Spatiotemporal single cell analysis of human colonic macrophages

Victoria Therese Karlsen

Master in Bioscience
Cell biology, Physiology and Neuroscience
60 credits

Department of Biosciences
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

October 2021

Spatiotemporal single cell analysis of human colonic macrophages

Victoria Therese Karlsen
Department of Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo

October 2021

© Victoria Therese Karlsen

2021

Spatiotemporal single cell analysis of human colonic macrophages

Victoria Therese Karlsen

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo

Acknowledgements

This project was performed in the Jahnsen group at the Laboratory of Immunohistochemistry and Immunopathology (LIIPAT), Institute of Pathology, Oslo University Hospital Rikshospitalet, University of Oslo. The work was carried out from August 2019 to October 2021 under the supervision of Diana Domanska and Espen Bækkevold, with Finn-Eirik Johansen as internal supervisor.

Foremost, I would like to thank my two supervisors, Diana Domanska and Espen Bækkevold, for their great support through my thesis work. Thank you Diana for guiding me through the computational analyses. Espen, thank you for all your help in the lab, for always being available to answer my questions and for all your feedback during the process of writing the thesis. I would also like to thank Umair Majid for guiding me through the process of tissue dissection, Hogne Røed Nilsen for teaching me immunohistochemistry, and Kjersti Thorvaldsen Hagen for always making sure I have everything I need in the lab and teaching me immunofluorescent staining. I would also like to thank Frode Jahnsen for giving me the opportunity to be a part of the Jahnsen group and Finn-Eirik Johansen for being my internal supervisor.

Abstract

The adult human gut spans an area of about 200–400 m^2, and this huge surface is covered by a single layer of epithelial cells creating a physical barrier against the environment. To maintain the integrity of this critical interface, it is crucial for the immune system to provide protection against intestinal pathogens, while at the same time avoiding excessive immune responses. Failure of intestinal immune regulation may lead to severe and debilitating chronic inflammatory disorders, like inflammatory bowel disease (IBD). Maintenance of homeostasis in the intestinal tract requires a critical regulation of host immune responses to the luminal contents, and intestinal macrophages are essential in this process by their production of immunoregulatory factors and potential to propagate regulatory T cells (Tregs) in the gut.

However, most of our knowledge to this end is derived from experimental animal models and in vitro cell cultures, and studies of human tissues are highly warranted. Moreover, most experimental and clinical results from tissues are based on bulk-analysis of cells or tissue samples. Such “averaging” of data inadvertently misses critical information because the heterogeneity of the samples is ignored, while the nature of biology is highly diverse.

To fully capture the heterogeneity and functional capacities of intestinal macrophages, we performed an unbiased single-cell transcriptomic analysis of macrophages isolated from human colon. This study shows that the colon mucosa contains monocyte-like cells that appear to differentiate into multiple transcriptionally distinct subsets, including subsets with proinflammatory properties and subsets with high antigen presenting and phagocytic capacity. Moreover, the various macrophage subtypes appeared to occupy distinct niches in the tissues. Detailed understanding of the intestinal macrophage landscape in healthy tissues may guide development of new anti-inflammatory treatment approaches.

Table of Contents

Abbreviations XI
Thesis aim 1
1 Introduction 3
  1.1 The immune system 3
  1.2 The immune response 3
  1.3 T lymphocytes 4
  1.4 Antigen-presenting cells 5
  1.5 Dendritic cells 6
  1.6 Macrophages 6
    1.6.1 Macrophage subgroups 8
    1.6.2 Intestinal macrophages 9
    1.6.3 Toll-like receptors 11
    1.6.4 Antigen sampling in the intestines 12
  1.7 The colon 13
    1.7.1 Colonic microbiota 14
    1.7.2 Macrophages in intestinal diseases 15
  1.8 RNA 16
  1.9 Bulk RNA sequencing 16
  1.10 Single-cell RNA sequencing 16
    1.10.1 Workflow 17
    1.10.2 10x genomics workflow 18
    1.10.3 Technical and biological variation 19
    1.10.4 Analysis 19
    1.10.5 Single-cell RNA sequencing compared to bulk RNA sequencing 21
  1.11 Bioinformatic tools for scRNA-seq analysis 21
    1.11.1 Seurat 22
    1.11.2 Monocle 3 23
2 Materials and methods 25
  2.1 Patient information 25
    2.1.1 Ethical considerations 25
  2.2 Tissue preparation 25

2.3 Isolation of mononuclear cells from peripheral blood ... 27
2.4 Cell staining ... 28
2.5 Flow cytometry ... 33
2.5.1 Analysis of flow cytometry data ... 33
2.6 Cell sorting ... 34
2.7 10x genomics ... 35
2.8 Computational analysis of scRNA-seq data ... 36
2.8.1 Analysis with Cell Ranger ... 36
2.8.2 Analysis with Seurat ... 36
2.8.3 Analysis with Monocle 3 ... 38
2.9 Immunohistochemistry ... 39
2.9.1 Immunofluorescent staining of formalin fixated slides ... 39
2.9.2 Enzymatic staining of formalin fixated slides ... 41
2.9.3 Immunofluorescent staining of cryopreserved sections ... 43
3 Results ... 45
3.1 Colonic macrophages are confined to the HLA-DR$^+$CD14$^+$ cell population ... 45
3.2 Macrophage sorting for single-cell RNA sequencing ... 47
3.3 Colonic macrophages separate into distinct clusters based on differently expressed genes ... 48
3.3.1 Cluster annotation ... 51
3.4 Trajectories ... 53
3.5 Identification of mature and immature macrophages in colonic tissue ... 56
3.5.1 DC3s – A newly described macrophage-related DC subset ... 57
3.6 Localization of colonic macrophages *in situ* ... 58
4 Discussion ... 65
4.1.1 Methodological consideration ... 68
5 Conclusion and further perspectives ... 71
Appendix 1 ... 73
Appendix 2 ... 75
Appendix 3 ... 77
Appendix 4 ... 81
References ... 83

Abbreviations

APC	Antigen presenting cell
BSA	Bovine Serum Albumin
CCL	Chemokine ligand
CD	Cluster of differentiation
cDNA	Complementary DNA
CST	Cytometer setup and tracking
CXCL	Chemokine (C-X-C motif) ligand
DC	Dendritic cell
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence-Activated Cell Sorting
Fc	Fragment, crystalline
FCS	Fetal calf serum
FMO	Fluorescence minus one
g	Gravitational force
HIER	Heat induced epitope retrieval
HLA	Human leucocyte antigen
IBD	Inflammatory bowel disease
Ig	Immunoglobulin
IHC	Immunohistochemistry
IL	Interleukin
IFN	Interferon
Lin	Lineage
LPS	Lipopolysaccharide
M	Molar
MDP	Macrophage and Dendritic cell precursor
mg	Milligram
MHC	Major histocompatibility complex
ml	Milliliter
mM	Millimolar
Na	Sodium
nt	Nucleotide
PAMPs	Pathogen-associated molecular patterns
PBMCs	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PRR	Pattern recognition receptor
PVA	Polyvinyl alcohol
rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute medium
TBS	Tris-buffered saline
TCR	T cell receptor
Th	T helper cell
TLR	Toll-like receptor
TNF	Tumor necrosis factor
μl	Microliter
μm	Micrometer

Thesis aim

Single cell RNA sequencing enables unbiased untangling of cell heterogeneity, discovery of novel cell types, and uncovers the critical roles of cellular heterogeneity in biological processes. By looking at the gene expression in individual cells it is possible to reveal differences between tissues as well as differences between layers and compartments within the tissue. In clinical settings, single cell sequencing allows comparison of healthy and non-healthy cells and tissues. Here, new therapeutic targets may be uncovered, which might be used to develop novel immunotherapies and other disease treatments. Selection of specific cell types for further analysis based on gene expression may also be done. It is then possible to look into expression of specific markers and investigate the gene expression landscape. This might lead to the identification of “new” cell types or subtypes.

Macrophages are critically important in the immune regulation within the mucosal layer of the colon, and it is therefore essential to understand the real complexity of such cells in tissues where they reside. The aim of the thesis is to map the landscape of macrophages in human colon. In this project, single cells were isolated from healthy colon for single-cell RNA sequencing (scRNAseq). The scRNAseq data was then analyzed in order to investigate the gene expression landscape of macrophages in the mucosal layer of the colon. Isolated cells were also stained for flow cytometry to verify the presence of specific markers identified in the scRNAseq data analysis. Finally, immunostaining of colon tissue was performed to assess the *in situ* expression of markers by macrophage subsets.

Investigation of the macrophage landscape generates an important foundation from which new knowledge might emerge. To get a better understanding of diseases affecting the colon, like inflammatory bowel disease (IBD), it is vital to have a solid understanding of the colon in steady state. Since resident tissue macrophages play such an important role in the tissue, it is crucial to understand their distribution throughout the tissue and how they have adapted to their local environment.

1 Introduction

1.1 The immune system

The immune system developed to protect the host from foreign bodies, such as microbes. It is composed of several different cell types (e.g. macrophages, dendritic cells, B lymphocytes and T lymphocytes), blood proteins (e.g. antibodies and complement system components), as well as physical barriers such as epithelial surfaces [1].

The immune system is commonly divided into two main parts, namely the innate and the adaptive immune system. The innate immune system consists of components already present before the infection. Here, we find phagocytic cells (e.g. macrophages and dendritic cells) in addition to the physical barriers and components of the complement system. Innate immune responses are the primary response to infection and is crucial for host defense in the first hours or days. Following the innate response is the adaptive immune response. As the name suggests, this response is adaptive, meaning it is acquired by interaction with different microbes over the course of life. Consequently, the adaptive immune response varies within and between human populations. Here, we find both the B and T lymphocytes expressing specific receptors that enables binding to specific antigens (i.e. substances being nonmicrobial or microbial). Adaptive immune responses are therefore more specific than the innate response. The adaptive immune system can further be divided into two parts – cell-mediated and humoral immunity. T lymphocytes are responsible for the cell-mediated immunity, where microbes are eliminated by phagocytes. Humoral immunity depends on antibodies produced by B lymphocytes to neutralize and clear extracellular pathogens. After being introduced to a new microbe, the adaptive immune system is able to create an immunological memory where long-lived memory cells are generated after the first encounter with the microbe [1].

1.2 The immune response

The innate and adaptive immune system work together to generate an effective immune response. When a foreign body has entered the host, dendritic cells, mast cells and macrophages recognize pathogen-associated molecular patterns (PAMPs) on the pathogen through pattern recognition receptors (PRRs) and become activated. This activation initiates an inflammation with an influx of phagocytes into the tissue. For the phagocytes to be able to enter the inflamed tissue, the endothelium must be activated. TNF and IL-1 (tumor necrosis factor and interleukin-1, respectively) produced by activated macrophages facilitates the activation of endothelial cells, stimulating their production of selectin. Leukocytes are then able to engage in selectin-ligand interactions with the vessel surface. This interaction causes the leukocytes to “slow down” in the blood stream and roll on the vessel wall. As the leukocytes roll along the endothelial surface, chemokines produced by the endothelial cells are able to activate integrins on the leukocytes, enabling their firm adhesion on the vessel surface and finally extravasation of leukocytes through junctions between the endothelial cells [1].

At the site of inflammation, dendritic cells take up antigen and become activated by PAMPs. The cells migrate to the draining lymph node to present the antigen and activate naïve lymphocytes (i.e. B cells and T cells) which will become antibody-producing cells and effector T cells, respectively. The effector T cells migrate back to the inflamed tissue, where they facilitate the killing of infected cells and recruit additional phagocytes, such as macrophages. The antibodies produced by activated B cells may leave the blood stream to fight of the foreign body in collaboration with components of the innate immune system [1].

1.3 T lymphocytes

T lymphocytes (or T cells) are part of the adaptive immune system, where their introduction to foreign bodies elicit a cell-mediated immune response, resulting in destruction of infected cells. T cells develop from bone marrow precursors that migrate to and mature in the thymus. During their maturation process T cells undergo positive and negative selection. The first step of T cell development is production of T cells positive for both CD4 and CD8, by rearrangement of their T cell receptor (TCR). When expressing both CD4 and CD8, the double-positive thymocytes undergo positive selection where all cells able to bind MHC molecules displaying self-peptides with low affinity are selected for. From here, double-positive thymocytes recognizing MHC class I become CD4^{-}CD8^{+} T cells and the ones recognizing MHC class II develop into CD4^{+}CD8^{-} T cells. The double-positive thymocytes unable to bind the self-peptide MHC complex undergo apoptosis. During the T cell development there is also a negative-selection step, where T cells binding to self-peptide MHC complexes with high affinity are terminated through apoptosis, this is an important step to avoid autoimmunity. The resulting cells, termed naïve CD4 or CD8 T cells, leave the thymus and circulate through the secondary lymphoid organs [2].

The primary site for activation of naïve T cells is in the secondary lymphoid organs (e.g. lymph nodes) were they encounter the specialized migratory antigen-presenting cells (APCs, see below) termed dendritic cells. When naïve T cells bind to APCs presenting an antigen, proliferation and cytokine production is initiated leading to differentiation into mature, activated T cells with specific functions – either effector cells that eliminate microbes or long-lived memory cells. The effector T cells are able to interact with and activate macrophages, resulting in microbial killing [1].

In response to various cytokines produced by dendritic cells, naïve CD4$^+$ T cells may differentiate into different subset of effector T cells – Th1, Th2 or Th17. Differentiation into the Th1 subset, involved in infections by activation of macrophages, is mediated by the cytokines IFN-γ and IL-12. Key transcription factors for this subset are STAT4, T-bet and STAT1, with IFN-γ as the main cytokine produced. The cytokine involved in Th2 differentiation is IL-4, which leads to expression of GATA-3 by activation of the STAT6 transcription factor. This subset, producing the cytokines IL-4, IL-5 and IL-13, mediates defense against helminthic infections. Th17 cells derive from CD4$^+$ T cells stimulated by proinflammatory cytokines like IL-6 together with TGF-β, and depends on STAT3 and RORγt, which both are transcription factors. The principal cytokines produced by this subset are IL-17 and IL-22. These cells are able to recruit leukocytes to the infection site and thereby participate in microbial defense [1,3].

1.4 Antigen-presenting cells

Antigen-presenting cells (APCs) are cells able to activate T cells through antigen presentation on MHC molecules. There are three types of professional APCs, namely B lymphocytes, dendritic cells (DCs) and macrophages. B lymphocytes present antigens to T helper cells in humoral immunity, while macrophages present to effector T cells in cell-mediated immunity. DCs, on the other hand, initiate priming of T cells in secondary lymphoid tissues. When the APCs come into contact with antigens in the tissue, they present antigens on a major histocompatibility complex (MHC) molecule, also named human leucocyte antigen (HLA) [4]. There are two classes of MHC molecules. MHC class I is present on all nucleated cells, while MHC class II is only present on selected cells such as the professional APCs. The MHC molecules always presents peptides on the cells surface. Usually, self-peptides are presented which the immune cells, under normal conditions, do not react to. But when foreign microbes are present, cells may present microbe peptides on the MHC molecule which the immune cells will react to and initiate an immune response [1].

1.5 Dendritic cells

Dendritic cells (DCs) are the most efficient APCs that constantly migrate from blood to peripheral tissues where they sample antigens by cellular uptake. Subsequently, they migrate to regional lymph nodes through the draining lymph. Within lymph nodes, DCs initiate primary T cell responses by activating both $CD4^+$ and $CD8^+$ T cells through antigen presentation on MHC class II and MHC class I, respectively. DCs are traditionally divided into conventional and plasmacytoid dendritic cells, both of which derive from hematopoietic stem cells (HSCs) in the bone marrow [5,6].

The conventional DC (cDC) subtype derive from pre-DCs in the bone marrow migrating to the tissue, where they fully mature into cDCs [7]. Conventional DCs may be further divided into cDC1s and cDC2s. The cDC1 subgroup, identified as $CD14^-CD1c^-CD141^+$, is highly important in response to intracellular pathogens where they secrete IL-12 through toll-like receptor 3 (TLR3) stimulation, and facilitate differentiation of T helper cells to the Th1 subtype. The $CD14^-CD1c^+CD141^-$ cDC2s, on the other hand, initiate differentiation towards Th2 and Th17, making them important in defense against extracellular pathogens [7–9].

Plasmacytoid DCs (pDCs) develop in the bone marrow, and are located in the secondary lymphoid tissues (e.g. lymph nodes) as well as in the thymus. During a viral infection, type I interferon (IFN) is produced by pDCs enabling activation of both the innate and adaptive immune system. [8,10].

1.6 Macrophages

Macrophages are mainly tissue-resident professional APCs with different functions such as cytokine production (i.e. proteins facilitating immune responses), phagocytosis (i.e. engulfment of particles) and antigen presentation. They play an important role in both the innate and adaptive immune system and are found in all tissues of the body. When antigens are detected in the host, macrophages phagocytose the foreign body and present the antigen peptide on MHC molecules. By presenting this antigen peptide, the macrophage can activate T cells, which are part of the adaptive immune system. This activation creates an important link between the innate and adaptive immune system, which is crucial in host defense against foreign bodies. Maintenance of the tissue by clearance of apoptotic cells through phagocytosis is also an important function of macrophages [1,11,12].

Classically, tissue macrophages were thought to derive from blood monocytes that constantly seed tissues and mature to macrophages locally [13]. However, recent data indicate that macrophages may also derive from fetal precursors in the yolk sac or fetal liver that maintain themselves through local proliferation. In fact, mouse studies indicate that subsets of self-propagating fetal-derived intestinal macrophages are involved in maintenance of intestinal homeostasis [14]. However, the exact contribution of these different precursor sources to the macrophage network in adult human tissues is not clear, and the functional differences between fetal- and monocyte-derived macrophages is currently elusive.

In the case of bone marrow derived macrophages, a macrophage and dendritic cell precursor (MDP) cell in the bone marrow generates a monocyte able to leave and enter blood stream in a CCR2-dependent manner. Through extravasation, circulating monocytes may enter tissues, where local stimulation such as growth factors drive the differentiation into macrophages. In the intestines there is a constant recruitment of circulating monocytes entering the tissue and differentiating into tissue-resident intestinal macrophages [15–17].

In mice, embryonically derived macrophages from fetal yolk sac give rise to a large number of the tissue-resident macrophages in adults. These macrophages are likely there prior to birth and have the ability to self-renew, meaning that they are not dependent on peripheral blood monocytes. Seeding of embryonically derived macrophages occurs in two waves. The first wave generates macrophages migrating to the brain, giving rise to microglia (i.e. macrophages in the brain). This wave is independent of monocytes and appears in the early embryonic yolk sac. Migration of macrophages to the fetal liver occurs in the second wave. When the macrophages enter the fetal liver, they function as precursors for monocytes. Prior to birth, these monocytes can enter other tissues and differentiate into tissue-resident macrophages. It is debatable if these yolk sac derived macrophages are also present in adult humans, as most of the studies are performed in laboratory mice with limited exposure to pathogens. However, there is some evidence suggesting the presence of human macrophages with yolk sac origin [1,11,18]. In the case of intestinal macrophages, it is recognized that the macrophages are mainly monocyte-derived. This has been shown both in mice [19] and humans [20].

Macrophages express several different surface receptors important for their effector functions. In humans, macrophages are usually defined by the expression of CD14, CD68, CD163 and CD209, as well as MHC class II. CD14 is a receptor for lipopolysaccharide (LPS), playing an important role in immune recognition. Together with TLR4 (described below), CD14 initiate secretion of pro-inflammatory cytokines, such as IL-1 and IL-6, and pro-inflammatory chemokines. The lysosomal transmembrane glycoprotein CD68 is a frequently used marker for macrophage identification, even though its function in inflammation is not fully understood. Studies have shown that CD68 is greatly upregulated during inflammation, stimulated by neutrophils in inflammatory bowel disease and TLR4. CD163 is a transmembrane protein involved in uptake of hemoglobin, as well as binding to virus and some bacteria. Macrophages positive for CD163 have been located in sites of inflammation. IL-6 and IL-10 (among others) stimulate CD163 expression, while interferon-γ and LPS reduce the expression of CD163. Regulation of the phagocytic capacity of macrophages have been shown to be mediated by CD209 [21–24].

Tissue-resident macrophages (TRMs) are closely coupled with the tissue and reside in a spesific tissue over time. They experience a dynamic mosaic in the tissue, where the local environment plays an important role in TRM function and phenotype. Local stimuli mediate macrophage polarization, in which the macrophages generate tissue-specific traits. Here, the local environment gives maturation signals to generate functional TRMs, in addition to working on a epigenetic level. Tissue environment appears to shape the functions of the local macrophage network by shaping their chromatin landscape through epigenetic changes, making the TRMs heterogeneous [11,25].

1.6.1 Macrophage subgroups

Commonly, macrophages are divided into two subgroups, namely the classical activated macrophages and the alternative activated macrophages. The classically activated macrophages, also referred to as M1, are activated by TNF-α and IFN-γ (tumor necrosis factor-α and interferon gamma, respectively) produced by a subset of T helper cells, called Th1. The M1 macrophages are characterized as inflammatory macrophages, having increased antigen-presentation and antimicrobial features. The alternatively activated macrophages (M2), on the other hand, are anti-inflammatory and activated by IL-4 and IL-13 (interleukin 4 and interleukin 13, respectively) produced by a different subset of T helper cells, namely Th2s. M1 macrophages are associated with intracellular infection, whereas M2 macrophages are involved in phagocytosis, tissue maintenance and dampening inflammation by the release of anti-inflammatory subsets [1,11,26,27].

As shown in this thesis, and by many others, this division is too simplistic and outdated. Several studies have displayed the vast heterogeneity within a macrophage population, identifying numerous different subgroups with diverse functions and locations in the tissue. This has for instance been shown by Kang *et al.*, who found seven distinct macrophage populations in their study of the mouse colon [28], Bujko *et al.* found four subgroups of macrophages in the human small intestine [20], and Cochain *et al.* identified several subpopulation with distinct function within murine atherosclerosis [29]. The M1/M2 subset may be useful as an initial broad division of macrophages, but it is far from absolute. If this duality were to be used definitively, all other macrophage subsets would have to be placed in relations to the M1s or M2s. As Nahrendorf *et al.* mentions, this need for placement of all macrophages on a scale ranging from pro-inflammatory M1 macrophages to the anti-inflammatory M2s is problematic [30]. As first pointed out by Xue *et al.*, macrophages react with specific transcriptional programs upon distinct signals resulting in a very complex spectrum of macrophage activation states [31].

1.6.2 Intestinal macrophages

Through displaying samples of the intestinal microbiota to dendritic cells and regulatory T cell expansion, the intestinal macrophages are able to generate tolerance against the numerous bacteria within the microbiota (Figure 1). The macrophages are influenced by this microbiota, as its components induce differentiation in gene expression. An example of such a component is short chain fatty acids (SCFAs), generated by fermentation of indigestible fibers by microbiotic bacteria and may induce IL-10 production by the intestinal macrophages. It is partly due to this IL-10 production as well as secretion of PGE2 , working in an anti-inflammatory manner by reducing TNF (a pro-inflammatory cytokine) production in addition to increasing IL-10 production, that the intestinal macrophages are able to maintain their unresponsiveness. Consequently, they are able to maintain homeostasis within the intestine by removal of apoptotic cells as well as bacteria and other pathogens without generating an immune response. These macrophages are therefore often referred to as “silent killers”. The importance of this is dramatically demonstrated in a rear but severe and life-threatening case of IBD caused by gene defects in the IL-10 receptor, showing that intestinal macrophages are dependent on autocrine

IL-10 to maintain their tolerogenic and homeostatic properties [32]. Intestinal macrophages also contribute to renewal of epithelial cells in the mucosa by secretion of mediators such as epidermal growth factor (EGF) and hepatocyte growth factor (HGF). In addition, they participate in mucosal defense against pathogens, as well as nutrient uptake, fluid balance and peristalsis. They might also contribute to remodeling of the tissue by metalloproteinase production, which is a zinc-dependent endopeptidase (enzyme hydrolyzing peptide bonds) able to alter the extracellular matrix. Intestinal macrophages are found in every layer in the colon and small intestine. The lamina propria has the biggest population, with a larger population in the colon compared to the small intestine [17,26,33–37].

