# Functional and cellular heterogeneity of the myeloid cell system

**Patrick Günther**
from Stendal, Germany

ORCID: 0000-0002-9135-7819

Submitted in total fulfillment of the requirements of the joint degree of

Doctor of Philosophy (Ph.D.)

of

The Medical Faculty
The Rheinische Friedrich-Wilhelms-Universität Bonn
and
The Department of Microbiology and Immunology
The University of Melbourne

Bonn/ Melbourne, 2020

Performed and approved by the Medical Faculty of the Rheinische Friedrich-Wilhelms-Universität Bonn and the University of Melbourne

| | | |
|---|---|---|
| **1.** | **Supervisor:** | **Professor Dr. Eicke Latz** |
| **2.** | **Supervisor (shared):** | **Professor Dr. Joachim L. Schultze** |
| | | **Professor Dr. Stephen J. Turner** |

Date of submission: September 20th, 2019

Date of the oral examination: August 6th, 2020

Life and Medical Sciences Institute Bonn, Managing Director: Professor Dr. Waldemar Kolanus

## Contents

## Abbreviations

**A**

| | |
|---|---|
| AAAS | American Association for the Advancement of Science |
| AS-DC | $AXL^{+}Siglec6^{+}$ DC |
| AML | Acute myeloid leukemia |
| ATAC-seq | Assay for transposase-accessible chromatin using sequencing |

**B**

| | |
|---|---|
| BCG | Bacille Calmette–Guérin |

**C**

| | |
|---|---|
| C2TA | Class 2 transcription activator |
| CD | Chow diet |
| CDC | Classical DC |
| cDNA | Complementary DNA |
| CLP | Common lymphoid progenitor |
| CMML | Chronic myelomonocytic leukemia |
| COPD | Chronic obstructive pulmonary disease |
| CVD | Cardiovascular disease |
| CXCR3 | CXC chemokine receptor |
| CyToF | Cytometry by time-of-flight |

**D**

| | |
|---|---|
| DBSCAN | Density-based spatial clustering with noise |
| DC | Dendritic cell |
| DNA | Deoxyribonucleic acid |
| DOI | Digital Object Identifier |

**E**

| | |
|---|---|
| E.g. | *Exempli gratia* |
| EMP | Erythroid-myeloid progenitors |
| Et al. | *Et alii* |

**F**

| | |
|---|---|
| FACS | Fluorescent activated cell sorting |
| FC | Flow cytometry |

**G**

| | |
|---|---|
| GMP | Granulocyte-monocyte progenitor |

**H**

| | |
|---|---|
| HET-E | Heterokaryon incompatibility |
| HLA-DR | Human Leukocyte Antigen – DR isotype |

**I**

| | |
|---|---|
| IFN | Interferon |
| IL | Interleukin |

**L**

| | |
|---|---|
| LDL | Low-Density Lipoprotein |
| LDLR | LDL-receptor |
| LIN | Lineage marker |
| LPS | Lipopolysaccharides |
| LRR | Leucine-rich repeat |

**M**

| | |
|---|---|
| MARS-seq | Massively parallel single-cell RNA-sequencing |
| MDS | Myelodysplastic syndromes |
| MPS | Mononuclear phagocyte system |
| MRNA | Messenger RNA |

**N**

| | |
|---|---|
| NACHT | NAIP, C2TA, HET-E and TP1 |
| NAIP | Neuronal apoptosis inhibitor protein |
| NCD | Non-communicable diseases |
| NGS | Next-generation sequencing |
| NK | Natural killer |
| NLRP3 | NACHT, LRR and PYD domains-containing protein 3 |

**O**

| | |
|---|---|
| OxLDL | Oxidized LDL |

**P**

| | |
|---|---|
| PBMC | Peripheral blood mononuclear cell |
| PCR | Polymerase chain reaction |
| PDC | Plasmacytoid DC |
| Ph.D. | Doctor of Philosophy |
| PreDC | Predendritic cells |
| PreMacs | Precursor of macrophages |
| PYD | PYRIN domain |

