., Ibrahim, J.G., Love, M.I., 2019. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics 35, 2084–2092. https://doi.org/10.1093/bioinformatics/bty895
Ziegler-Heitbrock, L., Ancuta, P., Crowe, S., Dalod, M., Grau, V., Hart, D.N., Leenen, P.J.M., Liu, Y.-J., MacPherson, G., Randolph, G.J., Scherberich, J., Schmitz, J., Shortman, K., Sozzani, S., Strobl, H., Zembala, M., Austyn, J.M., Lutz, M.B., 2010. Nomenclature of monocytes and dendritic cells in blood. Blood 116, e74–e80. https://doi.org/10.1182/blood-2010-02-258558