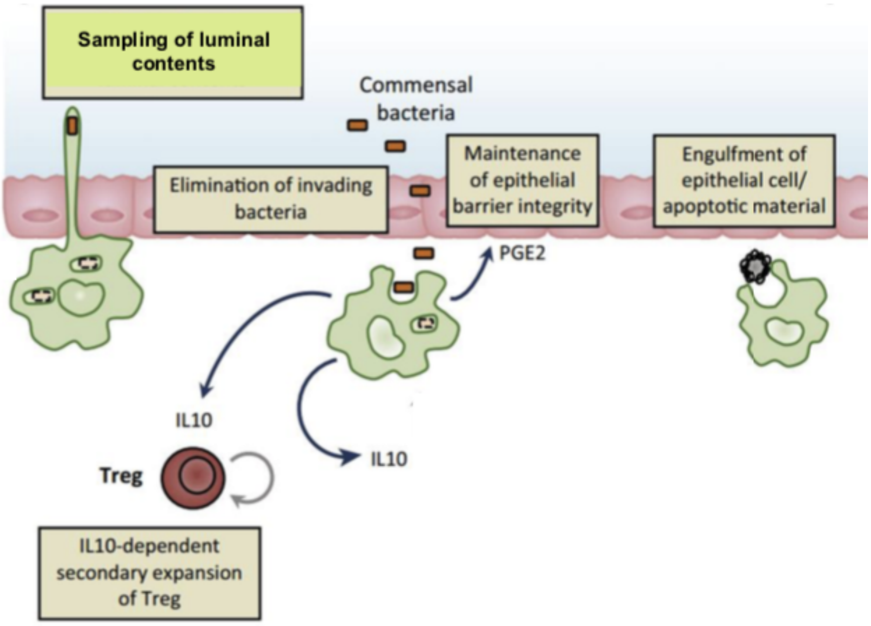
Sampling of luminal contents
Commensal bacteria
Elimination of invading bacteria
Maintenance of epithelial barrier integrity
Engulfment of epithelial cell/ apoptotic material
PGE2
IL10
IL10
Treg
IL10-dependent secondary expansion of Treg

Figure 1. Functions of intestinal macrophages.
Here, different functions of intestinal macrophages are depicted. They may sample the luminal content through protrusions, eliminate invading bacteria and clear apoptotic cells. Through the production of IL-10, intestinal macrophages may drive the expansion of regulatory T cells (Tregs). Secretion of PGE2 is important in the maintenance of the epithelial barrier. Illustration modified from [38].

1.6.3 Toll-like receptors

Toll-like receptors (TLRs) are pattern recognition receptors able to recognize molecules derived from stressed and dying cells (DAMPs) as well as from pathogens (PAMPs) in order to generate an immune response. In humans, there are 10 different TLRs identified that may be found as membrane-bound receptors or within endosomal compartments (Figure 2A), expressed on different cell types such as macrophages and dendritic cells. Each TLR recognize different PAMPs (Figure 2B). Their structure is composed of a Toll/interleukin-1 (IL-1) receptor (TIR) domain in the carboxyl-terminal and a leucine-rich amino-terminal able to bind PAMPs and DAMPs. The TIR domain functions as an adaptor on the cytosolic side of the TLR, engaging other signaling adaptor creating a signaling cascade resulting in initiation of an immune response.

TLR2 work together with TLR1 and TLR6 by the formation of a heterodimer (TLR2/TLR1 or TLR2/TLR6), resulting in recognition of a broad range of ligands. While TLR2 is mainly expressed on endothelial cells and APCs, TLR1 and TLR6 are found on a broad range of cells. TLR3 recognizes double-stranded RNA, possibly playing an important role in antiviral immunity. TLR4 was the first TLR to be characterized in humans and function by binding LPS. Both TLR2 and TLR4 interact with CD14, found on macrophages, working as co-receptors in recognition of Gram positive (TLR2) and Gram negative bacteria (TLR4). Flagellin in bacterial flagella is recognized by TLR5, while TLR9 recognizes CpG DNA from bacteria. CpG-motifs are a distinct sequence of unmethylated DNA [39–43].

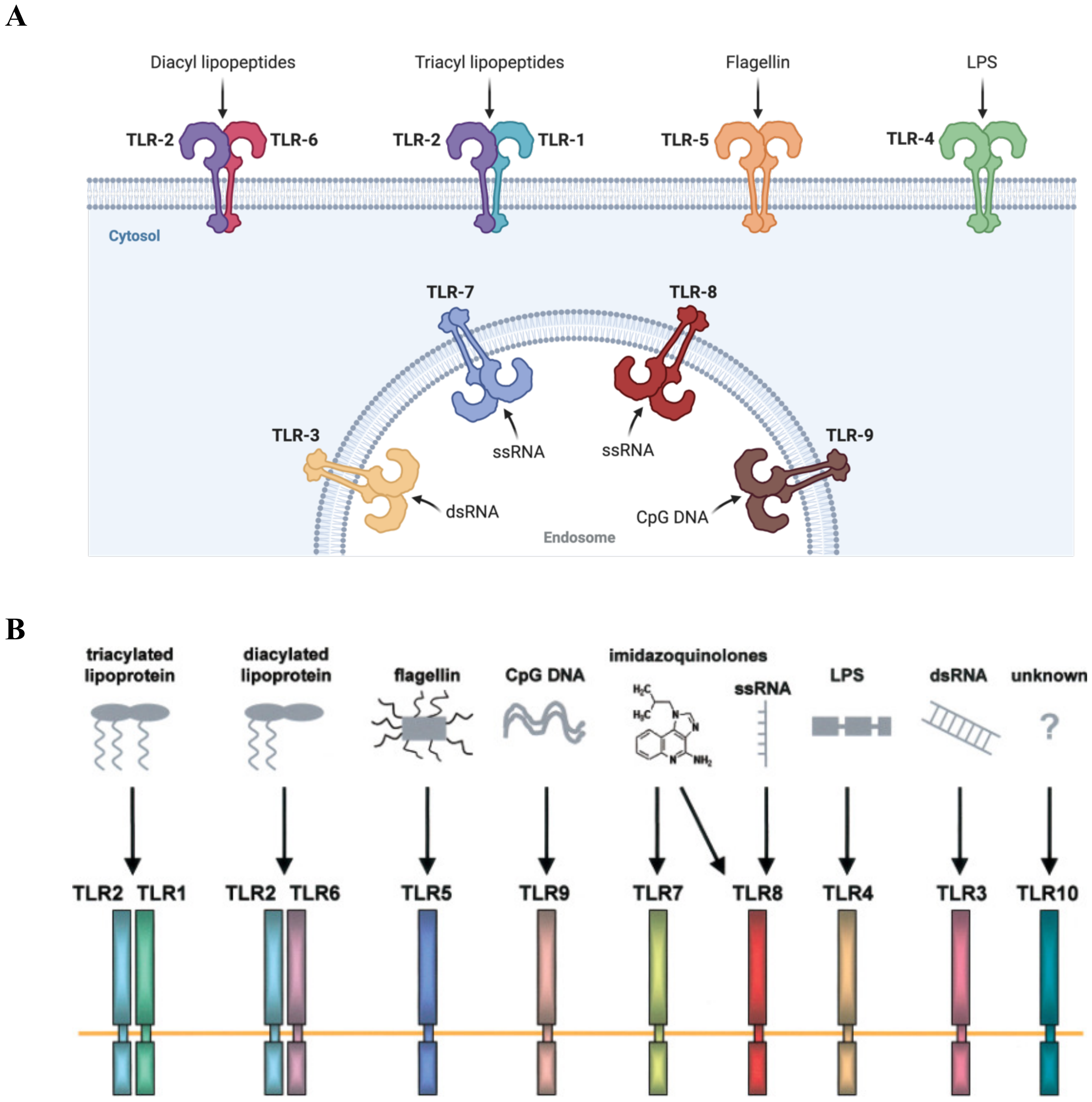
A
Diacyl lipopeptides
Triacyl lipopeptides
Flagellin
LPS
TLR-2
TLR-6
TLR-2
TLR-1
TLR-5
TLR-4
Cytosol
TLR-7
TLR-8
TLR-3
ssRNA
ssRNA
TLR-9
dsRNA
CpG DNA
Endosome
B
triacylated lipoprotein
diacylated lipoprotein
flagellin
CpG DNA
imidazoquinolones
ssRNA
LPS
dsRNA
unknown
?
TLR2 TLR1
TLR2 TLR6
TLR5
TLR9
TLR7
TLR8
TLR4
TLR3
TLR10

Figure 2. Toll-like receptors.
A Toll-like receptors (TLRs) are either membrane-bound receptors or within endosomal compartments. (Illustration created with BioRender). **B** The different molecules recognized by TLRs to initiate an immune response. Illustration from [44].

1.6.4 Antigen sampling in the intestines

Macrophages can sample antigens and phagocytose foreign bodies through different pathways (Figure 3). In the intestine there are multiple ways for antigens to be transported across the epithelial layer by the means of different cell types, such as goblet cells and M cells. In the case of goblet cell-dependent antigen transport, the mucus producing goblet cells may transport antigens of low molecular weight so that it can be taken up and presented by APCs such as macrophages. M cells (epithelial cells specialized in transcytosis) are able to internalize luminal antigens through phagocytosis, endocytosis or macropinocytosis. The internalized antigen is then made available for uptake in intraepithelial pockets. Another cell type found in the intestinal epithelium is enterocytes. These cells express neonatal Fc receptors (FcRn) which is transcytotic, enabling antigen uptake. Antigen may additionally be sampled through phagocytosis of apoptotic cells. As mentioned above, macrophages are also able to sample antigens through projecting protrusions in the junctions between epithelial cells to phagocytose luminal antigen. By receptor-ligand interaction, macrophages may recognize these antigens. A signaling cascade within the macrophages is initiated when receptors on their surface recognize and bind ligands on the surface of the antigen, resulting in phagocytosis [45–47].

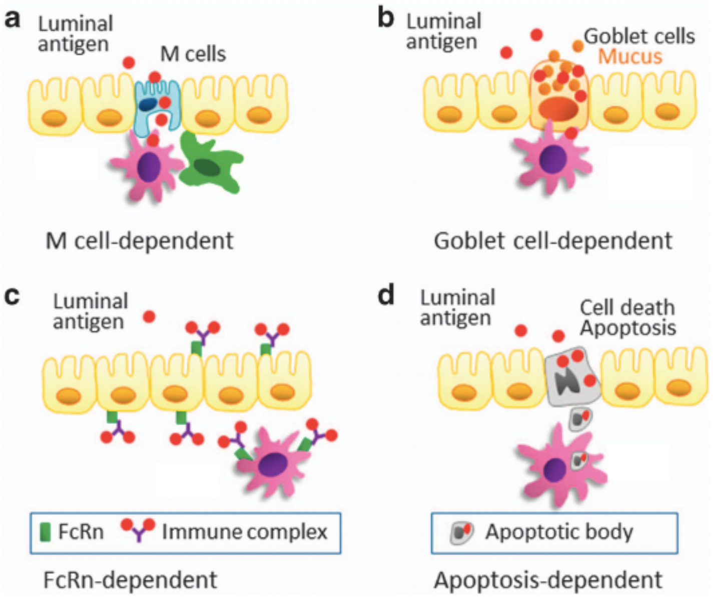
a
Luminal antigen
M cells
M cell-dependent
b
Luminal antigen
Goblet cells
Mucus
Goblet cell-dependent
c
Luminal antigen
FcRn
Immune complex
FcRn-dependent
d
Luminal antigen
Cell death
Apoptosis
Apoptotic body
Apoptosis-dependent

Figure 3. Different pathways for antigen sampling in the intestines.
Macrophages may sample antigens in the intestines through different pathways. **A** M-cell dependent antigen sampling, **B** Goblet cell-dependent sampling of antigen. **C** FcRn-dependent antigen sampling. **D** Antigen sampling through apoptotic cells. Illustration adapted from [45].

1.7 The colon

The large intestine (also called colon) is part of the gastrointestinal tract and measures around 1.5 meters in adults. The colonic epithelium is responsible for absorption of 90% of the water and salt from the proximal colon. The absorption of sodium is important for the water absorption. Bicarbonate is secreted in the colon to maintain the acid-base homeostasis, since the bacteria in the colon produce organic acids which needs buffering. The colon may be divided into four parts: the ascending colon measuring about 15 centimeters, the transverse colon (around 45 centimeters), the descending colon (about 25 centimeters) and the sigmoid colon (around 35-40 centimeters). Within the colon there are different layers, namely the mucosa, submucosa, muscularis propria, and serosa. The mucosal layer is composed of a single layer of epithelial cells facing the lumen, and subepithelial connective tissue in the submucosa, followed by a muscle layer named muscularis mucosae (Figure 4). This mucosal layer is responsible for the water and electrolyte absorption. Covering the mucosal layer is a layer of beneficial bacteria, termed commensal bacteria, generating the colonic microbiota [48–50].

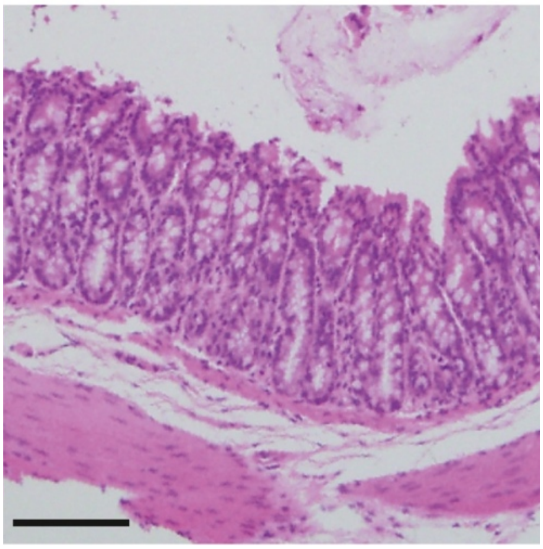

Figure 4. Overview of the colon.
Here, the histology of the colon is shown where the mucosal layer is at the top, followed by a thin layer of submucosa covering the muscular layer termed muscularis. Picture adapted from [17].

1.7.1 Colonic microbiota

The colonic microbiota is composed of a vast amount of bacteria living in a symbiotic relationship with the host. The microbiota arises just after birth by colonization of bacteria derived from the mother and the surroundings. Different bacterial species appears at different timepoints during the first bacterial colonization, as the newborn is exposed to diverse events such as *Escherichia coli* in the birth canal and *Lactobacillus* when nursing [51].

Bacteria within the colon contribute to nutritional uptake and metabolism, as well as stimulation of the intestinal tissue to mature. For instance, they play an important role in the breakdown of undigestible fibers. In return, the bacteria are given access to nutrients and a place to grow. This bacterial growth is closely monitored by the host immune system together with genetics. By allowing commensal bacteria to cover the colonic lumen pathogenic bacteria are unable to settle and expand. In this way colonic microbiota functions as a protective barrier against pathogens. Commensal bacteria also contribute to increased secretory IgA, mucin and antibacterial peptide production within the colon [51].

Genetic alterations influencing the immune responses in combination with a dysbiotic microbiota pose a threat to the host, in form of potential chronic inflammation known as inflammatory bowel disease (IBD). IBD includes diseases such as ulcerative colitis and Chron’s disease. Here, there is insufficiency in the immune system caused by genetic alterations, resulting in generation of immune responses to the commensal bacteria. In such diseases, macrophages play an important role, and it is therefore important to understand their biology in order to identify new therapeutic targets.

1.7.2 Macrophages in intestinal diseases

In IBD, macrophage function is disturbed where the process of generating anti-inflammatory tolerogenic macrophages are failing, leading to high amounts of pro-inflammatory monocyte-like cells. These cells also have an increased TNF and IL-23 production, contributing to intestinal inflammation as they both are pro-inflammatory cytokines. Intracellular bacteria are able to survive for a longer period of time within macrophages under IBD conditions, enabling them to generate pathogenic Th17 expansion. Interestingly, genetic alterations affecting genes associated with macrophage functions, including PRRs and proteins involved in cytokine signaling, are highly associated with IBD development and are most likely the cause of this failing of monocyte-like cells to differentiate into macrophages, which they do during homeostasis in healthy intestines [16,52].

Macrophages may be used as targets in IBD treatment. These treatments might for instance focus on transcriptional inactivation of pro-inflammatory molecules by the use of corticosteroids. There are several anti-inflammatory drugs, as well as biological drugs aiming for inhibition of binding of certain pro-inflammatory agents such as TNF. Further investigation and discovery of new therapeutic targets and techniques is continuously emerging in the field [16].

1.8 RNA

RNA (ribonucleic acid) is a polymer with ribose in the backbone, as opposed to DNA having deoxyribose as a backbone component. It is synthesized from a template DNA; the four bases in RNA are uracil (U), adenine (A), guanine (G), and cytosine (C). Reverse transcriptase is an enzyme capable of synthesizing double-stranded DNA from a single-stranded RNA template [53]. RNA has different functions; messenger RNA (mRNA) encodes proteins, transfer RNA (tRNA) and ribosomal RNA (rRNA) are important for ribosomal function [54].

1.9 Bulk RNA sequencing

To assess the gene expression in cells, the traditional option is to use bulk RNA sequencing. In this method, thousands or millions of cells are pooled and sequenced and the resulting data contains an average of the gene expression patterns for all the sequenced cells [55]. The sequencing starts with a sizable cell population or tissue that has been homogenized, from which RNA is extracted. After sequencing, it is possible to investigate several properties such as alternative splicing and gene expression, but the use of average expression might contribute to loss of central biological information [54]. The heterogeneity within the cell population is not uncovered when using this sequencing method, which might be disadvantageous.

1.10 Single-cell RNA sequencing

The whole genome expression profile depicts which genes each cell expresses. In order to obtain a measure of this expression profile, single-cell RNA sequencing (scRNA-seq) is used [55]. From the measure of gene expression it is possible to investigate gene expression information such as pattern of co-expression and splicing. The gene expression data might also give information about networks of gene regulation and genes that are co-regulated. Heterogeneity and rare cell populations within cell types may be identified and studied when performing this sequencing technique, as it opens up for transcriptomal comparison. Sequencing depth and number of sequences are two variables in scRNA-seq that need to be considered in order to perform an optimal sequencing; the amount of unique cell states (e.g. mature and immature cells) determine the number of cells needed for sequencing, while the extent of variance between the states determines the sequencing depth [56].

1.10.1 Workflow

The sequencing workflow consists of 5 steps; isolation of cells, lysis, transcription from mRNA to cDNA, amplification of cDNA, and finally, sequencing [56].

Step 1. The first step in cell isolation is preparation of the tissue, which must be performed shortly after surgery and in a cool environment to prevent autolysis. During tissue preparation transcription inhibitors are added, e.g. by the use of flavopiridol, followed by digestive enzymes or mechanical force to obtain single cells in suspension. When performing this separation there may be some complications along the way, such as cells might be damaged and transcription may be altered as a result of the isolation, making this part of the sequence preparation challenging. Here, enzymatic digestion by the use of Liberase is a well suited choice as it contains small amounts of lipopolysaccharides (LPS) and has the ability to avoid destruction of epitopes. Resulting from this tissue digestion is single cells in suspension. From this, single cells must be isolated, for instance by using the 10x platform. This method is based on droplets, where the cells attach to beads with distinct primers on them. The primers are barcoded, linking the transcripts to the cells of origin. When the cells are connected to the beads, they become imbedded in an oil emulsion, creating a droplet. Another way to isolate single cells from suspension is by the use of fluorescence-activated cell sorting (FACS). This method uses flow cytometry to sort cells based on phenotypes. When using FACS, one is also able to perform sorting of several populations of cells at the same time, which is highly convenient in many cases. To sort the different cells in suspension, the cells are fluorescently marked using fluorescently tagged antibodies directed at particular cell markers [54,57,58].

Step 2. When the cells of interest are sorted and captured the RNA needs to be exposed to perform the sequencing analysis, this is done by lysing the cells. Lysis is the process in which the outer membrane of the cells is broken, exposing the intracellular components. In general, there are two types of lysis; complete lysis, where the whole membrane is destroyed, and partial lysis, where only a part of the membrane is broken. For scRNA-seq, cells are for instance lysed enzymatically or by the use of detergents. When detergents are used, the membrane is not only destroyed, but the membrane proteins also become soluble. Enzymatic lysis, on the other hand, uses enzymes directed at spesific components of the membrane in order to destroy it. When the lysis is performed as in preparation from scRNA-seq a special lysis buffer for single cell lysis is used, made specifically for extraction of RNA from single cells. There are some advantages of using this particular lysis buffer in scRNA-seq, such as the sample loss is decreased when using this buffer, and it works well with reverse transcriptase (an enzyme enabling transcription of cDNA from mRNA) [59].

Step 3. After the cell lysis, the RNA is available for further analysis. Here, the RNA is of interest, but in order to perform the sequencing it has to be converted into DNA. This is done by transcribing the mRNA into cDNA, which is performed by the use of reverse transcriptase.

Step 4. The newly synthesized cDNA is now ready for amplification. To obtain adequate amount of sequences for the library preparation, it is essential to amplify the cDNA [54]. Template switching (TS) and *In vitro* transcription (IVT) [60] are the two principal approaches used for amplification . When TS is performed, it results in a cDNA of full length with a primer sequence attached. This primer sequence is used to synthesize the second strand, as well as the 5' end of the mRNA. Since the two ends of the sequence are identified, the amplification by PCR is more efficient. This is because the synthesis may start in both ends, not just one. IVT, on the other hand, performs linear amplification rather than PCR to decrease the amplification bias. After completed IVT, paired-end sequencing results in a read for the mRNA sequence and another one for the barcode of individual cells [54].

Step 5. After amplification, the sequences are aligned to a reference genome for gene name annotation [57]. To label every distinct mRNA molecule, unique molecular identifiers (UMIs) are used. They consist of 4-10 arbitrary nucleotides (nt) amplified with the cDNA. By using these UMIs it is possible to detect multiple amplifications of the same mRNA. Quantification of the amount of gene transcripts in a cell may also be obtained by calculating the number of UMIs present [55]; each UMI is only counted once [54].

1.10.2 10x genomics workflow

For this project the 10x genomics platform, specifically The Chromium Single Cell 3' v2 Reagent kit, was used to sequence single cells. This technology is based on microfluidics where cells in suspension are attached to gel beads with a unique barcode. Then, one cell will then be attached to one bead and consequently given its own specific barcode. Generation of cDNA from poly-adenylated mRNA is performed by incubation with particular primers containing a 10nt unique molecular identifier (UMI), a poly-dT primer sequence, an Illumina sequencing primer (Illumina R1 sequence), and 10x barcode with 16nt. The bead-cell interaction is then broken to capture the cDNA generated by each cell, which is achieved by separating the cell-containing beads by the use of an oil emulsion called GEM (Gel Bead-In-EMulsions). The captured cDNA is then sequenced with the barcodes, enabling traceback to the specific cell. Before cDNA amplification by PCR, removal of primers and reagents from the GEM mix is performed by using silane magnetic beads. Following, the sequencing library is prepared where cDNA with an optimized size is generated by size selection and enzymatic fragmentation. The resulting cDNA contains primers needed for Illumina bridge amplification. The sequencing libraries generated are compatible with Illumina, containing paired-end sequences. For this project, the Illumina NextSeq500 using HighOutput flow cells v2.5 was used for library sequencing. The sequencing raw data may then be analyzed in the 10x Genomic's Cell Ranger software. Further bioinformatic analyses might be performed following the initial Cell Ranger analysis [61].

1.10.3 Technical and biological variation

Technical and biological variation must be taken into account when performing scRNA-seq. Technical variation may occur as the lab work is performed; there might be several different persons preparing the samples slightly different. These small differences are almost impossible to avoid. Dropout events are also a technical variation to consider. Since beads do not capture the mRNA at the same proportion in every droplet, there might be some "empty" droplets where there are no mRNA present [57]. A method to reduce the technical variance is to use UMIs, which decrease the amplification bias [56]. Cell cycle effects are a biological variation that is usually corrected for, for instance by simple linear regression. The same method is used when correcting for mitochondrial gene expression. When correcting for biological variance there are some considerations to be aware of; cell cycle effects might give useful information, processes may be linked, and cell size can influence transcription effects [62]. Information about these processes may be lost due to these corrections.

1.10.4 Analysis

Computational analysis is necessary when scRNA-seq data is being analyzed. In general, there are 3 steps in the computational workflow used to analyze scRNA-seq; quality control and removal of low quality cells, normalization, finding highly variable genes and marker gene identification [55].

Step1. Cells that are damaged, stressed or dead are considered as low-quality cells, and removed from the sequencing data when the quality control (QC) is performed [56,63]. In order to identify these low-quality cells there are some metrices that are commonly looked at, aiming to eliminate them from additional analysis downstream [57], such as normalization, clustering and trajectory inference. These metrices include evaluation of the number of mitochondrial genes and the level of gene expression. If there are high levels of mitochondrial genes present, it may indicate stressed or dead cells and these should then be removed. Also, the gene expression gives an indication of the quality of the cells; low-quality cells will only express a small number of genes compared to high-quality cells. But cells of low quality might also have an extensively high amount of gene expression. This is due to the formation of multiples (i.e. several cells in one droplet) [64]. Principal component analysis (PCA) is a mathematical algorithm reducing data dimensionality, which can also be used [55,56]. This analysis conserves most of the variation within the sequencing data, while the highest variation, called principal component, is found and used to perform the reduction. This dimensionality reduction makes it easier to visualize the data in plots, as the number of components are greatly reduced [65].

Step 2. Normalization is a method used in order to compare results with different measures to one another by converting all the results to one common measure. E.g. when comparing two cell populations of different size to each other, the large cell population might appear to have a higher gene expression than the smaller population; this may be the case, but it is also likely that this is just due to the fact that there are more cells present. In order to investigate this, normalization is performed where the population size is corrected for in order to reveal gene expression differences between the two. Normalization enables removal of batch effect (i.e. systematic differences from one batch to another as a result of technical variations such as time of preparation, different people, and different protocols), and dropout events (where mRNA capture differs between droplets) [57,66]. One normalizing approach assumes that all cells start with the same amount of mRNA. Here, the differences in number of sequences are thought to be due to technical variations and do not represent actual biological variance [62]. These differences are therefore considered as bias and removed by scaling [55].