**Q**

| | |
|---|---|
| QTL | Quantitative trait locus |

**R**

| | |
|---|---|
| RNA | Ribonucleic acid |
| RNA-seq | RNA sequencing |

**S**

| | |
|---|---|
| ScRNA-seq | Single-cell RNA-seq |
| SIGLEC6 | Sialic Acid Binding Ig Like Lectin 6 |

| | | |
|---|---|---|
| **T** | | |
| | TET2 | Tet Methylcytosine Dioxygenase 2 |
| | TF | Transcription factor |
| | TP1 | Telomerase-associated protein 1 |
| | TSNE | T-stochastic neighbor embedding |
| **U** | | |
| | UMI | Unique molecular identifier |
| **W** | | |
| | WD | Western diet |
| **Y** | | |
| | YS | Yolk sac |

## Abstract

Cells of the myeloid lineage form the innate part of the immune system and are characterized by a high level of functional plasticity, which is required to address the diverse set of functions of these mononuclear cells. Monocytes, Macrophages and dendritic cells (DC) are collectively categorized as the mononuclear phagocyte system (MPS), to highlight their functional equipment that specializes them to the phagocytosis of pathogens as a starting point to elicit an immune response. Besides this role, cells of the MPS are also involved in a wide variety of homeostatic functions including early development and regulation of physiological processes. However, the multitude of mechanisms required to acquire this functional plasticity remains poorly understood. The work that has been performed in the scope of this dissertation aimed to advance current knowledge of the causes and consequences of functional and cellular plasticity of the myeloid immune system. High-dimensional characterization of the effects of a Western diet on myeloid immune cell progenitor cells revealed a long-term transcriptional and epigenetic reprogramming of the myeloid cell compartment. The formation of an innate immune memory in myeloid progenitor cells leads to lasting inflammatory priming of monocytes, which may directly contribute to the progression of myeloid cell-associated diseases.

In addition, single-cell RNA-seq elucidated unreported cellular heterogeneity of the monocyte and dendritic cell compartment in human peripheral blood. A combination of phenotypic and transcriptional analyzes resulted in a precise categorization of the human DC compartment consisting of pDCs, cDC1, two cDC2 subsets, and a deeply characterized preDC subset. Furthermore, a universal strategy for the integration of cellular atlases was conceptualized and applied to establish a consensus map of the human DC and monocyte cell space. This thesis provides mechanistic insights into the cellular composition of myeloid cells and their functional plasticity, which will form the foundation for further investigations into the dynamic changes of the immune cell compartment during diseases and will be critically relevant for designing effective treatments for a wide variety of pathologies linked to myeloid cells.

## Declaration

I hereby declare that the dissertation entitled “Functional and cellular heterogeneity of the myeloid cell system” was written entirely by myself. Further, I declare that no other sources or resources than those stated were used and that the parts of the thesis which were taken from other works including electronic media in terms of wording or meaning are indicated as such followed by the corresponding reference. The underlying thesis towards the degree of Doctor of Philosophy follows the guidelines of the Rheinische Friedrich-Wilhelms Universität and the University of Melbourne.

This is to certify that,

(i) the thesis comprises only my original work towards the Ph.D. except where indicated in the Preface,

(ii) due acknowledgment has been made in the text to all other material used,

(iii) the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies, and appendices.

## Preface

This thesis is based on the work that I have partially conducted in national and international collaborations. It consists of four original research articles and one review article dealing with the consequences of cellular and functional heterogeneity in the myeloid cell system. My contribution to the publications within each chapter was as follows:

Publication I: 80%
Published in *Frontiers in Immunology* on October 4th, 2019.

Publication II: 55%
Published in *Cell* on January 11th, 2018.

Publication III: 20%
Published in *Science* on June 9th, 2017.