Step 3. Highly variable genes (i.e. the genes that are the most differentially expressed) indicate which genes are important for the heterogeneity of the cell population [67]. Identification of marker genes is performed to find specific genes that are differentially expressed in subpopulations, and the marker genes are then used as an identifier for the cells within that subpopulation. In order to be a marker gene, the chosen gene has to have a robust differential expression compared to the other subpopulations [55].

1.10.5 Single-cell RNA sequencing compared to bulk RNA sequencing

Single-cell RNA sequencing utilize a low amount of captured input RNA and it is a lot noisier compared to bulk RNA sequencing. This noise is due to the low amount of input RNA and subsequently drop-out events where there are no captured RNA. As a result of low RNA capture, there are less gene expression detected, and transcripts per million (TPM) is used as normalizing unit instead of reads per kilobase per million (RPKM) used in bulk RNAseq. Since there are temporal fluctuations in transcription, meaning that there are some sporadic transcription, scRNA-seq will contain zero-observation where there are no mRNA. When scRNA-seq is used it is possible to uncover biological variability to a larger extent than with bulk RNA sequencing, and a more complex transcript distribution is obtained [55,56].

Table 1: A summary of the comparison between single-cell RNA sequencing and bulk RNA sequencing.

Bulk RNA sequencing	**Single-cell RNA sequencing**
• Measure the average gene expression pattern for all sequenced cells • Higher RNA capture • Normalizing unit: reads per kilobase per million (RPKM) • Unable to investigate heterogeneity • Loss of biological variability	• Measure the gene expression in each of the sequenced cells • Lower RNA capture • Normalizing unit: transcripts per million (TPM) • Possible to investigate heterogeneity • More variable data due to technical noise

1.11 Bioinformatic tools for scRNA-seq analysis

There are several different programming languages, such as R and Python, which enables analysis of the scRNA-seq data. When using these programming languages, toolkits designed for specific tasks may be installed to perform the desired analyses. These tasks might for instance be cell clustering- and trajectory analysis. Scanpy is a Python-based toolkit able to perform both clustering and trajectory analysis, as well as other analyses steps. With this toolkit it is possible to perform all the analyzing steps mention above (1.10.4) [68]. The analyses provided by Scanpy may also be done by using the programming language R. Here, different toolkits might be utilized, such as Seurat and Monocle 3, which were used in this project.

1.11.1 Seurat

Seurat is an R toolkit aiming to incorporate various data sets from scRNA-seq for analysis of heterogeneity within cell populations [69]. In addition to the general computational workflow (see 1.10.4), Seurat allows for several further analysis steps. After identifying the highly variable genes there is a scaling step, where the data undergoes a linear transformation. By scaling the data, the gene expression from each cell is equally considered in the further analysis (i.e. genes that are highly expressed is not overriding the less expressed genes) [64].

The dimensionality of the data is reduced after scaling. This is achieved by principal component analysis (PCA). As described above (0) this mathematical algorithm conserves the variation within the data while reducing the dimensionality. It is useful to find the significant principal components (PCs), which are chosen based on the p-value. Plots like JackStrawPlot visualize the p-value for all PCs and may enable a selection of the most significant PCs. Another visualization method for the PCs is ElbowPlot, where the variance for each PC is plotted. In this plot an “elbow” is created where the variance flattens out. This “elbow” indicates which PCs holds the greater part of the information (i.e. the PCs to the left of the elbow) [64].

Seurat uses the PCA scores in order to cluster the cells, which is made possible by the dimensionality reduction of the data. In order to perform this clustering, the Louvain algorithm is used. This algorithm moves the clustering nodes around to generate the optimal clustering [70]. The clustering might be visualized by performing Uniform Manifold Approximation and Projection (UMAP) or t-distributed Stochastic Neighbour Embedding (t-SNE). UMAP generates a graph representing the data with a high dimension. From this high dimension graph a new graph is made with a low dimension, which is then plotted in two dimensions [71]. The t-SNE plots the high dimensional data, where each point is first placed at random. Then the points interact with each other following the two principals that (1) all points are drawn to their closest neighboring point and (2) the points repel one another. Both these methods perform a non-linear dimensionality reduction. Here, cells are grouped together if they have a similar expression profile [64,72].

Seurat is able to identify the defining markers for the clusters created. In each cluster, Seurat may identify both the negative and the positive markers by comparing it to the other cluster. It is also possible to specify for instance only positive markers. This comparison and identification might be done for all clusters (i.e. each cluster is compared to all the other cluster) or specific clusters may be chosen for comparison to each other [64].

When the clustering is performed and the defining markers are identified, it is possible to give names to the clusters. Usually, the names given are cell types based on the defining markers for each cluster [64].

1.11.2 Monocle 3

Monocle 3 allows for trajectories to be made, in order to get a more dynamic view of the cells (e.g. how the cells differentiate). To create these trajectories the data needs to be preprocessed, normalized, and clustered, which may be done within the Monocle toolkit. An alternative to this approach is to use a data set that have already gone through preprocessing, normalization and clustering in a different toolkit, such as Seurat. This is an advantagous feature of Monocle since it enables comparison between the clusters made by Seurat, and the trajectories made by Monocle. If the clustering is performed in Monocle it will differ from the clustering done in other toolkits. This is because each clustering will give a different result. The principle for creating a trajectory is based on the order of changes in gene expression. When Monocle learns these changes, it is able to generate a trajectory and place the different cells along the trajectory line. If more than one trajectory is to be made, Monocle might make new subgroups with cells similar to each other, from the already clustered cells before creating the trajectories. By creating subgroups Monocle is able to identify several different paths in the gene expression [73].

Depending on the type of trajectory (i.e. looped or tree-like trajectories), there are two different ways for Monocle 3 to learn the principle graph(s). This is an important step in order for Monocle to place the cells in the trajectory. One method is by the use of L1 Graph, which is able to learn looped trajectories. The L1 Graph is robust to data noise as it looks at the data points in a context (i.e. looking at one point in relation to the rest) rather than comparing two and two points. As opposed to the *k*-nearest-neighbor, used for instance in clustering, the L1 Graph method is an automated method were the sparsity is not manually decided [74]. Another method, learning trajectories resembling trees, is SimplePPT. Here, principal points and the tree structure are both learnt at the same time. The dimensionality of the data is not reduced further when using this method [73,75].

Clustered cells and chosen genes may also be shown as a function of pseudotime. For the clustered cells this is visualized by a color scale ranging from dark purple (early) to yellow (late). Early pseudotime corresponds to cell in the early time of differentiation (e.g. immature cells) that are on their way to the differentiation endpoint. Late pseudotime is close to this endpoint. When visualizing spesific genes as a function of pseudotime, the same principles apply (i.e. if the gene of interest is highly expressed early in pseudotime it is likely a gene representing the early stages of differentiation) [76].

2 Materials and methods

2.1 Patient information

The patients included in this project underwent sigmoid colon cancer surgery performed at Akershus University Hospital (Lørenskog, Norway). Healthy colon at least 10 cm from the tumor was used for further analysis. For scRNA-seq, colons from 4 patients were used (age 62-78, 3 males), and colons from 3 patients were used for immunohistochemistry (age 62-78, 2 males).

2.1.1 Ethical considerations

All experiments described were performed within closed laboratories with virtually no effect on the external environment. This project conforms to EEC Directives concerning medical research and the corresponding Norwegian legislation. All biopsy and surgical resection specimens of human origin were taken after informed consent of the study subjects. The experiments and biobanks are approved by the Regional Committee for Medical Research Ethics (REK 2015/946).

2.2 Tissue preparation

In order to obtain single cells from the colon, tissue dissection was performed [20,77]. The colons we received from Akershus University Hospital were dissected within 2 hours after surgery. First, the colon was opened longitudinally to expose the mucosal layer and transferred onto a large petri dish. The colon was dipped a few times in PBS before being gently dried off with a paper towel, on a large cell culture dish, to remove any mucus and other residues. Secondly, the mucosal layer was excised by cutting off thin strips of the mucosal folds and transferring them into a 50 ml tube containing about 20 ml EDTA buffer (Appendix 1). The thin submucosal layer was subsequently removed. This layer was not used further on. To remove the teniae the colon piece was flipped over, where the teniae are clearly visible as three thick stipes running along the length of the colon. They were cut off and moved to a new tube with around 20 ml EDTA buffer. Then, only the muscularis layer was left which was cut into smaller pieces and, also here, moved to a tube containing 20 ml EDTA buffer. The three tubes with EDTA buffer and the different tissue layers were then filled up with additional EDTA buffer so that the total volume was 50 ml in each tube. Next, the tubes were placed in the CO_2 incubator on the Hula-mixer for 15 minutes at 37°C for tissue washing to remove epithelial cells.

While the tissue was being washed the enzyme mix for digestion of tissue was made in 100ml cups, containing RPMI without phenol red, liberase and DNase (Appendix 1). When the tissue was ready, the top layer of supernatant was gently removed so no tissue was lost. A 100 μm piece of mesh was then placed in a funnel to filter the remaining supernatant with the tissue. The mesh containing the tissue was then washed with PBS and squeezed to remove as much epithelial cells as possible. The mucosal layer needed to be washed three times because there are a lot of mucus and other intestinal residues present in the lumen of the colon, that needs to be removed in order to only isolate the cells of interest. Between each wash, the filtering step was performed, and the mucosal tissue was transferred to a new tube containing EDTA buffer. While the mucosal layer was being washed, the remaining two layers (which were only washed once) were transferred to a petri dish and minced by using scissors. The same procedure was done for the mucosal layer as well. The minced tissue was then transferred to a 100 ml cup, containing the enzyme mix; one cup for each layer. A magnet was added and the cups, with the lid on, were placed in the CO_2 incubator on a magnetic stirrer at 250 rpm for 60 minutes at 37°C.

After the incubation with the enzyme mix, the different layers were filtered through a 100 μm filter on top of a 50 ml tube (one filter and one tube for each layer). The tissue was dispersed by pipetting up and down and then put on the filter. The narrow part of the pipet opening was cut to make it bigger making it easier to pipet out the tissue mix, which contained some larger chunks as well. On the filter, the tissue was mixed around to speed up the filtration process; this is not a crucial step in the protocol, but if it is not done it will take hours before the tissue has passed. PBS was then added to the cup with the enzyme mix, to make sure that all of the tissue was collected, and transferred to the filter by pipetting.

When the filtration was complete, PBS was added to the three tubes so that the total volume in each tube was 50 ml. The cells were then spun down at 500g for 10 minutes. After spinning, the supernatant was discarded, leaving around 5 ml behind, the cell pellet was resuspended and filtered one more time (as described above) and spun again. Following the second spin, the supernatant was discarded so that approximately 5 ml was left in each tube. The cell pellet was resuspended and ready for counting.

The teniae cells were diluted 1:10 for the cell counting, by adding 90 μl of PBS to an Eppendorf tube and then adding 10 μl of the resuspended cells. The mucosa and muscularis cells were diluted 1:100 by first making a 1:10 dilution, as described, then adding 90 μl PBS to a new Eppendorf tube and finally adding 10 μl of the 1:10 cell dilution. The difference in dilution is due to the number of cells captured. When isolating cells from the mucosal and muscularis layer, other “unwanted” cells will also be included in the cell suspension. This is due to the close proximity to epithelial cells and the submucosa, respectively. As a consequence, there will be a lot more cells isolated from these layers. The cells were counted by the use of EVE™ Automatic cell counter. In order for the cell counter to detect the cells, 10 μl of the diluted cells was mixed with 10 μl of Trypan Blue stain 0.4%. 10 μl of the stained cells were then added to a cell counting chamber slide and placed in the cell counter. Dead cells are identified by high uptake of the live/dead marker and consequently a high amount of color detected. This is due to the permeability of dead cells (i.e. the live/dead marker can leak into dead cells, but not live cells with the cell membrane intact). To calculate the total number of live cells, the equation described in Equation 1 was used.

The following analysis was performed on the mucosal cells, whereas the cells form muscularis and teniae were used in other projects performed by the group.

Equation 1: Calculating the total number of live cells.

$$Number\ of\ live\ cells \cdot Cell\ resuspension\ volume \cdot Dilution = Total\ number\ of\ live\ cells$$

2.3 Isolation of mononuclear cells from peripheral blood

In order to isolate mononuclear cells from peripheral blood, a lymphoprep protocol was performed [78]. First, the blood was diluted with room temperature PBS in a 50 ml tube to a total volume of 40 ml and 10 ml lymphoprep was pipetted in a different 50 ml tube. The blood and PBS mix was then transferred carefully to the lymphoprep tube, holding the tube in a 60° angle and pipetting along the side, so that the blood lies on top of the lymphoprep. The tube containing blood and lymphoprep was spun for 20 minutes in the Beckman Model TJ-6 Centrifuge in order to separate the blood based on density. The buffy-coat (Figure 5) was pipetted and transferred to a new 50 ml tube and PBS was added for a total volume of 50 ml before the tube was spun at 250g for 10 minutes. After the spinning, the supernatant was discarded and the cell pellet resuspended in 10 ml PBS before doing a cell count as described above. Following the counting, PBS was added to the tube for a total volume of 50 ml and the tube was spun at 500g for 5 minutes. Finally, the supernatant was discarded and the pellet ,containing PBMCs, resuspended in 300 μl FCS.

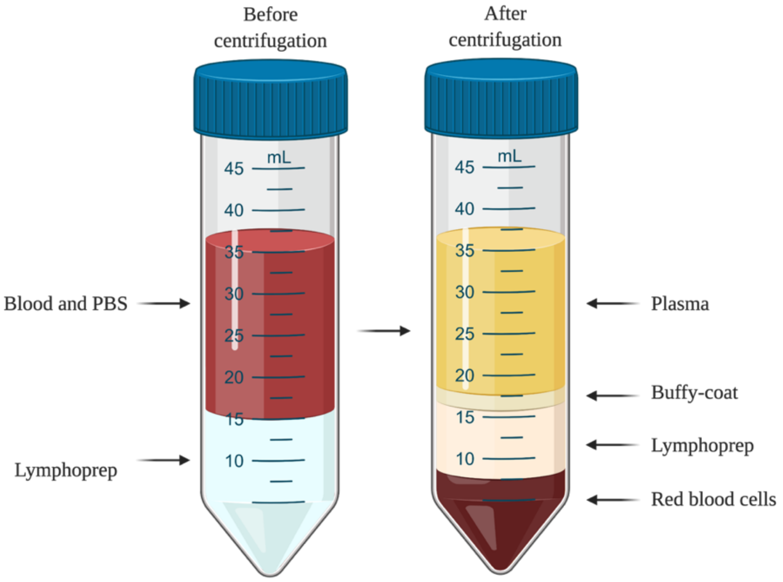

Figure 5. Lymphoprep of peripheral blood to isolate mononuclear cells.
Peripheral blood was diluted in PBS and transferred to a tube containing lymphoprep. After centrifugation the blood was separated based on density, where the plasma was the top layer follow by a buffy-coat (containing mononuclear cells), a lymphoprep layer and the red blood cells at the bottom. Figure created with BioRender.com.

2.4 Cell staining

Following the tissue preparation, the cells obtained were stained with antibodies for further analysis by flow cytometry. Different antibodies were used in order to identify several cell markers, both extra- and intracellularly. The aim here was to identify macrophages based on expression of CD45, HLA-DR and CD14 [20]. We also wanted to distinguish these cells from dendritic cells, which do not express CD14, in order to clearly show that CD14^{+} is indicative of macrophages. Antibodies for CD3 and CD19 conjugated with the same fluorochrome as the live/dead marker were used to exclude HLA-DR^{+} cells that were not macrophages, such as B and T cells (this exclusion is marked as lin^{-}).

First, approximately 1 million cells were transferred to 5 ml flow tubes, washed by adding 1 ml flow buffer (Appendix 1) and centrifuged at 400g for 4 minutes at 4°C. 1 million cells were used as it corresponds to one test for the antibodies (i.e. the volume per test for the antibody can be used). The supernatant was discarded, leaving only the cell pellet in the tube. Next, the cells were stained with only surface markers or a combination of surface and intracellular markers.

For the surface marker staining, 95 μl of the macrophages vs. dendritic cells or the DC3 master mix (Table 2) was added to each of the tubes containing the cell pellet. The cells then incubated on ice for 30 minutes in the dark. Following the incubation, the cells were washed again as described above. 400 μl of flow buffer was added to each tube after the supernatant was discarded, and then flow cytometry was performed. 10 μl of TO-PRO-1 was added to each tube prior to running the flow cytometer, working as a live/dead marker. Dead cells are identified by high uptake of the live/dead marker and consequently a high amount of fluorescence detected. This is due to the permeability of dead cells (i.e. the live/dead marker can bind intracellular proteins and give a stronger signal than when the marker only binds surface proteins on live cells).

Table 2: Antibodies used in the macrophages vs. dendritic cells and DC3 surface staining mixes for flow cytometry.

Macrophages vs. dendritic cells staining		**DC3 staining**	
Antibodies	*Volume per test (µl)*	*Antibodies*	*Volume per test (µl)*
Fc Block	10	Fc Block	10
BerEp4 FITC	10	BerEp4 FITC	10
CD3 FITC	5	CD3 FITC	5
CD19 Ax488	5	CD19 Ax488	5
CD45 BV510	5	CD45 BV510	5
HLA-DR PerCP-Cy5.5	2.5	HLA-DR PerCP-Cy5.5	2.5
CD14 PE-Cy7	5	CD14 APC	5
CD11c APC	5	FcϵRI PE	5
CD1c BV421	5	CD1c BV421	5
CD141 PE	10		
Total volume (antibodies):	62.5	Total volume (antibodies):	52.5
Added buffer:	37.5	Added buffer:	47.5
Total volume (master mix):	100	Total volume (master mix):	100

Isolated cells were also stained intracellularly in addition to the surface staining. In order to accomplish both intracellular and surface staining, the cells were treated with a fixation and permeabilization approach [79]. First, the cells (around 1 million) were washed, as described above, and then incubated on ice for 30 minutes in the dark with 1 ml of a Fixable Viability dye eFluor 780 live/dead master mix; this mix was made by adding 10 μl of Fixable Viability dye eFluor 780 to 10 ml PBS [80]. After incubating, the cells were washed with 1 ml PBS and centrifuged at 400g for 4 minutes at 4°C. The supernatant was discarded and 95 μl of the surface master mix (Table 3) was added to the cells, which then incubated for 30 minutes on ice in the dark. Following the incubation, the cells were washed two times with 1 ml flow buffer (according to the protocol) and spun as described above. The cells were resuspended in 250 μl Fix/Perm Solution and incubated for 20 minutes on ice in the dark. Next, the cells were washed two times with 1 ml of a Perm Wash solution containing 5 ml PermWash and 45 ml distilled H_2O. After the cells had been washed, 95 μl of the intracellular master mix (Table 3) was added and the cells incubated on ice for 30 minutes in the dark. Finally, the cells were washed with flow buffer as described above, and resuspended in 400 μl flow buffer before performing flow cytometry.

Table 3: Antibodies used in the mature/immature macrophages surface and intracellular staining mix for flow cytometry.

Surface staining		**Intracellular staining**	
Antibodies	*Volume per test (µl)*	*Antibodies*	*Volume per test (µl)*
Fc Block	10	C1Q FITC	1
CD45 BV510	5	Calprotectin PE	10
HLA-DR PerCP-Cy5.5	2.5		
CD14 APC	5		
Total volume (antibodies):	22.5	Total volume (antibodies):	11
Added buffer:	77.5	Added buffer:	89
Total volume (master mix):	100	Total volume (master mix):	100

2.5 Flow cytometry

Before the flow cytometry analysis was performed, a CST test was done on the BD LSRFortessa X20 machine. When performing this test beads specific for CST, that all have the same intensity in the fluorescence and are of the same size, are run. This test performs a quality control for the lasers, fluidics and electronics, as well as the optics within the flow cytometer [81].

Compensation beads were run prior to running the stained cells. These beads work as single-color controls used to calculate the compensation matrix in FlowJo later on. The beads were made by applying one drop of UltraComp eBeads™ to a flow tube, adding 1 μl of an antibody conjugated with the desired fluorescence and mixing well. The beads then incubated on ice for about 30 minutes in the dark. Compensation beads for all the different fluorescence mentioned above were made. 10 000 events of the beads were recorded.

The stained cells were then run, where 1 000 000 events were recorded. The forward scatter was adjusted so all events lied nicely inside the plot. All of the analyses were performed in FlowJo, but some gates were set when running the samples as well, in order to get an overview of the cell distribution (i.e. see that there were CD45$^+$ cells present, and that these cells also express CD14 and HLA-DR, which indicated that they were macrophages).

2.5.1 Analysis of flow cytometry data

The flow cytometry data were analyzed using FlowJo. A compensation matrix was calculated automatically in FlowJo by using the compensation beads. This matrix was applied to the sample data. When looking at the sample data, multiple gates were set in order to obtain the desired cells (Figure 6). First, a broad gate based on size (forward scatter) and granularity (side scatter) was set to include the region encompassing macrophages (Figure 6A). From the cells within this gate, a new gate was set to obtain all the live CD45$^+$ cells (Figure 6B). Here, we also remove epithelial cells, as well as T and B lymphocytes by using the antibodies BerEp4, CD3 and CD19, respectively. These three antibodies and the live/dead marker TO-PRO-1 are conjugated with fluorescence detected in the same channel, meaning that all of them will appear as one single parameter. Cells that were positive for either of the antibodies, including dead cells, were excluded from further analysis. Finally, HLA-DR$^+$ cells that ranged from CD14^{low} to CD14^{high} were gated for from the live CD45$^+$ cells (Figure 6C).

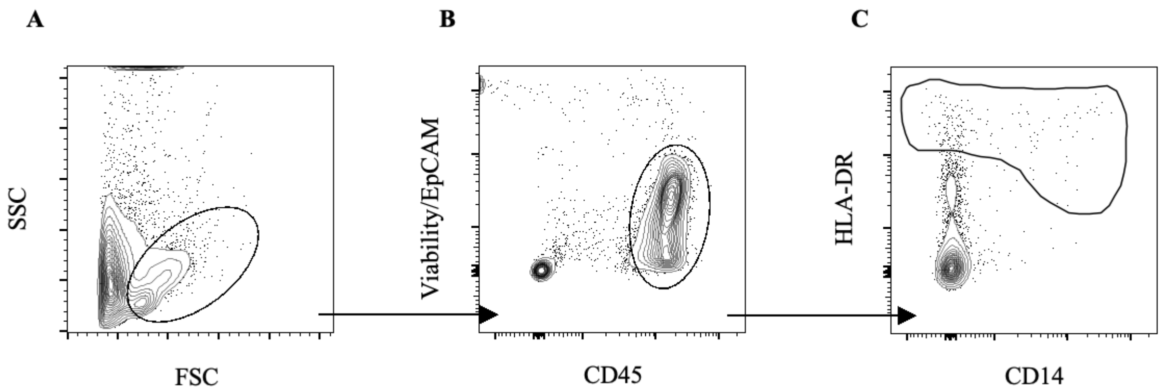

Figure 6. Representative flow cytometry general gating strategy for APCs.

Cells from the colon were stained with CD45-BV510, HLA-DR-PerCP-Cy5.5, and CD14-APC or PE-Cy7. Here, the general flow cytometry gating strategy, performed in FlowJo, is shown with the following gating strategy:

A All events are shown in a dot plot were, in the black circle, the region encompassing macrophages is gated out. Forward scatter is on the x-axis, indicating cell size, and side scatter on the y-axis, indicating cell granularity. **B** From the cell within gate A, a new gate is set for live cells (black circle). Here, CD45 is on the x-axis and the live/dead marker (TO-PRO-1 or Fixable Viability dye eFluor 780) on the y-axis. **C** Gated for CD14$^{\mathrm{lo\ to\ hi}}$ and HLA-DR^{+} (within the black lines) from the live cells. The black line is drawn around the cells that express HLA-DR and range from CD14$^{\mathrm{low}}$ to CD14$^{\mathrm{high}}$. Here, CD14 is on the x-axis and HLA-DR on the y-axis.

2.6 Cell sorting

The single cell sorting was performed by Espen Bækkevold. Single cells from the colon were isolated and washed in preparation for cell staining as described above (2.4). The cells were stained for CD45, HLA-DR and CD14 (Table 4) in order to sort out CD45^{+}HLA-DR^{+}CD14^{+} macrophages, captured in PBS with 0.5% BSA [20].

Table 4: Antibodies used for single cell sorting.

Antibodies	*Volume per test (µl)*
Fc Block	10
CD45 BV510	5
HLA-DR PerCP-Cy5.5	2.5
CD14 APC	5

2.7 10x genomics

Single cell gene expression analysis was performed by using the 10x genomics Chromium Single Cell 3’ Solution, and was outsourced to the Genomics Core Facility at Oslo University Hospital (Radium Hospital, Oslo, Norway). This part is only described in brief as these results are used for further analysis.

First, the FACS sorted CD45$^+$CD14$^+$HLA-DR$^+$ cells were merged with beads made of gel in an emulsion; one cell is attached to one bead. All cDNA produced from one cell would then have the same 10x barcode. Primers were added to the bead-cell solution to generate full-length cDNA and at the same time destroy the gel beads in order to retrieve the pooled fractions. The primers included, amongst other components, a 10 mt unique molecular identifier (UMI) and a 16nt 10x barcode. Disposal of primers and biochemical reagents that may be left in the suspension was done by the use of silane magnetic beads. To produce a large enough number of cDNA to create the library, PCR was used to amplify the barcoded cDNA. The library generated was sequenced on the Illumina NextSeq500 using HighOutput flow cells v2.5 (Figure 7) [61].

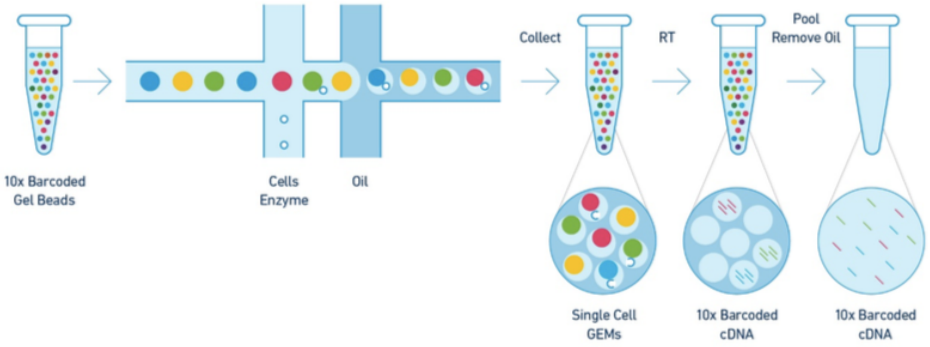

Figure 7. 10x genomics workflow.
Here, the workflow for single cell sorting by the use of 10x genomics is depicted. First, one cell is attached to a bead and imbedded in an oil emulsion. The Gel Bead-In-EMulsions (GEMs) are then collected and 10x barcoded cDNA is generated. Finally, the oil and biochemical reagents are removed. Illustration from [57].

2.8 Computational analysis of scRNA-seq data

The sequencing data generated from the 10x genomics was first preprocessed in Cell Ranger, before additional computational analyses in R were performed by the use of both Seurat and Monocle 3. The scripts used to perform the Seurat and Monocle 3 analyses are available on GitLab (https://gitlab.com/fjahnsenlab/master-thesis/single-cell-master-thesis).

2.8.1 Analysis with Cell Ranger

The preprocessing of the 10x genomics sequencing data was performed at the Genomics Core Facility at Oslo University Hospital (Radium Hospital, Oslo, Norway). First, FASTQ files were made before assessing the sequencing quality. The sequencing reads were then mapped to a reference genome and a gene-barcode matrix was generated. This gene expression matrix was then further used in Seurat to perform additional analyses.

2.8.2 Analysis with Seurat

First, the 10x genomics sequencing data from the four different patients was separately loaded into R by using the `Read10x()` function incorporated in the Seurat package. Quality control and filtration of the data was then performed, where cells having an RNA count of more than 200 and less than 2000 with a mitochondrial gene percentage of 5% or less was selected. Cells with an RNA count without this range were characterized as low-quality cells. Following filtration, the cells were normalized with a scale factor set to 10 000. This scale factor is multiplied with the gene expression measured from the normalization. The resulting value then becomes log-transformed [64].

When these initial analyses were done, the four different data sets (one for each patient) were merged together in order to perform an integrated analysis. First a list of all the data sets was created, the dimensions were set to 30, which largely capture the data, and then the Seurat function `FindIntegrationAnchors()` was used. This function identifies anchors between the different data sets. These anchors were then used to perform a principal component analysis (PCA) on the merged data. After the PCA the number of principal components (PCs) to use needed to be set. In order to choose the correct number some calculations were done. Firs, the percentage variation for each PC was found by dividing the standard deviation of each PC by the sum of all standard deviations multiplied by 100. The cumulative (i.e. increasing) percent for each PC was then used to find the PCs with a cumulative percentage higher than 90% and a variance of less than 5%. The last PC with a percentage variation of more than 0.1% was then calculated, before identifying the lowest PC between the two. This resulting number was the correct number of PCs to use.

The data was then clustered, based on the calculated number of PCs, and a UMAP plot of the clustered cells was generated. To obtain the correct resolution for the UMAP plot a clustering tree was made (Figure 8), depicting different resolutions ranging from 0.1 to 2. The resolution is a parameter determining the number of clusters in the downstream analysis. This selection of resolution is highly subjective, and each resolution will give a slightly different picture. For these analyses, the lowest resolution with fewest arrows going across (i.e. the resolution giving the most linear picture) was chosen.

Following the clustering of cells, the average expression within each cluster was calculated by using the Seurat function `AverageExpression()`. All positive variable markers in all clusters were also identified, based on RNA. These differently expressed genes (DEGs) were later used to annotate each cluster.

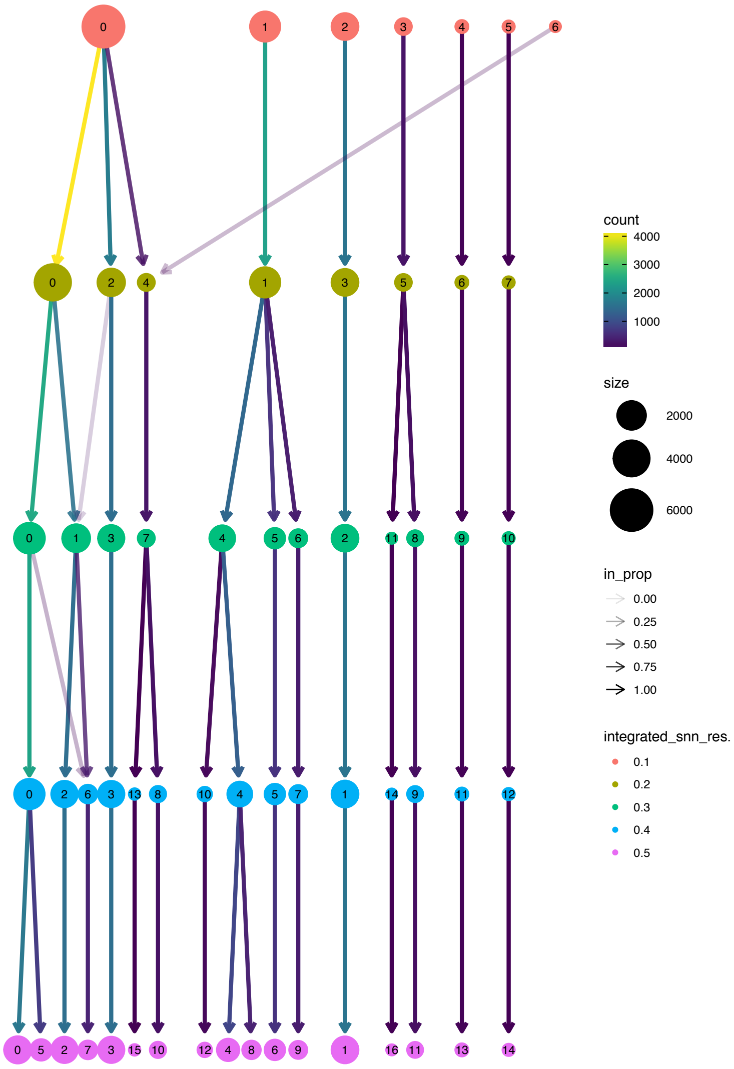

Figure 8. A representative clustering tree plot.
Here, a clustering tree plot is shown. Based on this tree the resolution for the UMAP plots may be chosen. The optimal resolution is the one where there are few dividing arrows, and at the same time having the lowest number of clusters (i.e. the resolution with the largest amount of arrows going straight down with lowest number of clusters is optimal).

2.8.3 Analysis with Monocle 3

The trajectory analysis was performed on the already clustered data set created by using the Seurat package. First, a cell data set (CDS) was created from the Seurat data, by using the gene expression. The Monocle 3 function `preprocess_cds()` was used to perform different preprocessing steps on the data, such as normalization. After the preprocessing, a principal component analysis was performed followed by a dimensionality reduction. To obtain the same information as from the Seurat analysis, the cell embeddings (i.e. cell coordinates) were not calculated by Monocle but rather imported from the Seurat data. The clustering was also based on the findings from the Seurat analysis.

Monocle then learns the principal graph (i.e. the biological differentiation based on gene expression), before being able to generate the trajectories within the clustered cells. These trajectories may be generated with or without defining a starting point (i.e. trajectory root). In this case, the trajectory root was defined as a single cell within the cluster identified as immature macrophages based on the differently expressed genes generated by Seurat.

2.9 Immunohistochemistry

In order to investigate the distribution of colonic macrophages *in situ*, immunohistochemistry was performed. The colon slides, being either formalin fixated or cryopreserved, were stained with different antibody mixes, to identify and localize immature and mature macrophages as well as DC3s. Both immunofluorescent and enzymatic staining methods were used.

2.9.1 Immunofluorescent staining of formalin fixated slides

First, the colon slides were deparaffinized by letting the slides sit or dipping in different solvent solutions for defined time, as described in Figure 9. For immunostaining of formalin-fixed tissues, heat induced epitope retrieval (HIER) is needed to unmask antigen epitopes that may be denatured, hidden or destroyed due to formaldehyde cross-linking [82]. The optimal protocol is usually determined empirically; by boiling in buffers with high or low pH for 20 minutes. The optimal HIER was established for all antibodies used in the study, as described below (2.9.2). After boiling, the slides are cooled down for 20 minutes in room temperature. Next, the slides were dipped in PBS before a DAKO grease pen was used to mark around the tissue. The slides were then blocked for 5 minutes with 10% Normal human serum (NHS).

Following the blocking, the primary antibodies (Table 5) were added to the tissue and incubated for 60 minutes at 37°C; one of the tissue samples worked as a negative control where only PBS was added, while the primary antibodies were applied to the other sample (Figure 10). To obtain the desired concentrations (1:100 for CD68, 1:400 for C1Q and 1:200000 for calprotectin), the antibodies were diluted; both C1Q and calprotectin had an initial dilution of 1:1 and in order to obtain the final dilution of 1:400 and1:200000, respectively, the antibodies had to be diluted in two steps. With C1Q, the antibody was first diluted 1:2 by adding 5μl of the C1Q antibody to

5 μl 1.25% BSA. For calprotectin, the antibody was diluted 1:1000 by applying 1μl of calprotectin to 1000 μl 1.25% BSA. Then the second dilution had to be 1:200 for both antibodies. To be able to make the dilution with C1Q or calprotectin in combination with CD68 in the same mix, the dilution for the CD68 antibody was calculated to match the C1Q/calprotectin dilution; since the start dilution of CD68 was 1:2 and the final dilution should be 1:100, the dilution made must be 1:50. In order to make this dilution and match the dilution of C1Q/calprotectin, the CD68 dilution was made as 4:200. This means that 1 μl of C1Q or calprotectin and 4 μl of CD68 was added to 195 μl 1.25% BSA for a total volume of 200 μl and a 1:200 dilution of C1Q/calprotectin and 1:50 of CD68.

After the incubation, the slides were washed twice in PBS for 5 minutes on a shaking table before the secondary antibodies (Table 5) were added and the slides incubated for 90 minutes at room temperature. The IgG3 antibody had an initial dilution of 1:10 while IgG was diluted 1:2, and to obtain the final dilutions (1:1000 and 1:200 respectively) a 1:100 dilution was made for both antibodies by adding 2 μl of each antibody to 196 μl 1.25% BAS for a total volume of 200 μl.

Next, the slides was again washed in PBS for 5 minutes and then in Hoechst (nuclear staining) for 5 minutes, both on the shaking table. The slides were then dipped in distilled H_2O and air dried so that the slides were almost completely dried. Finally, the coverslips were mounted by applying one drop PVA on top of the tissue.

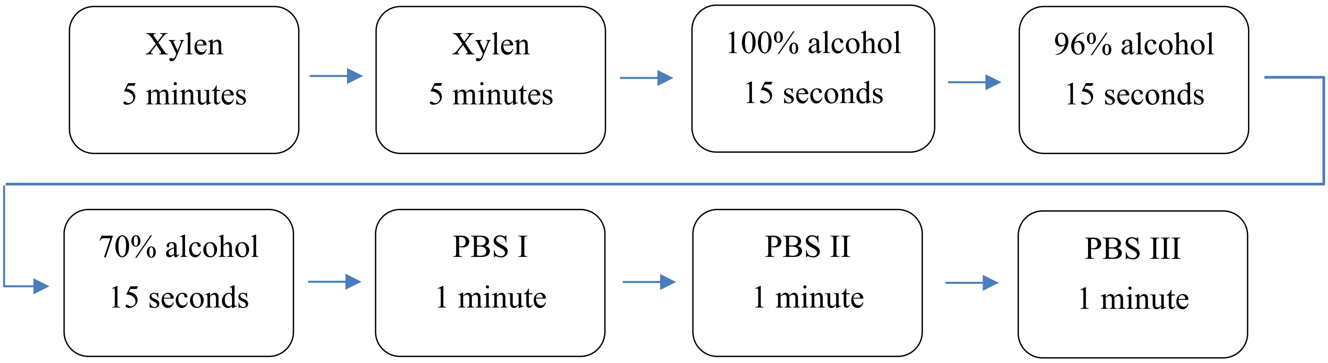

Figure 9. The reagents used for the deparaffination of slides for immunostaining.
Here, the workflow for the deparaffinization of the colon slides is shown. PBS I-III all contain PBS, the numbering only indicate that there are three different containers.

Table 5: Primary and secondary antibodies and their dilutions used for immunostaining of formalin fixated slides.

Primary antibodies			**Secondary antibodies**		
Antibodies	*Initial dilution*	*Final dilution*	*Antibodies*	*Initial dilution*	*Final dilution*
CD68	1:2	1:100	IgG3 Ax488	1:10	1:1000
C1Q	1:1	1:100	IgG (H+L) Cy3	1:10	1:200
Calprotectin	1:1	1:200000			

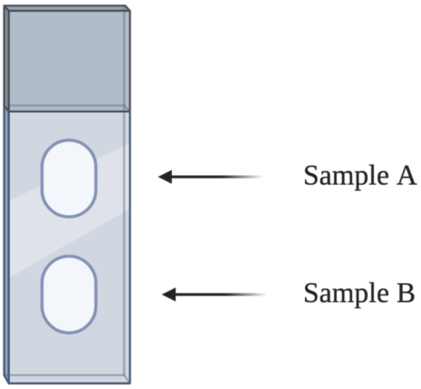

Figure 10. Immunostaining strategy.
On each slide, there are two tissue samples. In order to have a negative control, the two samples are treated differently; Sample A works as a negative control where only the secondary antibodies are applied, while both primary and secondary antibodies are applied to Sample B. Figure created with BioRender.com.

2.9.2 Enzymatic staining of formalin fixated slides

Enzymatic staining gives an increased antibody sensitivity compared to fluorescent staining. Therefore, formalin fixated slides was also stained following an enzymatic staining protocol. To find the optimal dilution and pH for HIER for the primary antibodies, manual staining with only one primary antibody was performed. The slides were first deparaffinized and boiled as described above (Figure 9), with a pH6 and a pH 9 buffer. Following 20 minutes of cooldown, the slides were put under running water for about 1 minute before marking around the tissue with a DAKO grease pen. Since the antibodies used for marker detection are peroxidase conjugated endogen peroxidase must be blocked to obtain expression only from the antibody-conjugated peroxidase. This was done by applying an endogen peroxidase block for 5 minutes. The slides were washed in TBS for about 1 minute before the primary antibody diluted in an antibody diluent (Table 6) was added and incubated for 60 minutes in room temperature. The secondary antibody DAKO DM822 was then applied after the slides were washed as described above and incubated for 30 minutes in room temperature. Another washing step followed the incubation prior to applying DAKO's DAB+ substrate for 5 minutes. Next, the slides were washed and put under running water for about 1 minute, before a nuclear staining was performed as described in Figure 11. Finally coverslips were mounted by using a xylen-based glue. Figure 12 shows representative images of the large difference between treating the slides with the same antibody dilution with different pH. The optimal pH was found to be pH 9 for both CD14 (Figure 12) and TREM1 (not shown) with the dilutions 1:50 and 1:100, respectively.

To investigate co-localization, two different antibodies were used. Since the secondary antibody used can bind primary antibodies from both mouse and rabbit, a sequential staining must be performed (i.e. the staining protocol described above have to be repeated two times, one for each primary antibody). This was performed on a Ventana DISCOVERY ULTRA automated slide preparation system. Also here, CD14 and TREM1 were used as primary antibodies, diluted in Discovery's antibody diluent. Here, the secondary antibody for CD14 was DISC Purple and Teal HRP Sub for TREM1 giving rise to a red and turquoise color, respectively.

Table 6: Primary antibodies and their dilution used for enzymatic staining of formalin fixated slides.

Antibodies	*Initial dilution*	*Final dilution*	*Optimal pH*
CD14	1:1	1:50	9
TREM1	1:1	1:100	9

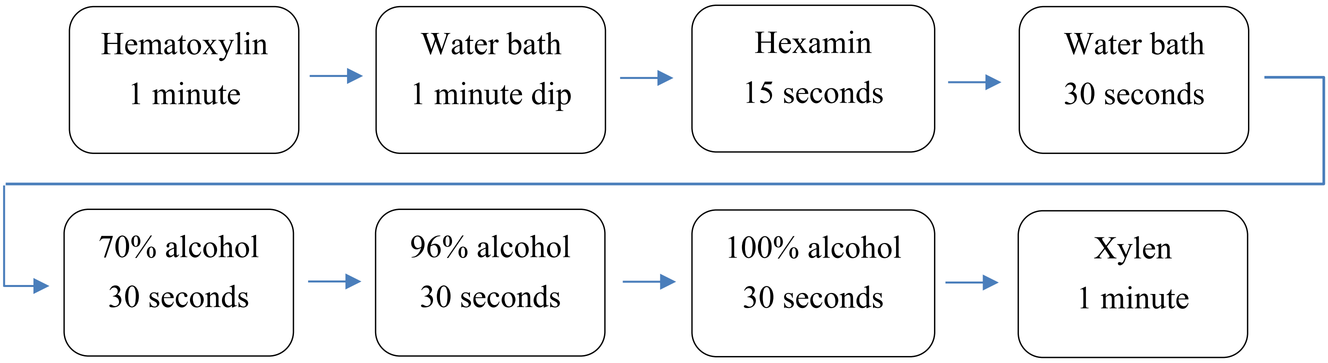
Hematoxylin
1 minute
Water bath
1 minute dip
Hexamin
15 seconds
Water bath
30 seconds
70% alcohol
30 seconds
96% alcohol
30 seconds
100% alcohol
30 seconds
Xylen
1 minute

Figure 11. The reagents used for the nuclear staining of enzyme stained slides.
Here, the workflow for the nuclear staining of enzymatically stained colon slides is shown

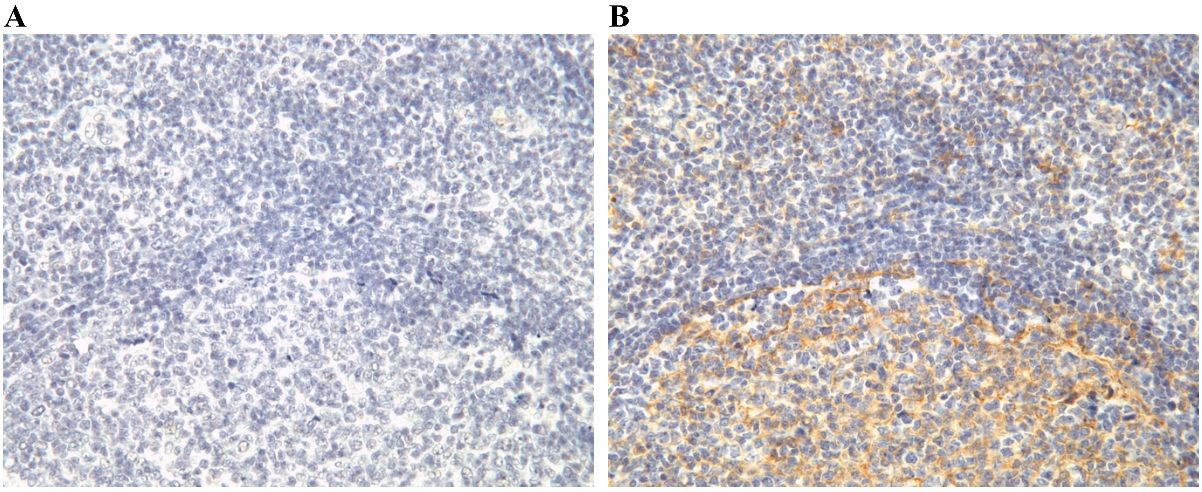

Figure 12. Representative image of enzyme staining of tonsils to determine the optimal pH.
Here, tonsil slides were stained for CD14 (shown in brown) with the dilution 1:50.
A This slide was treated with a pH 6 buffer. There are no visible CD14^{+} cells detected. **B** When treating the slide with a pH 9 buffer, several CD14^{+} cells are visible.

2.9.3 Immunofluorescent staining of cryopreserved sections

When staining cryosections, the pre-cut slides stored at -20ºC were first defrosted in room temperature for 5-10 minutes. A DAKO grease pen was used to circle around the slides, before the primary antibodies were added (Table 7). The antibodies incubated for 60 minutes in room temperature. Again, the top sample was used as a negative control. Following the primary antibodies incubation, the slides were washed two times for 3 minutes in PBS. After the wash, the secondary antibodies (Table 7) were added and incubated for 90 minutes in room temperature in the dark. Finally, the slides were washed for 3 minutes in PBS, then 3 minutes in Hoechst, before being dipped in dH_2O. After the slides were dried off, the coverslips were mounted with PVA.

Table 7: Primary and secondary antibodies and their dilution used for immunostaining of cryosections.

Primary antibodies			**Secondary antibodies**		
Antibodies	*Initial dilution*	*Final dilution*	*Antibodies*	*Initial dilution*	*Final dilution*
CD14	1:4	1:100	IgG3 Ax488	1:10	1:1000
CD1c	1:10	1:200	IgG1 Cy3	1:10	1:1500

3 Results

3.1 Colonic macrophages are confined to the HLA-DR$^+$CD14$^+$ cell population

In the small intestine Bujko *et al.* found macrophages to be CD14$^+$ [20]. Based on these findings we wanted to investigate if the macrophages within the human colon also could be identified as CD14$^+$. To do so we first selected all cells positive for CD45 (a pan leukocyte marker), which includes macrophages, but also other leukocytes such as dendritic cells and B and T cells. Consequently, a lineage dump (referred to as lin$^-$) was set up to filter out the B and T cells by the use of markers for CD19 (B cells) and CD3 (T cells). Macrophages also express HLA-DR, so from the filtered CD45$^+$ cells all cells positive for HLA-DR were selected (Figure 6). Since dendritic cells express both HLA-DR and CD45 as well, different markers have to be used to differentiate between them.

We first wanted to establish that CD14 confines all macrophages and distinguish them from dendritic cells (DCs), so we used antibodies for CD1c, CD141 and CD11c, as well as CD14, to stain cells from the mucosal layer of the colon and PBMCs. The PBMCs were used to set up the staining, as these cells are well characterized. By using this approach it is possible to depict the differences between macrophages and DCs, as well as distinguishing between mature and immature macrophages. Both CD1c and CD141 are markers for dendritic cells and by using them together, we were able to differentiate between DC1s (CD14$^-$CD1c$^-$CD141$^+$) and DC2s (CD14$^-$CD1c$^+$CD141$^-$) [9]. Since CD1c is also expressed on macrophages and monocytes, this marker cannot alone be used as a dendritic cell marker. The more mature macrophages will have downregulated CD11c as opposed to the more immature ones, which still express relatively high levels of CD11c [20]. In addition to CD1c, CD141 and CD11c, all cells were stained with antibodies for CD45, HLA-DR and CD14 according to Table 2 as these were shown to be expressed by macrophages in the small intestine [20]. To find the cells of interest (CD45$^+$HLA-DR$^+$lin$^-$), several gates were set as depicted in Figure 6.

From the viable CD45$^+$HLA-DR$^+$lin$^-$ cells ranging from CD14$^-$ to CD14$^+$ (Figure 6C), three new gates were set dividing the cells into CD11c$^-$CD14$^+$, CD11c$^+$CD14$^+$ and CD11c$^+$CD14$^-$. The CD11c$^-$CD14$^+$ cells correspond to more mature macrophages, while the CD11c$^+$CD14$^+$ were most likely immature macrophages. In the PBMCs, there were no CD11c$^-$CD14$^+$ cells present. Within the CD11c⁻CD14⁺ population in mucosa, the cells were either CD14$^{\text{intermediate}}$ or CD14$^{\text{high}}$ possibly similar to Bujko's findings of Mf3 and Mf4 macrophage subsets in the small intestine [20]. The CD11c$^+$CD14$^+$ cells were found to be HLA-DR$^{\text{intermediate}}$ or HLA-DR$^{\text{high}}$, comparable to the Mf1 and Mf2 macrophage subsets found by Bujko [20], although the Mf1 population contains quite few cells in our analysis. The majority of the CD14$^-$ cells were found to be either CD1c$^+$ or CD141$^+$, equivalent to DC1s or DC2s respectively (Figure 13).