Publication IV: 70%
Submitted for publication. A preprint version was uploaded to *biorxiv.org* on June 3rd, 2019.

Publication V: 15%
Published in *Science* on September 9th, 2016.

I acknowledge the valuable contributions of others to publications presented in this thesis:

Publication I: Prof. Joachim L. Schultze

Publication II: Dr. Anette Christ

Publication III: Dr. Branko Cirovic and Kevin Baßler

Publication IV: Peter See, Charles-Antoine Dutertre and Jinmiao Chen

Publication V: Elvira Mass, Ivan Ballesteros, Matthias Farlik and Florian Halbritter

## Acknowledgments

The work presented in this thesis would not have been possible without the incredible support I have received from many people, including supervisors, colleagues, friends, and family. I want to thank each of you for your continued support in accomplishing my goals.

First and foremost, I want to thank my supervisors Prof. Dr. Joachim L. Schultze and Prof. Dr. Stephen J. Turner for giving me the possibility of working on a fascinating scientific project. Joachim, thank you very much for your trust and support. I'm very grateful that you provided an environment, which allowed me to grow and to challenge myself. Thank you for your mentorship and inspiration, for giving me a lot of opportunities, for pushing me and for always offering a helping hand. Steve, I'm very grateful that you were my supervisor during my Ph.D. candidature. I'm thankful for the possibility to work with you, and I want to thank you for allowing me to collaborate on an exciting project. Thanks for your guidance and for being a great mentor. I want to show my gratitude to Prof. Dr. Katherine Kedzierska for being far more than the chair of my advisory committee. Katherine, I want to sincerely thank you for the warm welcome in Melbourne and for integrating me into your research group without any hesitation. It was an honor to learn from your exciting work.

I want to thank Prof. Eicke Latz for serving as my supervisor and for being an inspiring scientist. I'm very thankful for the great collaboration and for the opportunity you provided to get involved and to shape a fantastic project. In this regard, I want to express my gratitude to Dr. Anette Christ for being the best collaborator one could imagine.

I want to express my special thanks to Dr. Thomas Ulas for encouraging and mentoring my move into the field of computational biology. Thank you so much for being a terrific mentor and a wonderful friend. Furthermore, I would like to thank my colleague and friend Kevin Baßler for being constant support from the beginning to the end, at work and in private. I will be forever in your debt for turning my loss into a win.

Furthermore, I want to pay my regards to various national and international collaborators for sharing their exciting projects. Special thanks to Prof. Dr. Frederic Geissmann, Dr. Elvira Mass, Prof. Dr. Florent Ginhoux, Dr. Peter See, Prof. Dr. Vishwa Dixit, and Dr. Christina Camell for inspiring and successful collaborations. I want to show my gratitude to Prof. Dr. Andreas Schlitzer and Dr. Branko Cirovic for an outstanding partnership regarding our project on dendritic cells and monocytes. Branko, thank you for your inspiration, your patience, and your assistance sometimes even until sunrise.

I want to present my special thanks to Prof. Dr. Christian Kurts and Prof. Dr. Sammy Bedoui for initiating a terrific graduate school, providing me the opportunity to grow scientifically but also on a personal level. In this regard, I want to thank Lucie Delforge and Dr. Marie Greyer for the coordination of the graduate school and for being always helpful with any problem.

Thank you to all past and present members of the Schultze laboratory for creating a helpful and friendly atmosphere. Special thanks to Elke Meyer, Kathrin Klee, Jil Sander, Stefanie Warnat-Herresthal, Nico Reusch, Jonas Schulte-Schrepping, Lorenzo Bonaguro, Matthias Becker and Kristian Händler for all sorts of help ranging from administrative support, programming/ computational assistance to fruitful scientific and non-scientific discussions. You guys are great, and I couldn’t ask for better colleagues. Likewise, I want to thank all members of the Kedzierska and Turner laboratories for warmly welcoming me and integrating me into their lives. Besides, I want to thank Emily Hinkley for careful reading of the dissertation.