FMOs for CD1c and CD141 were used as a negative control to set the gates (Figure 13). Here, cells were stained with all the antibodies in the mix except CD1c and CD141 in order to detect potential autofluorescence and capture the true fluorescence originating from antibody binding.

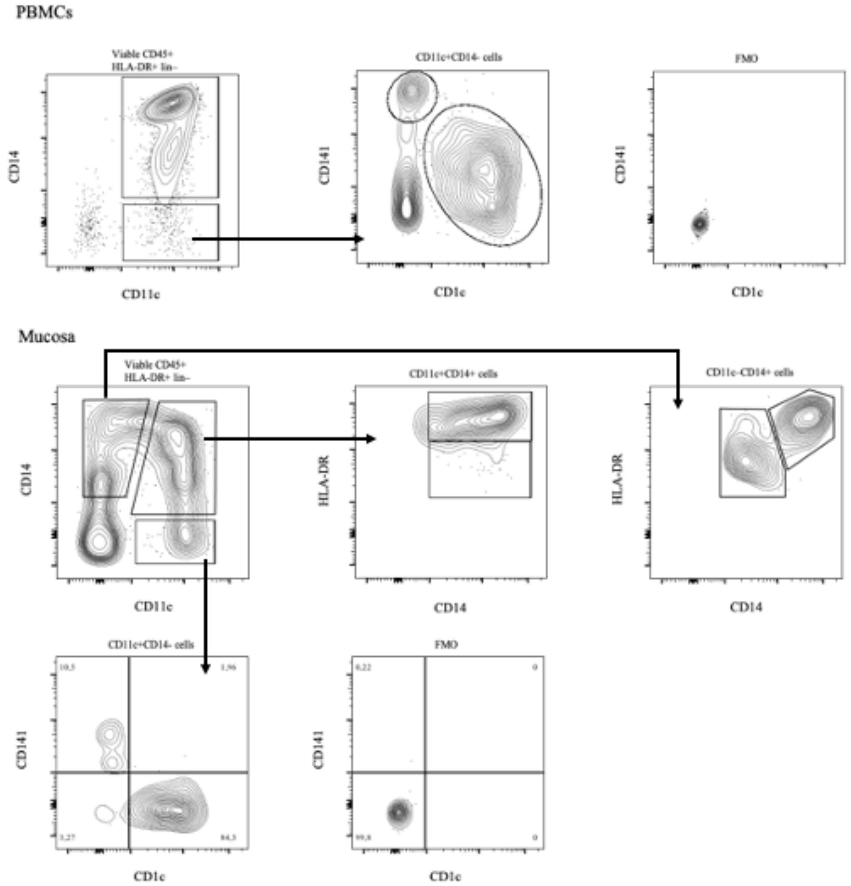

Figure 13. Representative flow cytometry data from the macrophages vs. dendritic cells assay.
Cells from the mucosal layer of the colon and PBMCs were stained with CD45 BV510, HLA-DR PerCP-Cy5.5, CD14 PE-Cy7, CD11c APC, CD1c BV421 and CD141 PE. BerEp4 FITC, CD3 FITC, CD19 Ax488 and TO-PRO-1 were used to exclude epithelial cells, T cells, B cells and dead cells, respectively, in addition to FcBlock which stops unspecific binding. This staining enables identification of macrophages (CD45$^+$HLA-DR$^+$lin$^-$CD14$^+$) and dendritic cells (CD45$^+$HLA-DR$^+$lin$^-$CD14$^-$). Here the cells are shown in a contour plot where outliers are indicated by dots. Prior to creating these plots, a general gating strategy, identifying viable CD45$^+$HLA-DR$^+$lin$^-$ cells, was performed (Figure 6).

From the CD45$^+$HLA-DR$^+$lin$^-$ cells, three separate subpopulations were identified when plotting CD11c (x-axis) against CD14 (y-axis). The subpopulations are divided into CD11c$^-$CD14$^+$, CD11c$^+$CD14$^+$ and CD11c$^+$CD14$^-$ cells. Within the CD11c$^+$CD14$^+$ subpopulation, the cells were found to be HLA-DR^{intermediate} or HLA-DR^{high}, similar to Bujko's findings of Mf1 and Mf2 macrophages, respectively [20]. Here, CD14 is on the x-axis and HLA-DR on the y-axis. The cells in the CD11c$^-$CD14$^+$ subpopulation were shown to generate two populations when plotting CD14 (x-axis) against HLA-DR (y-axis), namely CD14^{intermediate}HLA-DR$^+$ and CD14^{high}HLA-DR+. Within the CD11c$^+$CD14$^-$ subpopulation, the cells were identified as either DC1s (CD1c$^-$CD141$^+$) or DC2s (CD1c$^+$CD141$^-$), by plotting CD1c (x-axis) against CD141 (y-axis). FMOs for CD141 and CD1c were used in order to properly set the gates identifying dendritic cells.

3.2 Macrophage sorting for single-cell RNA sequencing

The findings of CD45$^+$HLA-DR$^+$lin$^-$ cells being either CD14$^-$ or CD14$^+$ and the CD14$^-$ cells comprising of CD141$^+$ DC1s and CD1c$^+$ DC2s, indicates that CD14 is a well suited marker for macrophage identification. Based on this, cell sorting was performed on mucosal cells from the colon. The cells were stained according to Table 4, and viable CD45$^+$HLA-DR$^+$CD14$^+$ cells were captured (Figure 14). The sorted macrophages were then used to generate single-cell RNA sequencing data, through the 10x platform, used to perform bioinformatic analyses. The raw sequencing data was preprocessed in Cell Ranger where cells with high expression of mitochondrial genes were filtered out, as high mitochondrial gene expression is a sign of dead cells, resulting in 5263 high quality cells used for further analyses.

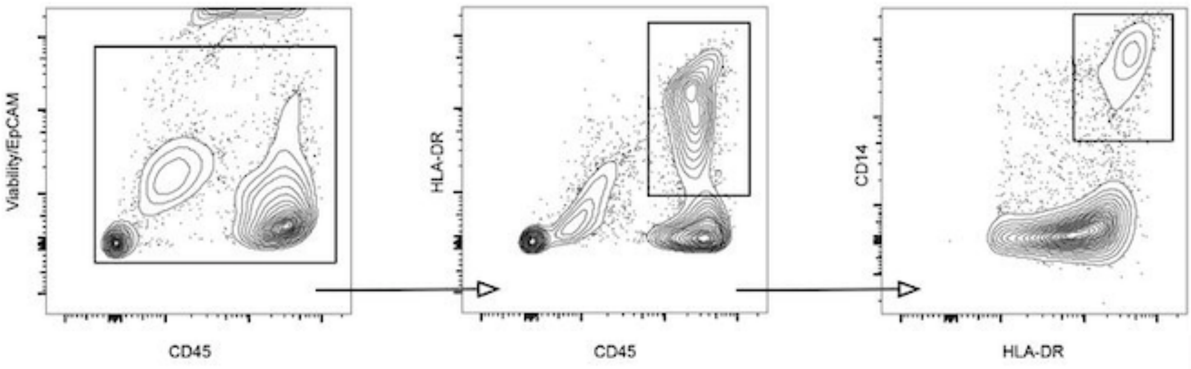

Figure 14. Single cell sorting plots.
Following the verification that macrophages are CD14$^+$, cells from the mucosal layer of the colon were stained with CD45 BV510, HLA-DR PerCP-Cy5.5, CD14 APC. BerEp4 FITC and TO-PRO-1 were used to exclude epithelial cells, T cells, B cells and dead cells, respectively, in addition to FcBlock which stops unspecific binding. CD45$^+$HLA-DR$^+$CD14$^+$ macrophages were sorted out for single cell RNA-sequencing.

First, viable CD45$^+$ cells are gated out, with CD45 on the x-axis and BerEp4, CD3, CD19 and TO-PRO-1 on the y-axis. From these cells, the HLA-DR$^+$ cells are selected, by plotting CD45 (x-axis) against HLA-DR (y-axis). Finally, the CD45$^+$HLA-DR$^+$CD14$^+$ cells are gated out, with HLA-DR on the x-axis and CD14 on the y-axis. (All gated cells are within the squares). Plots made by Espen Bækkevold.

3.3 Colonic macrophages separate into distinct clusters based on differently expressed genes

To visualize the single-cell RNA sequencing data generated from sorted macrophages, the R package Seurat was used. Graph-based clustering of the colonic mucosal layer from 4 patients was performed and visualized by the use of UMAP, a non-linear dimensional reduction method [64], based on differentially expressed genes (DEGs). Figure 15 shows the integrated UMAP, where the scRNA-seq data from all four patients were merged. The initial clustering identified some non-macrophage clusters in the UMAP. Cluster 3, 6 and 8 were identified as B cells, cluster 11 as T cells and cluster 13 contained endothelial cells (Figure 15A). This is based on the top differentially expressed genes (DEGs) in each cluster – genes such as JCHAIN and IGHA1 were found in the B cell clusters, CD3 and CCL5 in the T cell cluster, and SPARCL1 and PLVAP in the cluster with endothelial cells. To obtain a clear image of the differences within macrophages, these non-macrophage clusters were removed and the data re-clustered, before performing any further trajectory analysis (Figure 15B). The re-clustering is important because the clustering is based on the relationship between all cells in the data. It is therefore necessary to redo the clustering after the removal of non-macrophage cells. The differentially expressed genes gives the foundation for clustering the cells. In Figure 16, the top 10 DEGs for each cluster, after removing the non-macrophage clusters, are depicted in a heatmap. Based on the DEGs, each cluster was annotated, identifying different macrophage subtypes.

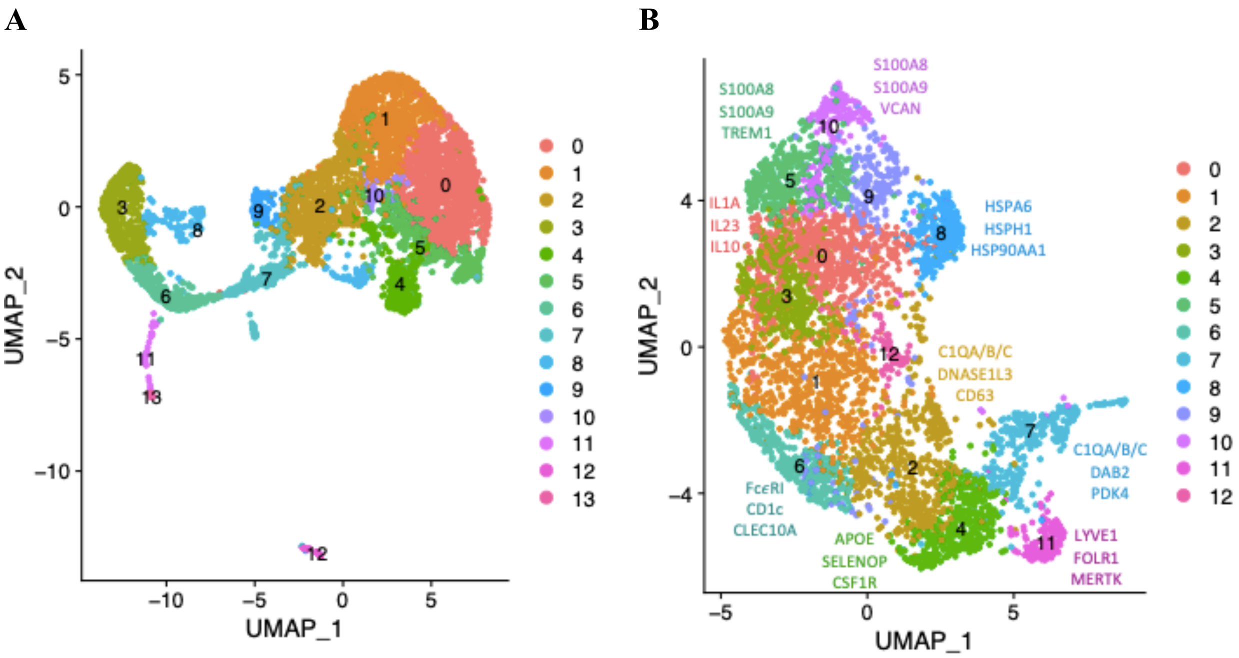

Figure 15. UMAP plots of cells from the mucosal layer of the colon, where the clustering is based on differentially expressed genes.

A The graph-based clustering of 5263 cells from the colonic mucosal layer, where 14 different clusters were identified. Before non-macrophage clusters are removed. **B** The graph-based clustering of 4251 cells from the mucosal layer, after non-macrophage clusters are removed, where 13 different clusters were identified. Some of the highly expressed genes in the clusters are also shown.

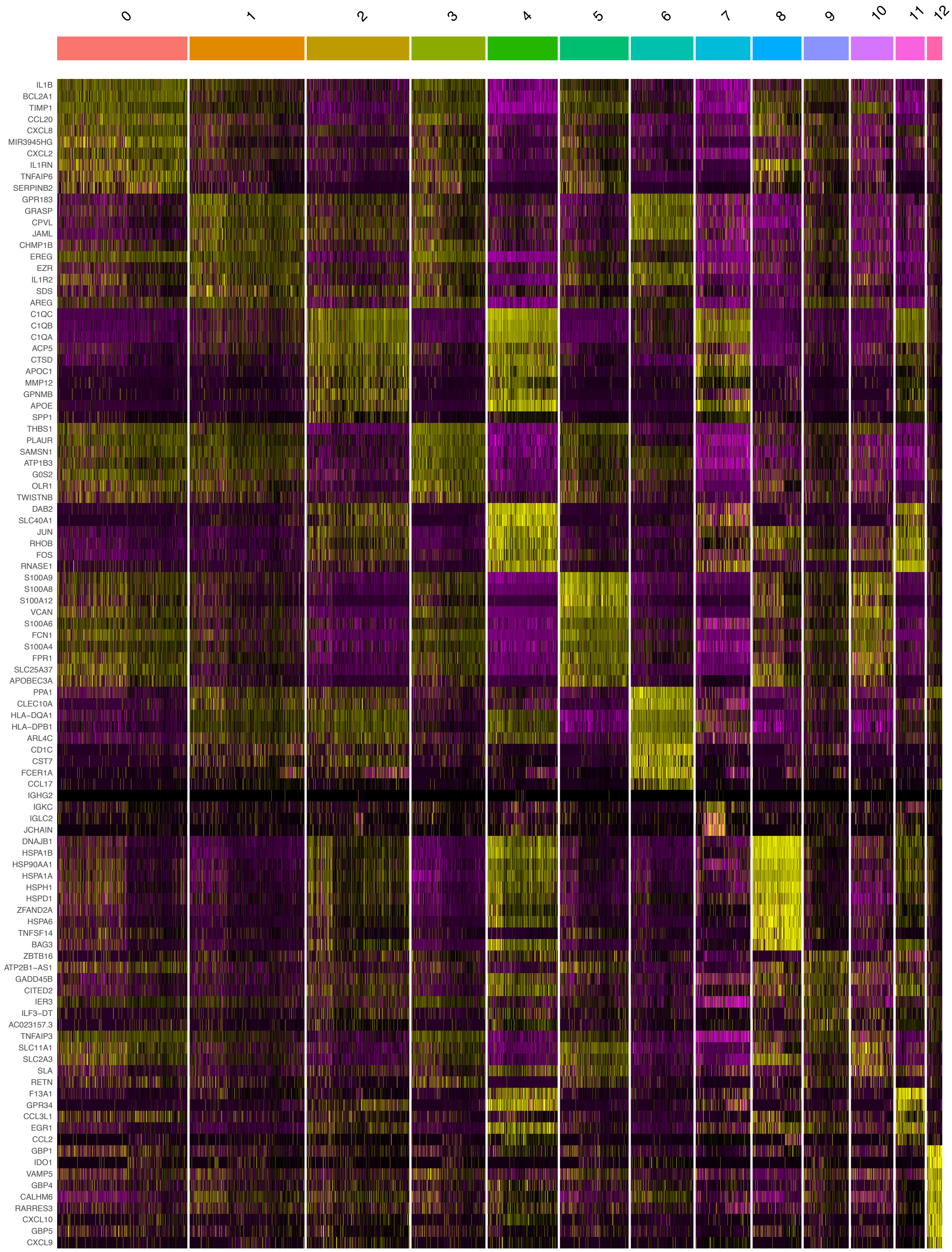

Figure 16. Heatmap of the top 10 differentially expressed genes in each cluster.
Here, the top 10 differentially expressed genes in each cluster are shown with the gene names on the left. Highly expressed genes are colored in yellow.

3.3.1 Cluster annotation

Immature macrophages are characterized by expression of certain markers, such as S100A8 and S100A9. Based on these markers, clusters 5 and 10 from the mucosal layer were identified as immature macrophages (Figure 17). FCN1, VCAN, AQP9 and TREM1 are genes associated with S100A8/S100A9$^+$ immature macrophages. When looking at the gene expression, it was evident that FCN1 and TREM1 were highly expressed in both the immature clusters, while VCAN and AQP9 were present in cluster 10 and cluster 5, respectively.

The mature macrophage clusters were identified by high expression of the different C1Q genes (i.e. C1QA, C1QB and C1QC), discovered in cluster 2, 4, 7 and 11. Within these clusters, other genes associated with mature macrophages, such as MRC1, APOE, SELENOP, CSF1R, MERTK and LYVE1, were also present. APOE was found to be highly expressed in cluster 4, while MRC1 and LYVE1 were highly expressed in cluster 11. SELENOP, CSF1R and MERTK were all highly expressed in both cluster 4 and 11. Tissue-resident macrophages (TRMs) are associated with genes such as LYVE1, COLEC12, F13A1 and FOLR2, having the highest expression in Cluster 11, indicating that this cluster contains tissue-resident macrophages (Figure 17).

The newly described macrophage-related DC subset, named DC3s (see below) [83], were identified in cluster 6, based on the expression of CD1c and FcϵRI (Figure 17). The cells within this cluster also expressed some genes associated with dendritic cells, such as CLEC10A and CD1E, in addition to expression of CD14.

IL-1A, IL-1B, IL6, IL23A, CXCL2, CXCL3 and CXCL8 are genes identified as pro-inflammatory found highly expressed in cluster 0, suggesting that this cluster contains pro-inflammatory macrophages (Figure 17). IL-10 is an immunoregulatory gene also found highly expressed in cluster 0.

CD63, DNASE1L3 and ADAMDEC1 are genes associated with embryonic origin [84], found to be expressed in cluster 2 (CD63 and DNASE1L3), 4, 7 (DNASE1L3 and ADAMDEC1) and cluster 11 and 12 (CD63).

These findings suggests that the immature macrophages are located in cluster 5 and 10, while the mature macrophages are situated in cluster 2, 4 and 7. The cells of cluster 11 was identified as TRMs, consistent with high expression of the C1Q markers and TRM genes. The pro-inflammatory macrophages were found to be located in cluster 0, while the DC3 subset were located in cluster 6.

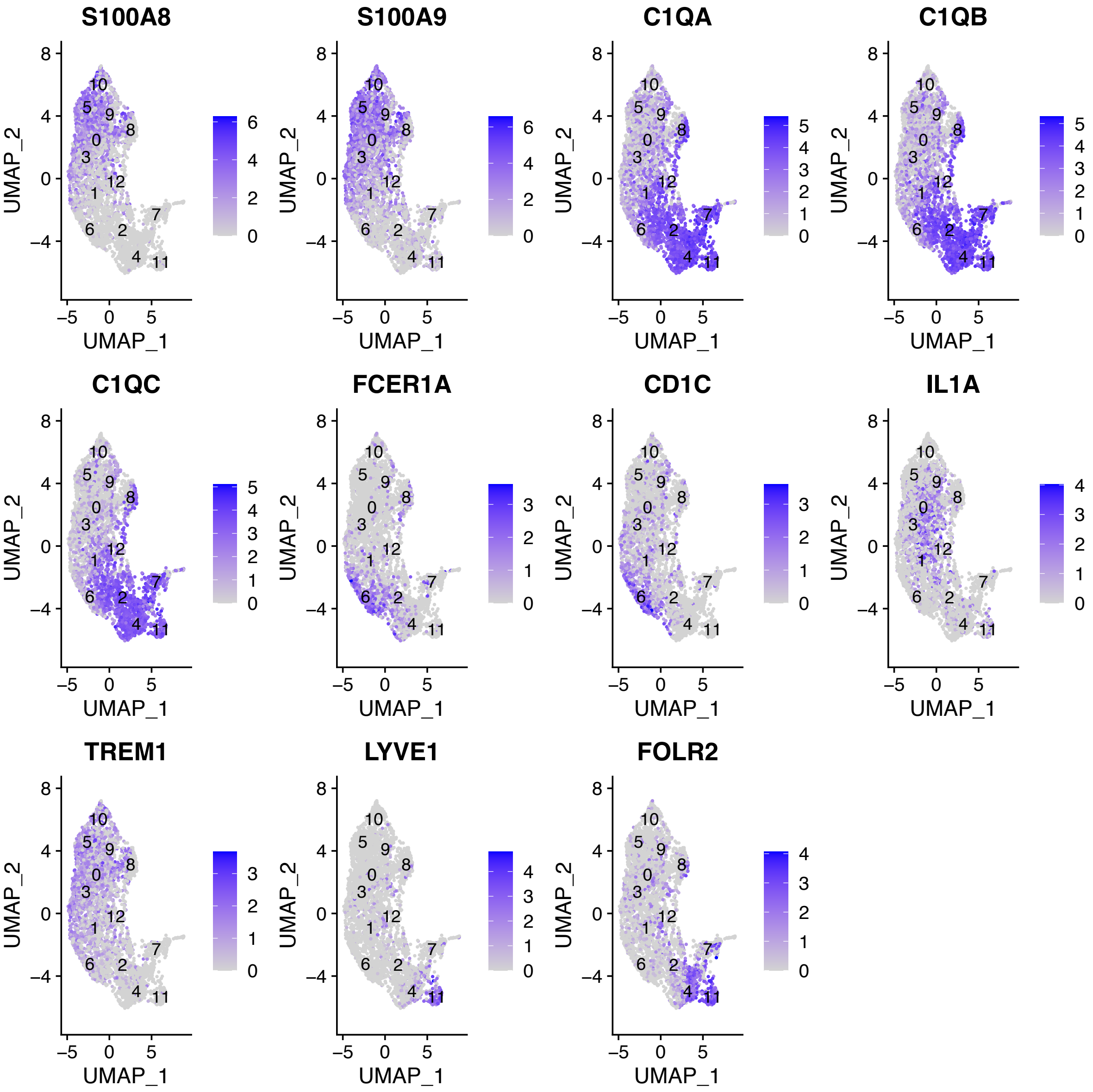

Figure 17. Feature plot of selected genes.
A feature plot showing the expression of selected genes and their distribution within the UMAP plot. The expression is shown as a gradient going from no expression (gray) to high expression (blue).

Here, the immature macrophage markers S100A8 and S100A9 are highly expressed in the clusters at the top of the UMAP plot, while a subset of these immature macrophages express TREM1. The mature macrophage markers C1QA/B/C are expressed in the clusters located at the bottom right in the UMAP plot. These cells also express LYVE1 and FOLR2. The DC3 markers FcϵRI and CD1c are highly expressed in cluster 6. The pro-inflammatory genes IL-1A is expressed mainly in cluster 0.

3.4 Trajectories

In order to investigate the trajectory inference in the clustered single-cell RNA sequencing data, Monocle 3 was used. Here, the trajectory root (i.e. starting point of the trajectories) was set based on the differentially expressed genes (DEGs) and clusters identified above. The DC3 cluster was deliberately removed, since studies have shown that this subset has a different origin, probably not classical monocytes [83].

Macrophages develop from recently elicited monocytes, through immature intermediates, and it is therefore reasonable to set the trajectory root in the immature macrophage cluster. Both cluster 5 and 10 were identified as clusters of immature macrophages, and since cluster 10 is situated at the top in the UMAP plot, this cluster was chosen as the trajectory root. Based on the assumption that the macrophages go from immature to mature, we would expect the end-point (in yellow) to be in the clusters located in the bottom right corner of the plot. The trajectory analysis showed a dynamic development going from blue at the root to yellow at the end-point, which was located in cluster 2, 4, 7 and 11 (Figure 18). There are several branches present in this trajectory plot. One branch has its endpoint in cluster 8 where we find multiple heat shock proteins, suggesting that these macrophages may be stressed. At the end-point (in yellow), the trajectory divides into two separate branches in cluster 7 and 11. Within cluster 7 we also find genes associated with B cells, such as IGKC and JCHAIN, indicating that these genes may be cargo-genes (i.e. genes expressed by phagocytosed cells within the macrophages). As mentioned above, the macrophages in cluster 11 express genes related to TRMs. This is most likely why there is two different end-points identified – both clusters contains mature macrophages, but with different additional gene expression. Another branch ends in cluster 3 where we find some pro-inflammatory genes such as IL-1A, IL-1B and IL-23.

Specific genes may also be plotted as a function of pseudotime, as shown in Figure 19. Here, we see that the immature macrophage markers S100A8 and S100A9 begin with a high expression, which decreases along the pseudotime. The mature macrophage markers (i.e. C1QA/B/C, FLOR2 and LYVE1), on the other hand, have an increased expression throughout pseudotime. This corresponds to the beforementioned assumption of differentiation from immature to mature macrophages. Moreover, IL-1 appears to be upregulated following early maturation, but is downregulated in fully mature cells.

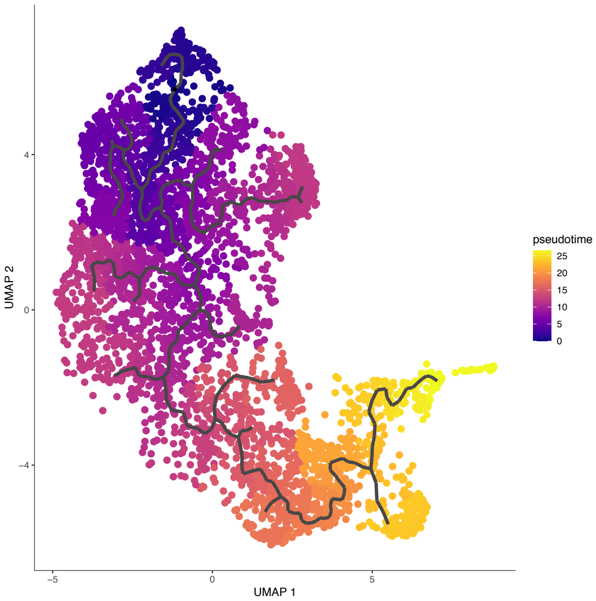

Figure 18. Trajectory plot of cells from the mucosal layer of the colon .
A trajectory plot showing the developmental pathway of the cells. Here, the plot is colored by pseudotime, meaning that the cells differentiate from blue to yellow. The root was set in cluster 10.

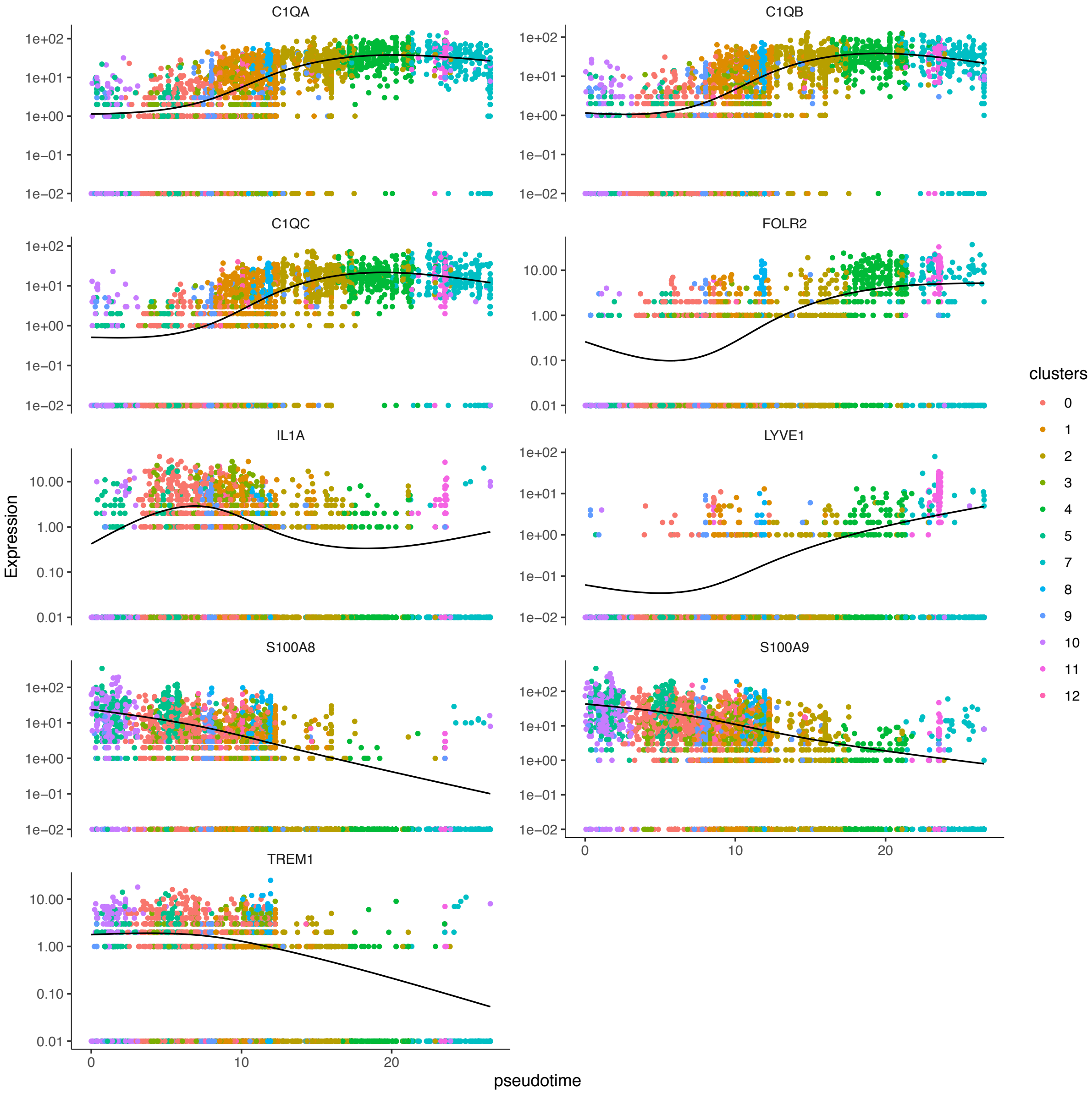

Figure 19. Selected genes plotted as a function of Pseudotime.

The dynamic expression of selected genes are shown as a function of pseudotime. Here, S100A8/9, TREM1and IL-1A are highly expressed in early pseudotime, while C1Q/A/B/C, FOLR2 and LYVE1 have a low expression. As the S100A8/9, TREM1and IL-1A expression decreases through pseudotime, the C1Q/A/B/C, FOLR2 and LYVE1 expression increases.

3.5 Identification of mature and immature macrophages in colonic tissue

Further, we wanted to validate the scRNA-seq data, by investigating the expression of specific markers directly in tissue-derived cells. This is important as there may be discrepancies between RNA expression and protein expression.

Here, we looked into the distribution of mature and immature macrophages by examining the expression of C1Q and calprotectin (markers for mature and immature macrophages, respectively). This is interesting because it depicts the dynamics within this cell population; a high amount of immature macrophages indicates that monocytes are coming in to the tissue where they will differentiate into macrophages, while the presence of many mature macrophages suggest that they have stayed there for some time, and might self-renew within the tissue.

C1Q is a marker for mature macrophages that we are able to stain for when performing flow cytometry. PBMCs and the mucosal layer of the colon were stained with the C1Q antibody in addition to an antibody for calprotectin. Calprotectin is a marker for the heterodimer of the S100A8 and S100A9 genes, which are expressed in immature macrophages. Also here, all cells were stained for CD45, HLA-DR and CD14 (Table 3) The cells shown in Figure 20 originated from the gating strategy described in Figure 6 . Mouse isotypes for both C1Q and Calprotectin were used as FMOs in order to properly set the gates, as described above (3.1, Figure 20B).

The majority of the macrophages are mature ($C1Q^{+}$) in the mucosal layer of the colon, while in the PBMCs there are almost solely monocytes present ($calprotectin^{+}$) as expected (Figure 20A). Immature macrophages are also present in the mucosal layer, but to a lower degree.

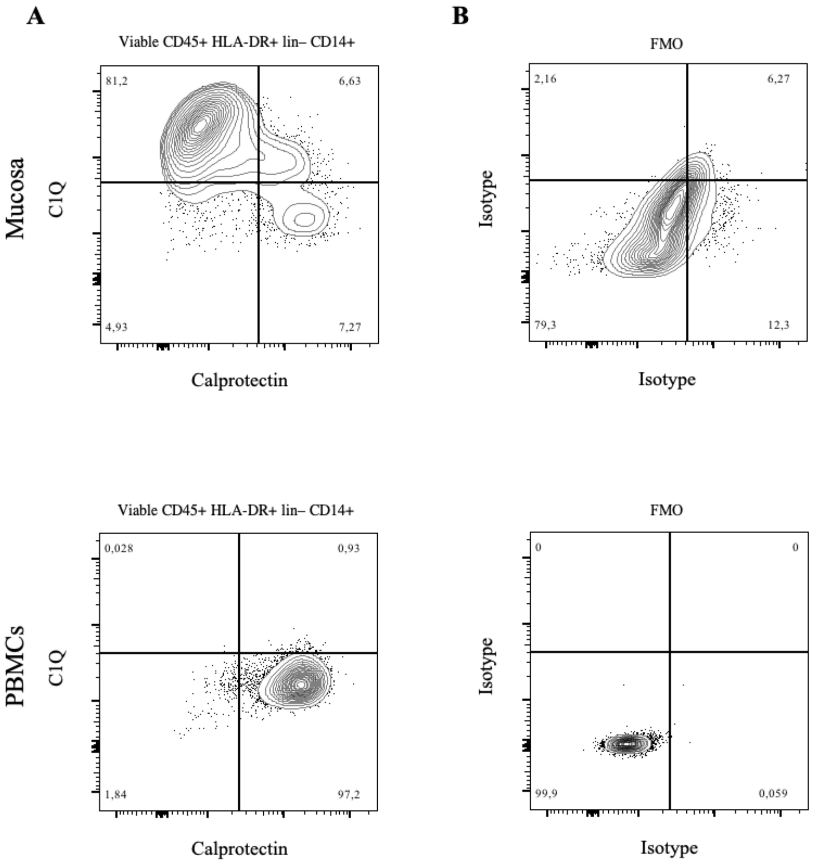

Figure 20. Representative flow cytometry data from the mature/immature macrophage assay.
Cells from the mucosal layer of the colon and PBMCs were stained with CD45 BV510, HLA-DR PerCP-Cy5.5, CD14 APC, C1Q FITC and Calprotectin PE. Fixable Viability dye eFluor 780 was used to exclude dead cells, in addition to FcBlock which stops unspecific binding. This staining enables identification of mature (CD45$^+$HLA-DR$^+$lin$^-$CD14$^+$C1Q$^+$Calprotectin$^-$) and immature macrophages (CD45$^+$HLA-DR$^+$ lin$^-$CD14$^+$C1Q$^-$Calprotectin$^+$). Here the cells are shown in a contour plot where outliers are indicated by dots. Prior to creating these plots, a general gating strategy, identifying viable CD45$^+$HLA-DR$^+$lin$^-$ cells, was performed (Figure 6).

A When plotting Calprotectin (x-axis) against C1Q (y-axis), two distinct populations within the viable CD45$^+$HLA-DR$^+$ cells are identified, being either C1Q$^+$ Calprotectin$^-$ (upper left quadrant) or C1Q$^-$Calprotectin$^+$ (lower right quadrant). **B** Mouse isotypes for C1Q and Calprotectin were used to properly set the gates in **A**.

3.5.1 DC3s – A newly described macrophage-related DC subset

Bourdely *et al.* have recently described a new subset of cells, termed DC3s. These cells have a transcriptional profile where some features are also found in the transcription profile of monocytes and DC2s, but they are thought to originate independently of classical monocytes and dendritic cells [83].

Markers for the DC3 subset are CD1c and FcϵRI, in addition to monocytic markers such as CD14, CD68, S100A8, and S100A9 [83]. Colon cells from the mucosal layer and PBMCs were stained with CD1c and FcϵRI in addition to other surface marker antibodies, described in Table 2 to detect DC3s. The flow cytometry analysis revealed a DC3 population which are CD1c$^+$ FcϵRI$^+$ within the CD45$^+$HLA-DR$^+$CD14$^+$ cells (Figure 21).

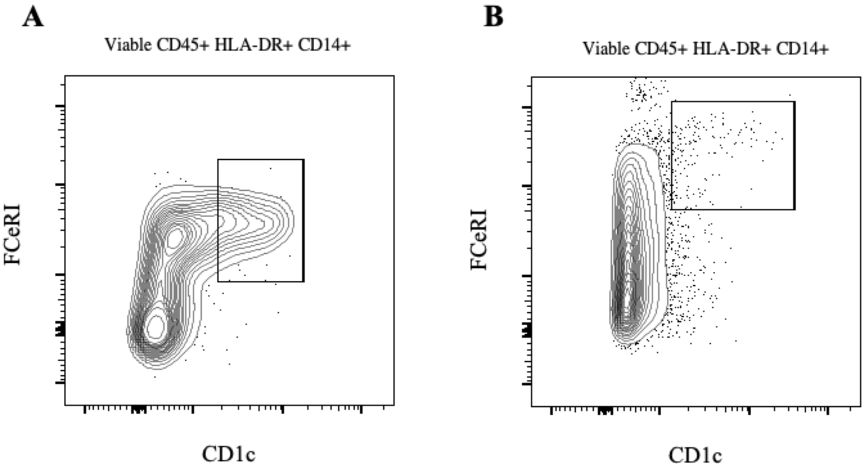

Figure 21. Representative flow cytometry data from the DC3 assay.
Cells from the mucosal layer of the colon and PBMCs were stained with CD45 BV510, HLA-DR PerCP-Cy5.5, CD14 APC, CD11c APC, FcϵRI PE and CD1c BV421. BerEp4 FITC, CD3 FITC, CD19 Ax488 and TO-PRO-1 were used to exclude epithelial cells, T cells, B cells and dead cells, respectively, in addition to FcBlock which stops unspecific binding. This staining enables identification of a DC3 subset (FCeRI$^+$CD1c$^+$). Here the cells are shown in a contour plot where outliers are indicated by dots. Prior to creating these plots, a general gating strategy, identifying viable CD45$^+$HLA-DR$^+$lin$^-$CD14$^+$ cells, was performed (Figure 6).

From the CD45$^+$HLA-DR$^+$CD14$^+$ cells, an FCeRI$^+$CD1c$^+$ subpopulation was identified and gated for (within the square) when plotting CD1c (x-axis) against FcϵRI (y-axis). This population, termed DC3s, was found both in the colonic mucosal layer (**A**) and in PBMCs (**B**).

3.6 Localization of colonic macrophages *in situ*

To acquire a deeper understanding of where the different subsets of macrophages (e.g. mature and immature) are located within the tissue, immunohistochemistry was performed. Colon slides were stained with antibodies for CD68 (a pan-macrophage marker) in combination with C1Q or calprotectin to localize mature and immature macrophages. Macrophages expressing TREM1 were also localized, as this gene was found to be highly expressed in a subset of immature macrophages. To identify DC3s, colon slides were stained for CD14 and CD1c.

For the immunofluorescent pictures, CD68 or CD14 is shown in green, C1Q, calprotectin, or CD1c in red, and a nuclear staining (Hoechst) in blue. The yellow coloring corresponds to an overlap between CD68 and C1Q or calprotectin, or CD14 and CD1c. A negative control, with the secondary antibodies alone, was used to verify that the true fluorescence is captured. The exposure time for each fluorochrome was set so that the negative control was barely visible (Figure 22). In the enzymatic staining, TREM1 is shown in turquoise, CD14 in red and the nuclear staining a weak blue.

From the *in situ* staining it is evident that the mature C1Q$^+$ macrophages localize to the upper part of the mucosa, underneath the epithelial layer facing the colonic lumen (Figure 23). The immature macrophages, on the other hand, are found in the lower part of the mucosa, closer to the bottom of the epithelial crypts (Figure 24).

As TREM1 was found to be highly expressed in the immature macrophage clusters, we assessed the *in situ* localization of these macrophages by enzymatic staining for CD14 and TREM1. The TREM1$^+$ macrophages were found scattered within the lamina propria between the crypts (Figure 25).

We also wanted to localize the DC3 population *in situ*. This was done by staining for CD1c and CD14. The CD1c-antibody available is not compatible with formalin-fixed tissue, hance cryopreserved sections were used for this analysis. Cells co-expressing these markers are likely to be DC3s, and were mostly found scattered in the lamina propria in close association with the basolateral face of the colonic epithelium (Figure 26).

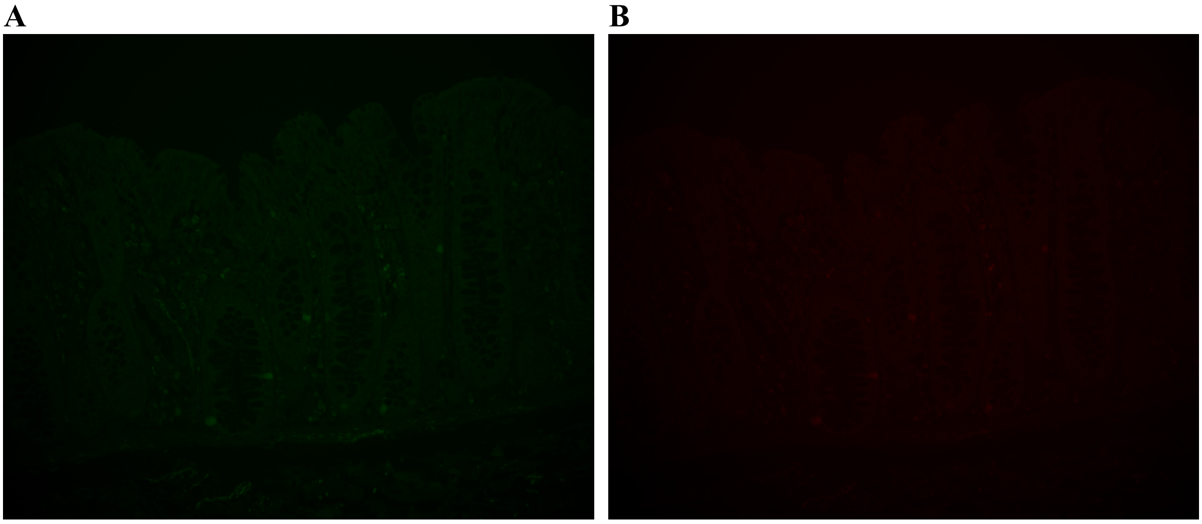

Figure 22. Representative image of *in situ* staining of colon, depicting a negative control.
The A sample of the stained colon slides were used as negative controls for the primary antibodies. From these controls, the exposure time for each fluorochrome were set to filter out any autofluorescence. Here, a negative control for Alexa fluor 488 in green (**A**) and Cy3 in red (**B**) are depicted.

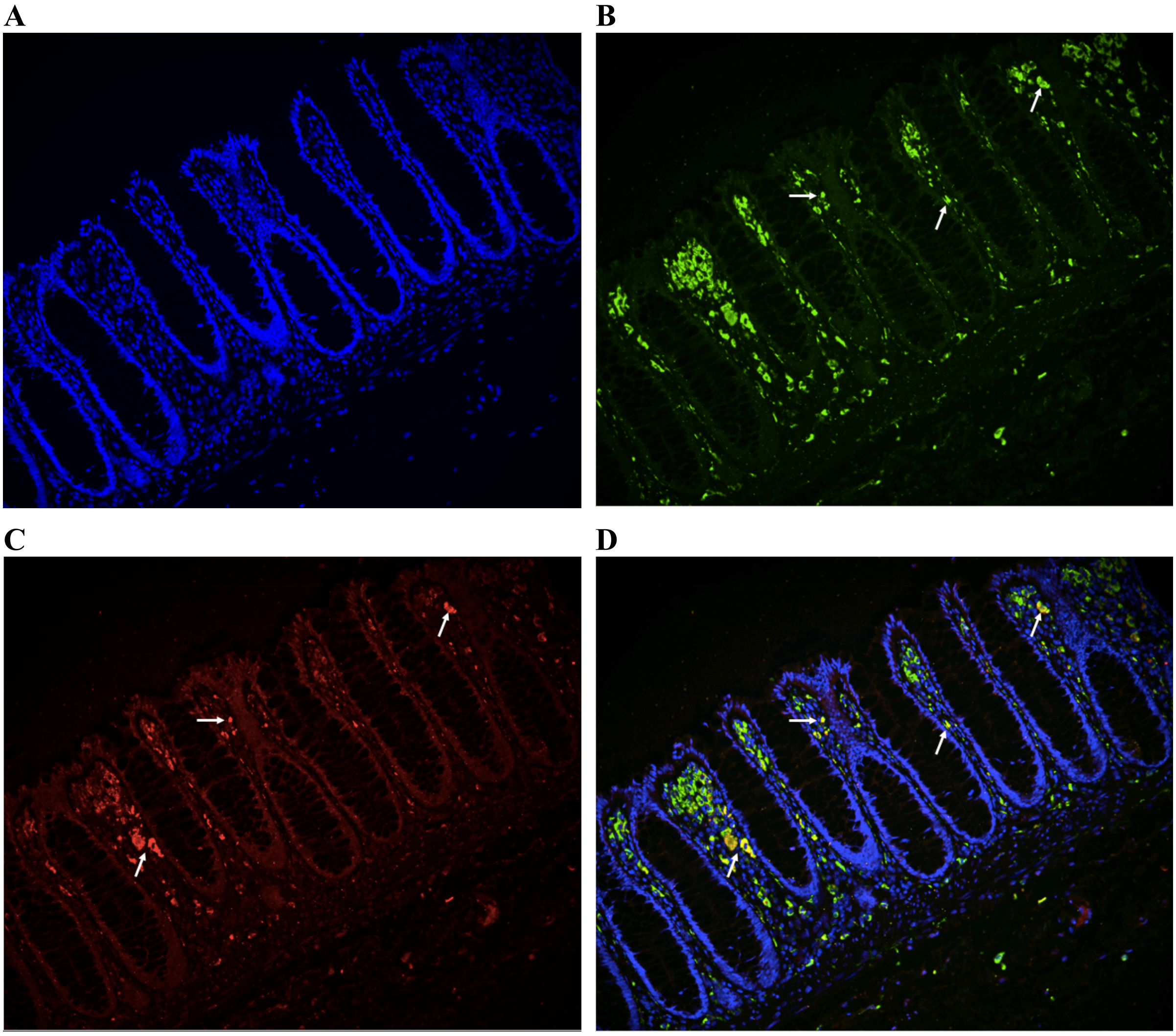

Figure 23. Representative image of immunofluorescent staining of colon with CD68 and C1Q. Staining of formalin fixated colon slides with fluorescently conjugated antibodies for CD68 and C1Q. In total, colon samples from three different patients were stained (n = 3).

A Only nuclear stain (DAPI) depicting the colonic morphology. **B** Only Alexa fluor 488 showing CD68^{+} cells, which are macrophages. **C** Only Cy3 staining C1Q^{+} cells. **D** All three colors combined (i.e. nuclear stain, CD68 and C1Q). The mature macrophages (CD68^{+}C1Q^{+}) are shown in yellow, arising from an overlap between CD68 in green and C1Q in red. The arrows point to the double-positive cells.

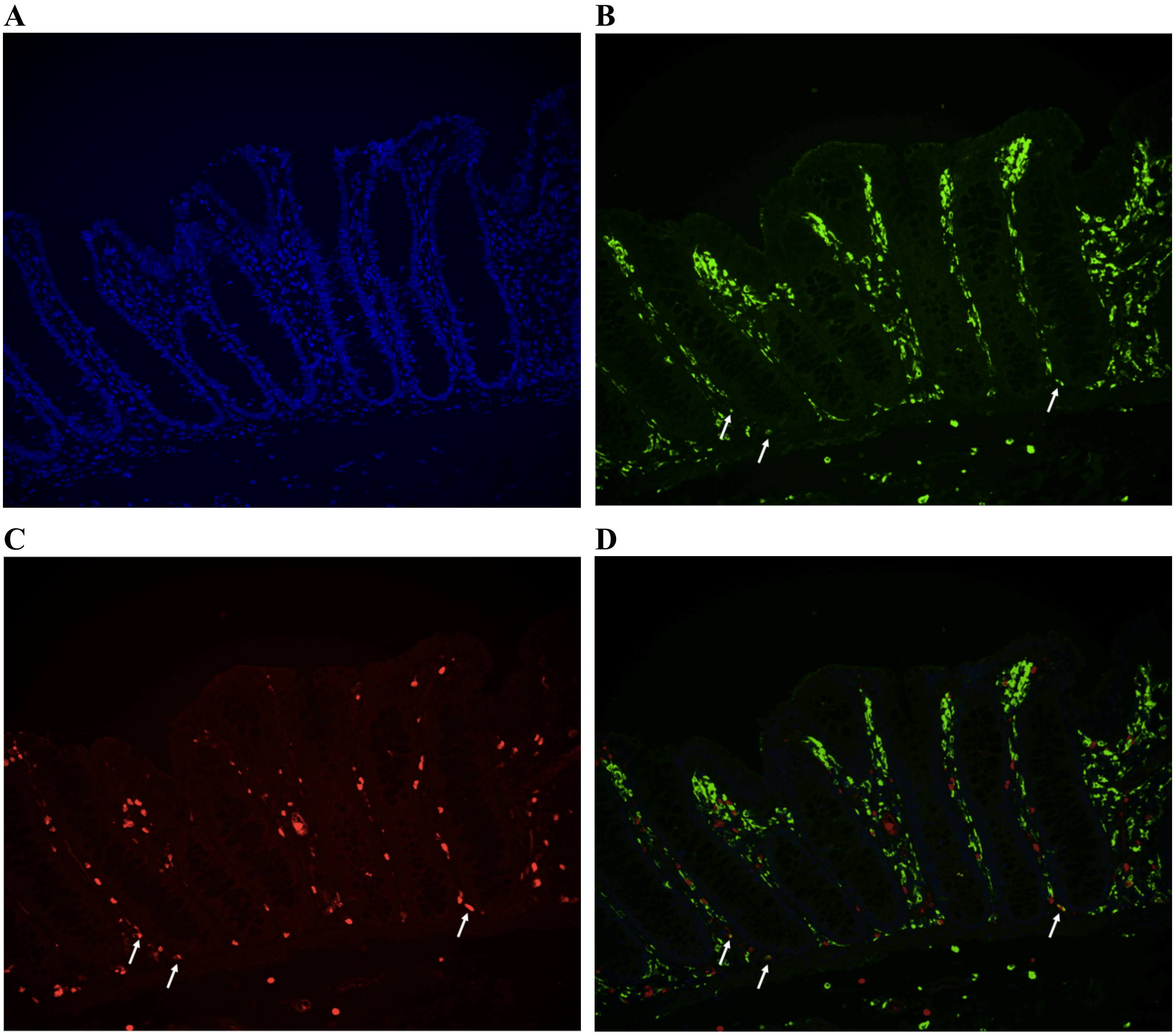

Figure 24. Representative image of immunofluorescent staining of colon with CD68 and calprotectin. Staining of formalin fixated colon slides with fluorescently conjugated antibodies for CD68 and calprotectin. In total, colon samples from three different patients were stained (n = 3).