I want to express my gratitude to the Hans und Eugenia Jütting Stiftung for providing me a stipend, which allowed me to focus on my scientific training.

This challenge would not have been possible without the support and encouragement of many important individuals in my life. Thank you very much.

Liebe Mama, Papa und David, ich bin euch unendlich dankbar, dass ihr mir immer den Rücken freigehalten habt. Ihr habt mich stets unterstützt und habt mich bestärkt meinen Weg zu gehen, auch wenn ihr nicht immer genau wusstet wie dieser aussieht. Ich danke euch von ganzem Herzen.

## List of publications

1. **Günther, P.** and Schultze, J.L. (2019) ‘Mind the Map: Technology Shapes the Myeloid Cell Space’. *Frontiers in Immunology*, 10, p. 2287. DOI: 10.3389/ FIMMU.2019.02287.

2. **Günther, P.*,** Cirovic, B.*, Baßler, K.*, Händler, K., Becker, M., Dutertre, C.A.A., Bigley, V., Newell, E., Collin, M., Ginhoux, F., Schlitzer, A. and Schultze, J.L. (2019) ‘A Rule-Based Data-Informed Cellular Consensus Map of the Human Mononuclear Phagocyte Cell Space’. *BioRxiv*, p. 658179. DOI: 10.1101/658179.

3. Camell, C.D., **Günther, P.**, Lee, L., Goldberg E.L., Spadaro O., Youm Y., Bartke A., Hubbard G.B., Ikeno Y., Ruddle N.H., Schultze J.L. and Dixit V.D. (2019) ‘Aging induces Nlrp3 inflammasome dependent adipose B cell expansion to impair metabolic homeostasis’. *Cell Metabolism*, 30(6), pp. 1024–1039.e6, DOI:https://doi.org/10.1016/j.cmet.2019.10.006

4. Baßler, K., **Günther, P.**, Schulte-Schrepping, J., Becker, M. and Biernat, P. (2019) ‘A Bioinformatic Toolkit for Single-Cell mRNA Analysis’. In Proserpio, V. (ed.) *Single Cell Methods: Sequencing and Proteomics, Methods in Molecular Biology*. Springer Nature, New York, NY, pp. 433–455. DOI: 10.1007/978-1-4939-9240-9_26.

5. Becker, M., Chabbi, M., Warnat-Herresthal, S., Klee, K., Schulte-Schrepping, J., Biernat, P., **Günther, P.,** Baβler, K., Craig, R., Schultze, H., Singhal, S., Ulas, T. and Schultze, J.L. (2019) ‘Memory-Driven Computing Accelerates Genomic Data Processing’. *BioRxiv*, p. 519579. DOI: 10.1101/519579.

6. Christ, A.*, **Günther, P.***, Lauterbach, M.A.R., Duewell, P., Biswas, D., Pelka, K., Scholz, C.J., Oosting, M., Haendler, K., Baßler, K., Klee, K., Schulte-Schrepping, J., Ulas, T., Moorlag, S.J.C.F.M., Kumar, V., Park, M.H., Joosten, L.A.B., Groh, L.A., Riksen, N.P., Espevik, T., Schlitzer, A., Li, Y., Fitzgerald, M.L., Netea, M.G., Schultze, J.L. and Latz, E. (2018) ‘Western Diet Triggers NLRP3-Dependent Innate Immune Reprogramming.’ *Cell*, 172(1–2), pp. 162-

175.e14. DOI: 10.1016/j.cell.2017.12.013.