A Only nuclear stain (DAPI) depicting the colonic morphology. **B** Only Alexa fluor 488 showing CD68$^+$ cells, which are macrophages. **C** Only Cy3 staining calprotectin$^+$ cells. **D** All three colors combined (i.e. nuclear stain, CD68 and calprotectin). The immature macrophages (CD68$^+$calprotectin$^+$) are shown in yellow, arising from an overlap between CD68 in green and calprotectin in red. The arrows point to the double-positive cells.

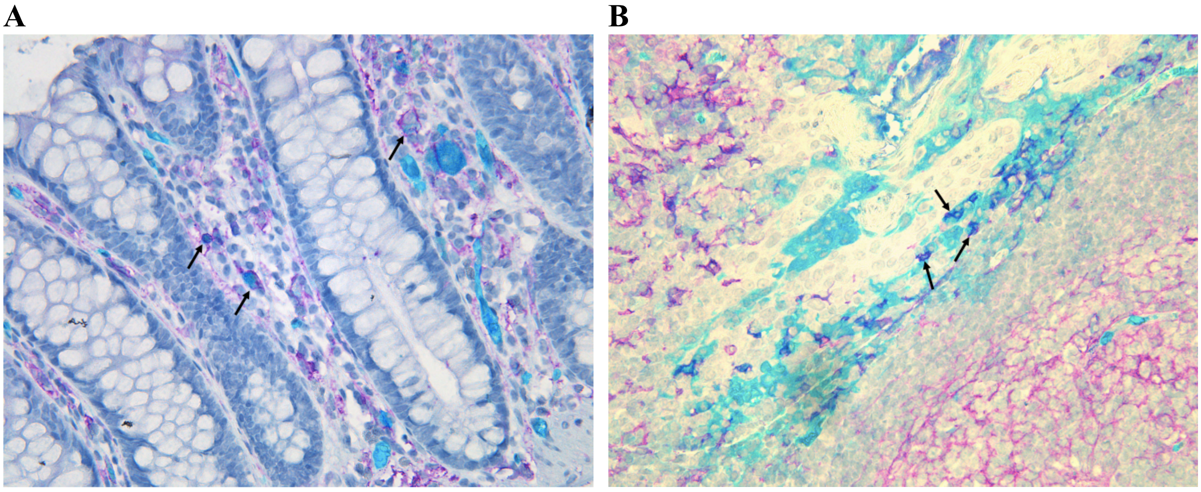

Figure 25. Representative image of enzyme staining of colon and tonsil with CD14 and TREM1. Enzyme staining of formalin fixated colon and tonsil slides with antibodies for CD14 and TREM1. Here, CD14 is shown in red and TREM1 in turquoise, while a nuclear staining gives the weak blue coloring. Double-positive (CD14^{+}TREM1^{+}) cells will appear with a darker blue/purple color. In total, colon samples from three different patients were stained (n = 3).

A Colon slide where CD14^{+}TREM1^{+} cells are indicated with arrows. **B** Doble-positive cells shown in tonsil, as reference.

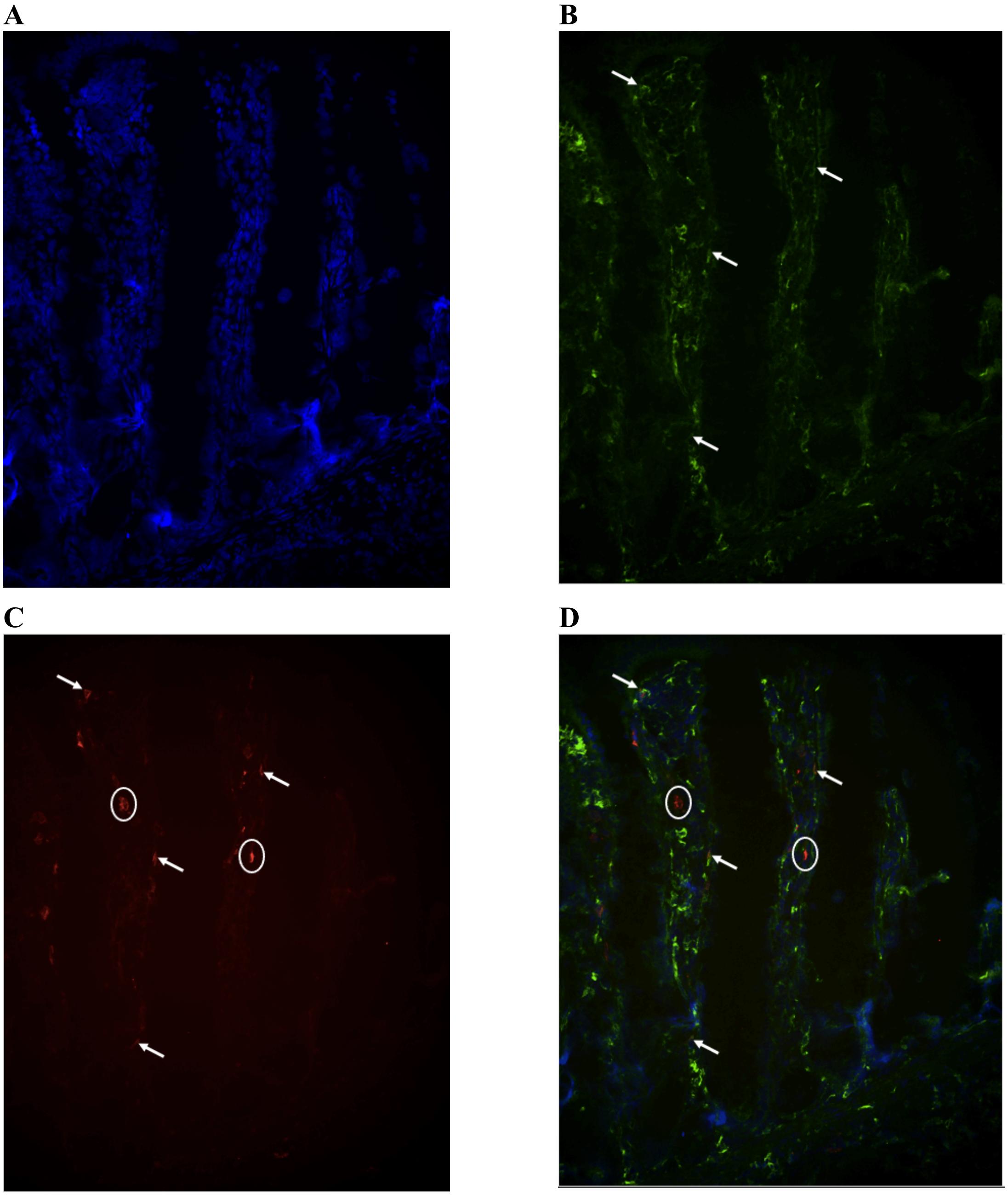

Figure 26. Representative image of immunofluorescent staining of colon with CD14 and CD1c. Staining of cryopreserved colon slides with fluorescently conjugated antibodies for CD14 and CD1c. In total, colon samples from three different patients were stained (n = 3).

A Only nuclear stain (DAPI) depicting the colonic mucosal morphology. **B** Only Alexa fluor 488 showing $CD14^+$ cells. **C** Only Cy3 staining $CD1c^+$ cells. **D** All three colors combined (i.e. nuclear stain, CD14 and CD1c). The double-positive cells ($CD14^+CD1c^+$) are shown in yellow, arising from an overlap between CD14 in green and CD1c in red, indicating cells of the DC3 subset. The arrows point to the double-positive cells. Within the circles there are $CD14^-CD1c^+$ cells, indicating DC2s.

4 Discussion

Macrophages are a heterogeneous population of immune cells scattered throughout the body. In the tissue, they play an important role when it comes to tissue maintenance and clearance of dead cells, as well as a bridge between the innate and adaptive immune system. Previous studies have investigated the macrophage landscape in the small intestine [20], but macrophages in the large intestine are not fully characterized. The aim of this thesis was to investigate the macrophage heterogeneity within the human colon.

In this study, we found CD14 to be a well suited marker for colonic macrophages. Through scRNAseq, flow cytometric and immunohistochemical analyses, several distinct subpopulations of macrophages were identified within the colonic mucosa, such as immature macrophages located at the lower part and mature macrophages found higher up in the mucosa underneath the epithelial layer. Pro-inflammatory subsets were also uncovered, as well as putative tissue-resident macrophages and the newly described DC3 population.

The flow cytometry analysis of isolated colonic mucosal cells showed a relatively large population of CD45^{+}HLA-DR^{+}CD14^{+}CD11c^{+} cells (Figure 13) among lamina propria leukocytes. These macrophages could be further divided into two macrophage subset, with high or intermediate expression of HLA-DR, similar to the findings of Bujko *et al.* (Figure 13) [20]. A majority of the macrophages were found to be of the Mf2 subset identified as CD14^{+}CD11c^{+}HLA-DR^{high}. We also found an Mf1 population, though smaller than the Mf2 population, in the colonic mucosa classified as CD14^{+}CD11c^{+}HLA-DR^{intermediate}. The macrophages in both these subpopulations have not yet downregulated their CD11c expression, suggesting that they are at an early stage of macrophage differentiation from monocytes. We looked into the monocyte population in blood as well, by isolation of PBMCs (peripheral blood mononuclear cells). Also here, we found the largest population of monocytes to be of the CD14^{+}CD11c^{+}HLA-DR^{high} subset, with a smaller population of the CD14^{+}CD11c^{+}HLA-DR^{intermediate} monocytes. Furthermore, a CD45^{+}HLA-DR^{+}CD14^{+}CD11c^{-} population of macrophages was identified (Figure 13). Their low levels of CD11c indicates that these macrophages have resided in the tissue for some time, suggesting that this population might consist of tissue-resident intestinal macrophages. Within the CD45^{+}HLA-DR^{+}CD14^{+}CD11c^{-} population a CD14^{intermediate} and a CD14^{high} macrophage population were uncovered, comparable to Mf3 and Mf4 found in the small intestine [20]. As expected, there were no

CD45$^+$HLA-DR$^+$CD14$^+$CD11c$^-$ monocytes detected in the PBMCs. This is in accordance with our hypothesis of circulating monocytes entering the tissue before becoming mature macrophages.

The identification of dendritic cells (DCs), classified as CD45$^+$HLA-DR$^+$CD14$^-$CD11c$^+$, within the CD14$^-$ cells (Figure 13) and the findings of different macrophage subsets within the CD14$^+$ cells, further supports the findings of Bujko *et al.* that CD14 is a well suited marker for macrophage identification, also in the colon.

Flow cytometry staining for immature and mature macrophages within the colonic mucosa was also performed to further investigate the macrophage heterogeneity. Here, we found the majority of the macrophages to express the mature C1Q marker. These results combined with the findings of a large macrophage population having a high CD11c expression, might indicate a rapid maturation process for the colonic macrophages compared to the small intestine. In fact, whereas Bujko *et al.* [20] found by analyzing the temporal macrophage turnover in transplanted small intestinal grafts that macrophages displayed a tissue half-life of approximately 6 months, our data indicate that the colonic macrophage pool may be replaced significantly faster. We also found a small population of cells expressing both immature and mature markers, suggesting that these macrophages are in an intermediate stage in their differentiation to mature macrophages. The population of immature macrophages identified by this staining are likely monocytes newly recruited to the tissue. PBMCs were stained for the same immature and mature markers, enabling a comparison between the blood and the colonic mucosa. Almost all of the cells within the PBMCs were only positive for the immature marker (S100A8/9$^+$). This further supports the hypothesis that monocytes circulate in the blood before extravasation into the tissue, where they differentiate into macrophages.

Macrophage heterogeneity have been shown in other tissues, such as lungs and liver, where several different subsets were identified [85–88]. This observed heterogeneity is likely due to the local microenvironment within the different parts of the tissue, where distinct environmental cues result in niche specific macrophage subsets [89]. These specific functions have been shown in tissues such as the heart, brain and lungs [90–92].

The analysis of the 10x genomics sequencing data revealed considerable heterogeneity within the sorted cells. From the clustering, we found additional subsets of macrophages, based on gene expression. Both immature and mature macrophages were present in the colonic mucosa, suggesting that there are monocytes coming in to the tissue were they differentiate to macrophages, which stay in the tissue over time as they mature. This was further supported by the identification of genes associated with tissue-resident macrophages, such as LYVE1 and COLEC12, within the mature macrophage clusters.

A subset of pro-inflammatory macrophages was also found based on different cytokines and chemokines. These pro-inflammatory macrophages may initiate an adaptive immune response by killing pathogens and presenting them on the HLA receptor (the human equivalent to MHC-II) to T cells [27], as well as clearance of apoptotic cells. Both of which are important functions in the colonic mucosa. The cytokine IL-10 was also found to be expressed in the same cluster as the pro-inflammatory genes. This might indicate the presence of a subset regulating both inflammatory and immune reaction by production of IL-10, IL-1β and IL-6 (among others) [27]. The identification of cargo-RNA within the macrophages in cluster 7 indicates ingestion of pathogens and a high phagocytic capacity, important for intestinal macrophages as they are involved in tissue homeostasis by eating dead and senescent cells.

The potential of bone marrow-derived monocytic precursors giving rise to mature intestinal macrophages was investigated in this study by computational analyses. Since the DC3s derive independently of the classical monocytic lineage, these cells were removed before performing the trajectory analysis. Our hypothesis regarding the dynamics within the macrophages, was that they differentiate from the immature S100A8/9^{+} cells to the C1QA/B/C^{+} cells. The trajectory analysis supported this assumption, depicting a dynamic differentiation from the immature blue starting point in cluster 10 to the mature yellow ending point in cluster 2, 4, 7 and 11. The pseudotime plotting of the immature and mature markers also stated that the cells went from S100A8/9^{+}C1QA/B/C^{-} to S100A8/9^{-}C1QA/B/C^{+}. Several studies in mice have also shown macrophages to be monocyte-derived [19,93–95]. However, genes associated with embryonic origin, such as CD63, DNASE1L3 and ADAMDEC1, were found in some of the macrophage clusters, suggesting the possibility of a small population of embryonically derived intestinal macrophages. Further analyses are needed to establish the origin of intestinal macrophages within the colon.

When looking at the mucosal colonic macrophages *in situ*, the immature macrophages were found to localize in the lower parts of the mucosa. Here, postcapillary venules are present where circulating monocytes are able to enter the colonic tissue from the blood stream. Mature macrophages are strategically located in the upper part of the mucosa, underneath the epithelial layer. These macrophages most likely play an important role in the high turnover of the colonic epithelia by clearance of apoptotic cells. Here, they also sample the microbiota and phagocytose bacteria that have entered. TREM1, thought to be a part of acute inflammation [96], was found to be expressed by macrophages scattered within the lamina propria between the crypts. Inflammatory chemokine and cytokine secretion is upregulated when TREM1 is activated. As intestinal macrophages do not normally express TREM1, the findings of TREM1$^+$CD14$^+$ cells within the colonic mucosa might suggest that these cells are newly recruited monocytes, since monocytes are known to have TREM1 expression [96]. This is in accordance with the scRNA-seq data showing a high TREM1 expression within the immature macrophage clusters.

From the gene expression within the sequenced cells, we uncovered the newly described DC3 subset. These cells are not thought to derive from the classical monocyte lineage, even though they share some common markers. The DC3s were found to secrete cytokines indicating that also they are involved in priming of T cells [83]. From the flow cytometry analysis, we found a DC3 subset within the CD45$^+$HLA-DR$^+$CD14$^+$ cells, by using the CD1c and FcϵRI markers. The same DC3 subset was also identified in blood. From the *in situ* analysis, we found the DC3 subset to localize in the lamina propria underneath the epithelium. Before, it was thought that DCs were able to sample antigens in the gut by extending protrusions between the epithelial cells, but these cells were later found to be macrophages [97]. Our findings may suggest that these cells, with phenotypic similarities to both macrophages and dendritic cells, are of the DC3 subset. Their close proximity to the colonic epithelial layer may indicate that these cells are sampling luminal antigens, possibly through extending protrusions.

4.1.1 Methodological consideration

All colonic tissue were obtained from patients that underwent sigmoid colon cancer surgery, posing a limitation to the study since the tissue was not obtained from healthy individuals.

When isolating single cells, the tissue was digested enzymatically with an enzyme mix containing proteases (which all enzyme mixes do). This may result in alterations of surface markers on the cells.

Compensation matrices used to analyze flow cytometry data were calculated automatically by FlowJo to minimize bias. To avoid unspecific antibody binding of Fc receptors strongly expressed on mononuclear phagocytes, Fc block was used. FMOs were also set up to filter out any background fluorescence. For the intracellular staining, the FMOs were set based on isotype-matched antibodies.

Single-cell RNA sequencing is a useful tool when cell heterogeneity is of interest, but the main downside is the zero-reads as a consequence of small quantities of starting material. Here, we are not able to differentiate between absence of gene expression because of technical limitations or resulting from no expression by the cell. Loss of rare cell populations may also occur, since the gene expression in these cells might appear as genes with low expression when compared to the other genes expressed by a majority of the cells. Sorting of numerous cells may minimize this problem. Computational analyses of the scRNA-seq data should be performed with caution, as biological importance is not synonymous with statistical significance.

For the immunohistochemistry several different techniques were used, as they all have their advantages and drawbacks. Enzymatic staining is more sensitive compared to fluorescent staining, but it is more difficult to interpret images of multiple antibodies. This is easier with fluorescence, were several different antibodies conjugated with distinct fluorochromes may be used simultaneously. Cryopreserved sections are more straightforward to stain as most antibodies work. However, the morphology is not so well conserved compared to formalin fixated slides.

5 Conclusion and further perspectives

In this study we found CD14 to be a well suited marker for macrophage identification in the colon, similar to the findings in the small intestine [20]. Through high-resolution spatiotemporal single-cell analysis we showed a significant heterogeneity within the colonic mucosal macrophages, identifying subsets with immature and mature phenotypes, in addition to pro-inflammatory and tissue-resident subsets. The newly identified DC3 population was also found in the colonic mucosa. We found the different subsets to localize in different parts of the mucosa, where the immature macrophages were situated in the lower parts compared to the mature macrophages found just underneath the colonic epithelial layer.

Further validation of colonic macrophage markers would be of value enabling investigation of the niche of all the clusters identified, by the use of multiplex immunohistochemistry. This could then be combined with spatial transcriptomics. From these findings novel knowledge about which cells contribute to the local macrophage differentiation may emerge. The same analyses may also be performed on tissue derived from diseased colon, such as from colon cancer or IBD, to identified spesific markers for dysregulated macrophages as they may be treatment targets.

Appendix 1

EDTA buffer:

Reagents	Volume	Final concentration
PBS 1x (Ca/Mg free)	500 ml	
FCS	5 ml	1%
0.5M EDTA stock	2 ml	2mM

Enzyme mix for digestion of tissue:

Reagents	Volume per tissue sample
PRMI medium with 10% FCS and 1% Penicillin/streptomycin solution (without phenol red)	10 ml
Liberase TL	2.5 mg/ml
DNase	10 mg/ml

Flow buffer:

Reagents	Volume	Final concentration
PBS	500 ml	
FCS	10 ml	2%
Na acid	500 μl	0.1%

Appendix 2

Supplies	**Product name**	**Company**	**Catalog nr.**
1.5 ml Eppendorf tube	SafeSeal reaction tube, 1.5 ml, PP, PCR Performance Tested	Sarstedt	72.706.400
5 ml flow tubes	Tube, 5 ml, (LxØ): 75 x 12 mm, PP	Sarstedt	55.1578
50 ml tube	Screw cap tube, 50 ml, (LxØ): 114 x 28 mm, PP, with print	Sarstedt	62.547.254
100 ml cups	Container with screw cap, 100 ml, (ØxH): 57 x 76 mm, PP, with safety label, transparent	Sarstedt	75.562.105
Cell sorter	FACS Aria Illu	BD Biosciences	N/A
Centrifuges	Beckman Model TJ-6 Tabletop Centrifuge	Analytical Instruments Brokers LLC	N/A
	Universal Refrigerated Centrifuge Model 5930	Kubota	
	Sigma 3-18K Centrifuge	Sigma	
Coverslips	Dekkglass, rektangulært 24*50mm, tykkelse 0,085-0,115	Marienfeld - VWR	0100222
DAKO grease pen	DAKO pen	Dako - Agilent	S2002
Flow cytometer	BD LSRFortessa X20	BD Biosciences	N/A
IHC machine staining	Ventana DISCOVERY ULTRA	Ventana	N/A
Large cell culture dish	Tissue culture dish, (ØxH): 150 x 20 mm, surface: Standard	Sarstedt	83.3903
Microscope slides	Thermo Scientific™ SuperFrost Plus©	Thermo Scientific	JM1800AMNZ

Appendix 3

Reagent	**Product name**	**Company**	**Catalog nr.**
Antibody Diluent	EnVision FLEX Antibody Diluent	Dako - Agilent	K800621-2
Antigen retrieval buffer pH 6	Target Retrieval Solution 10X Concentrate	Dako - Agilent	S1699
Antigen retrieval buffer pH 9	EnVision FLEX Target Retrieval Solution, High pH	Dako - Agilent	K8004
BerEp4 FITC (Clone Ber-EP4)	Mouse Anti-Human Epithelial Antigen/FITC,	Dako - Agilent	F086001-2
BSA	Bovine Serum Albumin	Sigma	A7511-25G
C1Q FITC (polyclonal)	C1q Complement/FITC (Conjugate)	Dako - Aglient	F025402-2
Calprotectin (immunostaining)	Anti-Human Calprotectin	Gift from I. Dahle (Calpro, Oslo, Norway)	N/A
Calprotectin PE (clone MAC387)	Mouse anti Human Macrophages:RPE	Bio-Rad	MCA874PE
CD1c BV421 (clone L161)	Brilliant Violet 421™ anti-human CD1c Antibody	BioLegend	331526
CD1c (clone M241)	Purified anti-CD1c	Ancell	146-020
CD3 FITC (clone UCHT1)	FITC anti-human CD3 Antibody	BioLegend	300406
CD11c APC (clone S-HCL-3)	APC Mouse Anti-Human CD11c	BD Biosciences	333144
CD14 APC (clone HCD14)	APC anti-human CD14 Antibody	BioLegend	325608
CD14 (clone EPR3653	CD14 (EPR3653) Rabbit Monoclonal Primary Antibody	Cell Marque	114R-14
CD14 Pe-Cy7 (clone HCD14)	PE/Cyanine7 anti-human CD14 Antibody	BioLegend	325618
CD19 Ax488 (clone HIB19)	Alexa Fluor® 488 anti-human CD19 Antibody	BioLegend	302219