7. Scholz, C.J., Biernat, P., Becker, M., Baßler, K., **Günther, P.**, Balfer, J., Dickten, H., Flöer, L., Heikamp, K., Angerer, P., Heilig, M., Karle, R., Köhler, M., Mazurkiewicz, T., Mönnighoff, M., Sauer, C., Schick, A., Schlemm, G., Weigelt, R., Winkler, M., Ulas, T., Theis, F., Huthmacher, S., Kratsch, C. and Schultze, J.L. (2018) 'FASTGenomics: An Analytical Ecosystem for Single-Cell RNA Sequencing Data'. *BioRxiv*, p. 272476. DOI: 10.1101/272476.

8. See, P.*, Dutertre, C.A.*, Chen, J.*, **Günther, P.**, McGovern, N., Irac, S.E., Gunawan, M., Beyer, M., Händler, K., Duan, K., Bin Sumatoh, H.R., Ruffin, N., Jouve, M., Gea-Mallorquí, E., Hennekam, R.C.M., Lim, T., Yip, C.C., Wen, M., Malleret, B., Low, I., Shadan, N.B., Fen, C.F.S., Tay, A., Lum, J., Zolezzi, F., Larbi, A., Poidinger, M., Chan, J.K.Y., Chen, Q., Rénia, L., Haniffa, M., Benaroch, P., Schlitzer, A., Schultze, J.L., Newell, E.W. and Ginhoux, F. (2017) 'Mapping the Human DC Lineage through the Integration of High-Dimensional Techniques'. *Science*, 356(6342), p. eaag3009. DOI: 10.1126/science.aag3009.

9. Beyer, M., Händler, K., **Günther, P.**, Baßler, K., Ulas, T., Becker, M., Klee, K., Schultze, J.L. and Schlitzer, A. (2017) 'Navigating Disease Phenotypes – A Multidimensional Single-Cell Resolution Compass Leads the Way'. *Current Opinion in Systems Biology*, 3, pp. 147–153. DOI: 10.1016/j.coisb.2017.05.004.

10. Sander, J.*, Schmidt, S. V.*, Cirovic, B., McGovern, N., Papantonopoulou, O., Hardt, A.-L., Aschenbrenner, A.C., Kreer, C., Quast, T., Xu, A.M., Schmidleithner, L.M., Theis, H., Thi Huong, L. Do., Sumatoh, H.R. Bin., Lauterbach, M.A.R., Schulte-Schrepping, J., **Günther, P.**, Xue, J., Baßler, K., Ulas, T., Klee, K., Katzmarski, N., Herresthal, S., Krebs, W., Martin, B., Latz, E., Händler, K., Kraut, M., Kolanus, W., Beyer, M., Falk, C.S., Wiegmann, B., Burgdorf, S., Melosh, N.A., Newell, E.W., Ginhoux, F., Schlitzer, A. and Schultze, J.L. (2017) 'Cellular Differentiation of Human Monocytes Is Regulated by Time-Dependent Interleukin-4 Signaling and the Transcriptional

Regulator NCOR2'. *Immunity*, 47(6), pp. 1051-1066.e12. DOI: 10.1016/ j.immuni.2017.11.024.

11. Mass, E.*, Ballesteros, I.*, Farlik, M.*, Halbritter, F.*, **Günther, P.***, Crozet, L., Jacome-Galarza, C.E., Händler, K., Klughammer, J., Kobayashi, Y., Gomez-Perdiguero, E., Schultze, J.L., Beyer, M., Bock, C. and Geissmann, F. (2016) 'Specification of Tissue-Resident Macrophages during Organogenesis.' *Science (New York, N.Y.)*, 353(6304), p. aaf4238. DOI: 10.1126/science. aaf4238.

12. Schmidt, S. V.*, Krebs, W.*, Ulas, T.*, Xue, J., Baßler, K., **Günther, P.**, Hardt, A.-L., Schultze, H., Sander, J., Klee, K., Theis, H., Kraut, M., Beyer, M. and Schultze, J.L. (2016) 'The Transcriptional Regulator Network of Human Inflammatory Macrophages Is Defined by Open Chromatin'. *Cell Research*, 26(2), pp. 151–170. DOI: 10.1038/cr.2016.1.