CD45 BV510 (clone 30-F11)	Brilliant Violet 510™ anti-human CD45 Antibody	BioLegend	368526
CD68 (clone PG-M1)	Anti-Human CD68	Dako - Agilent	M087601-2
CD141 PE (clone AD5-14H12)	CD141 (BDCA-3) Antibody, anti-human	Miltenyi Biotec	130-113-318
Compensation beads	UltraComp eBeads™ Compensation Beads	Invitrogen - ThermoFisher	01-2222-42
DAB Substrate	Liquid DAB+	Dako - Agilent	K346811-2
DISC Purple	DISCOVERY Purple	Ventana	253-4857
Discovery antibody diluent	DISCOVERY Ab Diluent	Ventana	760-108
DNase I	Deoxyribonuclease I from bovine pancreas	Sigma	D5025
EDTA	Ethylenediaminetetraacetic acid disodium salt solution	Sigma	E7889-100ml
Antibody diluent	EnVision FLEX Antibody Diluent, Diluent, Immunohistochemistry Visualization, 120 mL	Dako - Agilent	K800621-2
DAKO DM822	EnVision FLEX Mini Kit, High pH (Dako Omnis), HRP. Rabbit/Mouse. High pH	Dako - Agilent	K8010/K8012/K8024
Fc Block	FcR Bloking Reagent	Miltenyi Biotec	130-059-901
FcεRI PE (clone AER-37)	PE anti-human FcεRIα Antibody	BioLegend	334610
FCS	Fetal calf serum	Sigma	F2442-500ml
Fix/Perm Solution	BD Cytofix/Cytoperm™ Fixation and Permeabilization Solution	BD Biosciences	554722
Hematoxylin	Hematoksilin Puriss 100 G	Chemi-Teknik	5B-535
Hexamine	Heksamin P.A. 1 KG	Chemi-Teknik	24560-291
HLA-DR PerCP-Cy5.5 (clone L243)	PerCP/Cyanine5.5 anti-human HLA-DR Antibody	BioLegend	307630
IgG (H+L)	Donkey Anti-Rabbit IgG Secondary Antibody, Cy3	Jackson ImmunoResearch	711-165-152

IgG3	Goat anti-Mouse IgG3 Secondary Antibody, Alexa Fluor 488	Invitrogen - ThermoFisher	A-21151
Liberase TL	Liberase™ TL Research Grade	Roche	05401020001
Normal Human serum (NHS)	Human Serum	Sigma	H4522-20ml
PBS	Dulbecco's phosphate-buffered saline	Sigma	D8537-500ml
Penicillin-Streptomycin	Penicillin-Streptomycin Mixture	Lonza	DE17-603E
PermWash	BD Perm/Wash™ Buffer	BD Biosciences	554723
Peroxidase-Blocking Solution	Dako REAL Peroxidase-Blocking Solution	Dako - Agilent	S202386-2
PVA (Made at the lab Tris-Base and Tris-Phosphate)	Tris(hydroxymethyl)aminomethane (TRIS, Trometamol) ≥99%, Gen-Apex Molecular biology grade	VWR Chemicals	33621.260
	Trizma® phosphate monobasic	Sigma	T-1758
RPMI	RPMI 1640 Medium, no phenol red	Gibco	11835030
Teal HRP Sub	DISCOVERY Teal HRP Substrate	Ventana	253-6016
TO-PRO™-1	TO-PRO™-1 Iodide (515/531) - 1 mM Solution in DMSO	ThermoFisher	T3602
TREM1	IHC-plus™ Polyclonal Rabbit anti-Human TREM1 Antibody (aa162-191, IHC, WB)	LifeSpan Biosciences	LS-B10694
TBS wash buffer	Reaction Buffer (10x)	Ventana Medical Systems, Inc	950-300
Viability Dye eFluor 780	Fixable Viability Dye eFluor 780	eBioscience	65-0865-14
Xylen	Xylene (mixture of isomers)	VWR Chemicals	28975.291
10x Single Cell kit	10x Chromium Single Cell 3' kit	10x Genomics	120237
70% - 100% alcohol	Absolutt alkohol prima 99,9% 1 liter	Antibac	600068

Appendix 4

Software	**Company**	**Link**
FlowJo 10.6.1	Tree Star	N/A
Monocle 3 R package 0.1.3	Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014). “The dynamics and regulators of cell fate decisions are revealed by pseudo-temporal ordering of single cells.” Nature Biotechnology.	https://cole-trapnell-lab.github.io/monocle3/
R R Core https://www.r-project.org/	R R Core https://www.r-project.org/	R R Core https://www.r-project.org/
R package ggplot2 3.3.0	R CRAN	https://ggplot2.tidyverse.org/
R package cowplot 1.0.0	R CRAN	https://cran.r-project.org/web/packages/cowplot/index.html
R package Clustree 0.4.2	R CRAN	https://cran.r-project.org/web/packages/clustree/index.html
R package xlsx 0.6.5	R CRAN	https://cran.r-project.org/web/packages/xlsx/index.html
R package patchwork 1.1.1	R CRAN	https://cran.r-project.org/web/packages/patchwork/index.html
Seurat v3.0.0	Github; Butler et al., 2018	https://github.com/satijalab/seurat/releases/tag/v3.0.0

References

1. Abbas AK, Lichtman AH, Pillai S. Cellular and molecular immunology. Ninth edition. Philadelphia, Pa: Elsevier; 2018.
2. Kumar BV, Connors T, Farber DL. Human T cell development, localization, and function throughout life. Immunity. 2018 Feb 20;48(2):202–13.
3. Zhu J, Yamane H, Paul WE. Differentiation of Effector CD4 T Cell Populations. Annu Rev Immunol. 2010;28:445–89.
4. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. The major histocompatibility complex and its functions. Immunobiol Immune Syst Health Dis 5th Ed. 2001
5. Heath WR, Carbone FR. Dendritic cell subsets in primary and secondary T cell responses at body surfaces. Nat Immunol. 2009 Dec;10(12):1237–44.
6. Mildner A, Jung S. Development and Function of Dendritic Cell Subsets. Immunity. 2014 May 15;40(5):642–56.
7. Schraml BU, Reis e Sousa C. Defining dendritic cells. Curr Opin Immunol. 2015 Feb 1;32:13–20.
8. Schlitzer A, Sivakamasundari V, Chen J, Sumatoh HRB, Schreuder J, Lum J, Malleret B, Zhang S, Larbi A, Zolezzi F, Renia L, Poidinger M, Naik S, Newell EW, Robson P, Ginhoux F. Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. Nat Immunol. 2015 Jul;16(7):718–28.
9. Collin M, Bigley V. Human dendritic cell subsets: an update. Immunology. 2018 May;154(1):3–20.
10. Villadangos JA, Young L. Antigen-Presentation Properties of Plasmacytoid Dendritic Cells. Immunity. 2008 Sep 19;29(3):352–61.
11. Shapouri-Moghaddam A, Mohammadian S, Vazini H, Taghadosi M, Esmaeili S-A, Mardani F, Seifi B, Mohammadi A, Afshari JT, Sahebkar A. Macrophage plasticity, polarization, and function in health and disease. J Cell Physiol. 2018;233(9):6425–40.
12. Hilhorst M, Shirai T, Berry G, Goronzy JJ, Weyand CM. T Cell–Macrophage Interactions and Granuloma Formation in Vasculitis. Front Immunol. 2014;5.
13. van Furth R, Cohn ZA, Hirsch JG, Humphrey JH, Spector WG, Langevoort HL. The mononuclear phagocyte system: a new classification of macrophages, monocytes, and their precursor cells. Bull World Health Organ. 1972;46(6):845–52.
14. De Schepper S, Verheijden S, Aguilera-Lizarraga J, Viola MF, Boesmans W, Stakenborg N, Voytyuk I, Schmidt I, Boeckx B, Dierckx de Casterlé I, Baekelandt V, Gonzalez Dominguez E, Mack M, Depoortere I, De Strooper B, Sprangers B, Himmelreich U, Soenen S, Guilliams M, Vanden Berghe P, Jones E, Lambrechts D, Boeckxstaens G. Self-Maintaining Gut Macrophages Are Essential for Intestinal Homeostasis. Cell. 2018 Oct 4;175(2):400-415.e13.
15. Shi C, Pamer EG. Monocyte recruitment during infection and inflammation. Nat Rev Immunol. 2011 Nov;11(11):762–74.
16. Na YR, Stakenborg M, Seok SH, Matteoli G. Macrophages in intestinal inflammation and resolution: a potential therapeutic target in IBD. Nat Rev Gastroenterol Hepatol. 2019 Sep;16(9):531–43.
17. Berthold DL, Jones KDJ, Udalova IA. Regional specialization of macrophages along the gastrointestinal tract. Trends Immunol. 2021 Sep 1;42(9):795–806.
18. Bian Z, Gong Y, Huang T, Lee CZW, Bian L, Bai Z, Shi H, Zeng Y, Liu C, He J, Zhou J, Li X, Li Z, Ni Y, Ma C, Cui L, Zhang R, Chan JKY, Ng LG, Lan Y, Ginhoux

F, Liu B. Deciphering human macrophage development at single-cell resolution. Nature. 2020 Jun;582(7813):571–6.
19. Bain CC, Bravo-Blas A, Scott CL, Gomez Perdiguero E, Geissmann F, Henri S, Malissen B, Osborne LC, Artis D, Mowat AM. Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. Nat Immunol. 2014 Oct;15(10):929–37.
20. Bujko A, Atlasy N, Landsverk OJB, Richter L, Yaqub S, Horneland R, Øyen O, Aandahl EM, Aabakken L, Stunnenberg HG, Bækkevold ES, Jahnsen FL. Transcriptional and functional profiling defines human small intestinal macrophage subsets. J Exp Med. 2018 Feb 5;215(2):441–58.
21. Zamani F, Zare Shahneh F, Aghebati-Maleki L, Baradaran B. Induction of CD14 Expression and Differentiation to Monocytes or Mature Macrophages in Promyelocytic Cell Lines: New Approach. Adv Pharm Bull. 2013 Dec;3(2):329–32.
22. Chistiakov DA, Killingsworth MC, Myasoedova VA, Orekhov AN, Bobryshev YV. CD68/macrosialin: not just a histochemical marker. Lab Invest. 2017 Jan;97(1):4–13.
23. Etzerodt A, Moestrup SK. CD163 and Inflammation: Biological, Diagnostic, and Therapeutic Aspects. Antioxid Redox Signal. 2013 Jun 10;18(17):2352–63.
24. Schulz D, Severin Y, Zanotelli VRT, Bodenmiller B. In-Depth Characterization of Monocyte-Derived Macrophages using a Mass Cytometry-Based Phagocytosis Assay. Sci Rep. 2019 Feb 13;9(1):1925.
25. Chakarov S, Lim HY, Tan L, Lim SY, See P, Lum J, Zhang X-M, Foo S, Nakamizo S, Duan K, Kong WT, Gentek R, Balachander A, Carbajo D, Bleriot C, Malleret B, Tam JKC, Baig S, Shabeer M, Toh S-AES, Schlitzer A, Larbi A, Marichal T, Malissen B, Chen J, Poidinger M, Kabashima K, Bajenoff M, Ng LG, Angeli V, Ginhoux F. Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches. Science [Internet]. 2019 Mar 15 [cited 2020 Nov 2];363(6432). Available from: https://science.sciencemag.org/content/363/6432/eaau0964
26. Gordon S, Plüddemann A. Tissue macrophages: heterogeneity and functions. BMC Biol. 2017 Dec;15(1):53.
27. Viola A, Munari F, Sánchez-Rodríguez R, Scolaro T, Castegna A. The Metabolic Signature of Macrophage Responses. Front Immunol. 2019 Jul 3;10:1462.
28. Kang B, Alvarado LJ, Kim T, Lehmann ML, Cho H, He J, Li P, Kim B-H, Larochelle A, Kelsall BL. Commensal microbiota drive the functional diversification of colon macrophages. Mucosal Immunol. 2020 Mar;13(2):216–29.
29. Cochain C, Vafadarnejad E, Arampatzi P, Pelisek J, Winkels H, Ley K, Wolf D, Saliba A-E, Zernecke A. Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis. Circ Res. 2018 Jun 8;122(12):1661–74.
30. Nahrendorf M, Swirski FK. Abandoning M1/M2 for a Network Model of Macrophage Function. Circ Res. 2016 Jul 22;119(3):414–7.
31. Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, De Nardo D, Gohel TD, Emde M, Schmidleithner L, Ganesan H, Nino-Castro A, Mallmann MR, Labzin L, Theis H, Kraut M, Beyer M, Latz E, Freeman TC, Ulas T, Schultze JL. Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation. Immunity. 2014 Feb 20;40(2):274–88.
32. Mantovani A, Marchesi F. IL-10 and Macrophages Orchestrate Gut Homeostasis. Immunity. 2014 May 15;40(5):637–9.
33. Mowat AM, Scott CL, Bain CC. Barrier-tissue macrophages: functional adaptation to environmental challenges. Nat Med. 2017 Nov;23(11):1258–70.
34. Iyer SS, Cheng G. Role of Interleukin 10 Transcriptional Regulation in Inflammation

and Autoimmune Disease. Crit Rev Immunol. 2012;32(1).
35. Chang C, Werb Z. The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis. Trends Cell Biol. 2001 Nov;11(11):S37–43.
36. Cui N, Hu M, Khalil RA. Biochemical and Biological Attributes of Matrix Metalloproteinases. Prog Mol Biol Transl Sci. 2017;147:1–73.
37. Tang T, Scambler TE, Smallie T, Cunliffe HE, Ross EA, Rosner DR, O'Neil JD, Clark AR. Macrophage responses to lipopolysaccharide are modulated by a feedback loop involving prostaglandin E2, dual specificity phosphatase 1 and tristetraprolin. Sci Rep. 2017 Jun 28;7(1):4350.
38. Cerovic V, Bain CC, Mowat AM, Milling SWF. Intestinal macrophages and dendritic cells: what's the difference? Trends Immunol. 2014 Jun 1;35(6):270–7.
39. Lim K-H, Staudt LM. Toll-Like Receptor Signaling. Cold Spring Harb Perspect Biol. 2013 Jan;5(1):a011247.
40. O'Neill LAJ, Bowie AG. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. Nat Rev Immunol. 2007 May;7(5):353–64.
41. Satoh T, Akira S. Toll-Like Receptor Signaling and Its Inducible Proteins. Microbiol Spectr. 2016 Dec 23;4(6):4.6.41.
42. Arroyo-Espliguero R, Avanzas P, Jeffery S, Kaski JC. CD14 and toll-like receptor 4: a link between infection and acute coronary events? Heart. 2004 Sep;90(9):983–8.
43. Medzhitov R. Toll-like receptors and innate immunity. Nat Rev Immunol. 2001 Nov;1(2):135–45.
44. McInturff JE, Modlin RL, Kim J. The Role of Toll-like Receptors in the Pathogenesis and Treatment of Dermatological Disease. J Invest Dermatol. 2005 Jul 1;125(1):1–8.
45. Chang S-Y, Ko H-J, Kweon M-N. Mucosal dendritic cells shape mucosal immunity. Exp Mol Med. 2014 Mar;46(3):e84–e84.
46. Schulz O, Pabst O. Antigen sampling in the small intestine. Trends Immunol. 2013 Apr 1;34(4):155–61.
47. Joffe AM, Bakalar MH, Fletcher DA. Macrophage phagocytosis assay with reconstituted target particles. Nat Protoc. 2020 Jul;15(7):2230–46.
48. Poesen R, Meijers B, Evenepoel P. The Colon: An Overlooked Site for Therapeutics in Dialysis Patients. Semin Dial. 2013;26(3):323–32.
49. Mahadevan V. Anatomy of the caecum, appendix and colon. Surg Oxf. 2020 Jan;38(1):1–6.
50. Rao JN, Wang J-Y. Regulation of Gastrointestinal Mucosal Growth [Internet]. San Rafael (CA): Morgan & Claypool Life Sciences; 2010. (Integrated Systems Physiology: from Molecule to Function to Disease).
51. Kataoka K. The intestinal microbiota and its role in human health and disease. J Med Invest. 2016;63(1.2):27–37.
52. Naser SA, Arce M, Khaja A, Fernandez M, Naser N, Elwasila S, Thanigachalam S. Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. World J Gastroenterol WJG. 2012 Feb 7;18(5):412–24.
53. Alberts B, Wilson JH, Hunt T. Molecular biology of the cell. 6th ed. New York: Garland Science; 2015. XXXIV, 1342.
54. Olsen TK, Baryawno N. Introduction to Single-Cell RNA Sequencing. Curr Protoc Mol Biol. 2018;122(1):e57.
55. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data. F1000Research. 2016 Aug 31;5:2122.
56. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med. 2017 Aug 18;9(1):75.

57. AlJanahi AA, Danielsen M, Dunbar CE. An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. Mol Ther - Methods Clin Dev. 2018 Sep 21;10:189–96.
58. Liao X, Makris M, Luo XM. Fluorescence-activated Cell Sorting for Purification of Plasmacytoid Dendritic Cells from the Mouse Bone Marrow. JoVE J Vis Exp. 2016 Nov 4;(117):e54641.
59. Shehadul Islam M, Aryasomayajula A, Selvaganapathy PR. A Review on Macroscale and Microscale Cell Lysis Methods. Micromachines. 2017 Mar;8(3):83.
60. Ma C, Lyons-Weiler M, Liang W, LaFramboise W, Gilbertson JR, Becich MJ, Monzon FA. In Vitro Transcription Amplification and Labeling Methods Contribute to the Variability of Gene Expression Profiling with DNA Microarrays. J Mol Diagn. 2006 May 1;8(2):183–92.
61. Single Cell 3' Reagent Kit v.2 [Internet]. [cited 2021 Jan 18]. Available from: https://assets.ctfassets.net/an68im79xiti/RT8DYoZzhDJRBMrJCmVxl/6a0ed8015d89bf9602128a4c9f8962c8/CG00052_SingleCell3_ReagentKitv2UserGuide_RevF.pdf
62. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019 Jun 1;15(6):e8746.
63. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, Teichmann SA. Classification of low quality cells from single-cell RNA-seq data. Genome Biol. 2016 Feb 17;17(29).
64. Steuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck III WM, Hao Y, Stoeckius M, Smibert P, Satija R. Seurat [Internet]. [cited 2021 Mar 4]. Available from: https://satijalab.org/seurat/articles/pbmc3k_tutorial.html
65. Ringnér M. What is principal component analysis? Nat Biotechnol. 2008 Mar;26(3):303–4.
66. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol. 2018 May;36(5):421–7.
67. Yip SH, Sham PC, Wang J. Evaluation of tools for highly variable gene discovery from single-cell RNA-seq data. Brief Bioinform. 2019 Jul 19;20(4):1583–9.
68. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018 Feb 6;19(1):15.
69. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 May;36(5):411–20.
70. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019 Mar 26;9(1):5233.
71. Parra-Hernández RM, Posada-Quintero JI, Acevedo-Charry O, Posada-Quintero HF. Uniform Manifold Approximation and Projection for Clustering Taxa through Vocalizations in a Neotropical Passerine (Rough-Legged Tyrannulet, Phyllomyias burmeisteri). Anim Open Access J MDPI. 2020 Aug 12;10(8). A
72. Kobak D, Berens P. The art of using t-SNE for single-cell transcriptomics. Nat Commun. 2019 Nov 28;10(1):5416.
73. Monocle [Internet]. [cited 2021 Feb 5]. Available from: http://cole-trapnell-lab.github.io/monocle-release/monocle3/#citation
74. Cheng B, Yang J, Yan S, Fu Y, Huang TS. Learning with L1-graph for image analysis. IEEE Trans Image Process. 2010 Apr;19(4):858–66.
75. Mao Q, Yang L, Wang L, Goodison S, Sun Y. SimplePPT: A Simple Principal Tree Algorithm. Proc 2015 SIAM Int Conf Data Min. 2015 Jun 30;792–800.
76. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. Pseudo-temporal ordering of individual cells reveals

dynamics and regulators of cell fate decisions. Nat Biotechnol. 2014 Apr;32(4):381–6.

77. Richter L, Landsverk OJB, Atlasy N, Bujko A, Yaqub S, Horneland R, Øyen O, Aandahl EM, Lundin KEA, Stunnenberg HG, Bækkevold ES, Jahnsen FL. Transcriptional profiling reveals monocyte-related macrophages phenotypically resembling DC in human intestine. Mucosal Immunol. 2018 Sep;11(5):1512–23.
78. Bøyum A, Løvhaug D, Tresland L, Nordlie EM. Separation of Leucocytes: Improved Cell Purity by Fine Adjustments of Gradient Medium Density and Osmolality. Scand J Immunol. 1991;34(6):697–712.
79. Fixation/Permeabilization protocol [Internet]. [cited 2021 Feb 17]. Available from: https://www.bdbiosciences.com/ds/pm/others/554714_554715_555028_Book_Website.pdf
80. Fixable Viability Dye Cell Staining Protocol [Internet]. [cited 2021 Sep 21]. Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/65-0865.pdf
81. CS&T Beads [Internet]. [cited 2021 Sep 21]. Available from: https://www.bdbiosciences.com/en-ca/products/reagents/flow-cytometry-reagents/clinical-diagnostics/process-and-quality-controls/cs-t-beads.656505
82. Krenacs L, Krenacs T, Stelkovics E, Raffeld M. Heat-Induced Antigen Retrieval for Immunohistochemical Reactions in Routinely Processed Paraffin Sections. Methods Mol Biol Clifton NJ. 2010;588:103–19.
83. Bourdely P, Anselmi G, Vaivode K, Ramos RN, Missolo-Koussou Y, Hidalgo S, Tosselo J, Nuñez N, Richer W, Vincent-Salomon A, Saxena A, Wood K, Lladser A, Piaggio E, Helft J, Guermonprez P. Transcriptional and Functional Analysis of CD1c+ Human Dendritic Cells Identifies a CD163+ Subset Priming CD8+CD103+ T Cells. Immunity. 2020 Aug 18;53(2):335-352.e8.
84. Fawkner-Corbett D, Antanaviciute A, Parikh K, Jagielowicz M, Gerós AS, Gupta T, Ashley N, Khamis D, Fowler D, Morrissey E, Cunningham C, Johnson PRV, Koohy H, Simmons A. Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell. 2021 Feb 4;184(3):810-826.e23.
85. Tacke F, Zimmermann HW. Macrophage heterogeneity in liver injury and fibrosis. J Hepatol. 2014 May 1;60(5):1090–6.
86. Blériot C, Ginhoux F. Understanding the Heterogeneity of Resident Liver Macrophages. Front Immunol. 2019 Nov 19;10:2694.
87. Morales-Nebreda L, Misharin AV, Perlman H, Budinger GRS. The heterogeneity of lung macrophages in the susceptibility to disease. Eur Respir Rev. 2015 Sep 1;24(137):505–9.
88. Xu-Vanpala S, Deerhake ME, Wheaton JD, Parker ME, Juvvadi PR, MacIver N, Ciofani M, Shinohara ML. Functional heterogeneity of alveolar macrophage population based on expression of CXCL2. Sci Immunol. 2020 Aug 7;5(50):eaba7350.
89. Guilliams M, Thierry GR, Bonnardel J, Bajenoff M. Establishment and Maintenance of the Macrophage Niche. Immunity. 2020 Mar 17;52(3):434–51.
90. Hong S, Beja-Glasser VF, Nfonoyim BM, Frouin A, Li S, Ramakrishnan S, Merry KM, Shi Q, Rosenthal A, Barres BA, Lemere CA, Selkoe DJ, Stevens B. Complement and Microglia Mediate Early Synapse Loss in Alzheimer Mouse Models. Science. 2016 May 6;352(6286):712–6.
91. Hulsmans M, Clauss S, Xiao L, Aguirre AD, King KR, Hanley A, Hucker WJ, Wülfers EM, Seemann G, Courties G, Iwamoto Y, Sun Y, Savol AJ, Sager HB, Lavine KJ, Fishbein GA, Capen DE, Da Silva N, Miquerol L, Wakimoto H, Seidman CE, Seidman JG, Sadreyev RI, Naxerova K, Mitchell RN, Brown D, Libby P, Weissleder R, Swirski FK, Kohl P, Vinegoni C, Milan DJ, Ellinor PT, Nahrendorf M. Macrophages Facilitate Electrical Conduction in the Heart. Cell. 2017 Apr

20;169(3):510-522.e20.
92. Guilliams M, De Kleer I, Henri S, Post S, Vanhoutte L, De Prijck S, Deswarte K, Malissen B, Hammad H, Lambrecht BN. Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. J Exp Med. 2013 Sep 23;210(10):1977–92.
93. Bain CC, Scott CL, Uronen-Hansson H, Gudjonsson S, Jansson O, Grip O, Guilliams M, Malissen B, Agace WW, Mowat AM. Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6Chi monocyte precursors. Mucosal Immunol. 2013 May;6(3):498–510.
94. Tamoutounour S, Henri S, Lelouard H, de Bovis B, de Haar C, van der Woude CJ, Woltman AM, Reyal Y, Bonnet D, Sichien D, Bain CC, Mowat AM, Reis e Sousa C, Poulin LF, Malissen B, Guilliams M. CD64 distinguishes macrophages from dendritic cells in the gut and reveals the Th1-inducing role of mesenteric lymph node macrophages during colitis. Eur J Immunol. 2012 Dec;42(12):3150–66.
95. Schridde A, Bain CC, Mayer JU, Montgomery J, Pollet E, Denecke B, Milling SWF, Jenkins SJ, Dalod M, Henri S, Malissen B, Pabst O, Mcl Mowat A. Tissue-specific differentiation of colonic macrophages requires TGFβ receptor-mediated signaling. Mucosal Immunol. 2017 Nov;10(6):1387–99.
96. Schenk M, Bouchon A, Seibold F, Mueller C. TREM-1–expressing intestinal macrophages crucially amplify chronic inflammation in experimental colitis and inflammatory bowel diseases. J Clin Invest. 2007 Oct 1;117(10):3097–106.
97. Rescigno M. Before They Were Gut Dendritic Cells. Immunity. 2009 Sep 18;31(3):454–6.