13. Shahbazi, E., Moradi, S., Nemati, S., Satarian, L., Basiri, M., Gourabi, H., Zare Mehrjardi, N., **Günther, P.**, Lampert, A., Händler, K., Hatay, F.F., Schmidt, D., Molcanyi, M., Hescheler, J., Schultze, J.L., Saric, T. and Baharvand, H. (2016) 'Conversion of Human Fibroblasts to Stably Self-Renewing Neural Stem Cells with a Single Zinc-Finger Transcription Factor.' *Stem Cell Reports*, 6(4), pp. 539–551. DOI: 10.1016/j.stemcr.2016.02.013.

14. Vacchelli, E.*, Ma, Y.*, Baracco, E.E., Sistigu, A., Enot, D.P., Pietrocola, F., Yang, H., Adjemian, S., Chaba, K., Semeraro, M., Signore, M., Ninno, A. De., Lucarini, V., Peschiaroli, F., Businaro, L., Gerardino, A., Manic, G., Ulas, T., **Günther, P.**, Schultze, J.L., Kepp, O., Stoll, G., Lefebvre, C., Mulot, C., Castoldi, F., Rusakiewicz, S., Ladoire, S., Apetoh, L., Pedro, J.M.B.-S., Lucattelli, M., Delarasse, C., Boige, V., Ducreux, M., Delaloge, S., Borg, C., André, F., Schiavoni, G., Vitale, I., Laurent-Puig, P., Mattei, F., Zitvogel, L. and Kroemer, G. (2015) 'Chemotherapy-Induced Antitumor *Immunity* Requires Formyl Peptide Receptor 1'. Science, 350(6263), pp. 972–978. DOI: 10.1126/SCIENCE.AAD0779.

15. Mitschka, S., Ulas, T., Goller, T., Schneider, K., Egert, A., Mertens, J., Brüstle, O., Schorle, H., Beyer, M., Klee, K., Xue, J., **Günther, P.**, Bassler, K., Schultze, J.L. and Kolanus, W. (2015) 'Co-Existence of Intact Stemness and Priming of Neural Differentiation Programs in MES Cells Lacking Trim71'. *Scientific Reports*, 5(1), p. 11126. DOI: 10.1038/srep11126.

* Shared first authorship

# Chapter 1 Literature review

## 1.1 Introduction

A multicellular organism is continuously challenged by a diverse set of agents, including invading pathogens like viruses and bacteria, but also by malignant changes of its tissue or environmental toxins. To adapt to this broad repertoire of challenges, the immune system has developed, so that it cannot only recognize these dangers but also integrate, communicate, and eliminate them (Parkin and Cohen, 2001). The immune system has evolved basic principles ranging from the production of antimicrobials in simple organisms to the existence of a highly complex network of signaling and effector molecules and various immune cells in vertebrates (Buchmann, 2014; Zou *et al.*, 2016; Müller *et al.*, 2018). In addition, this host defense system is not only limited to fighting invading pathogens but also plays crucial roles in sensing non-microbial dangers, regulating physiological processes, orchestrating early development, and maintaining homeostasis.

In order to fulfill this diverse set of functions, the immune system has developed a variety of approaches. One of the primary strategies is the functional compartmentalization of tasks, which has led to the development of specialized immune cell types (Arendt, 2008; Buchmann, 2014; Arendt *et al.*, 2016). These specialized cell types can conduct particular tasks, like sensing and activating other cells, while others are equipped to attack invading pathogens. The cellular heterogeneity of the host defense system is one of the critical aspects that results in the incredible plasticity of the immune system of higher vertebrates.

Another essential strategy is the context-specific behavior of immune cell types that by integration of diverse signals including localization, signaling molecules, microbial substances, and the local microenvironment can respond adequately to the particular situation. This functional plasticity is crucial to balance the responses of immune cells varying between maintaining homeostasis and fighting pathogens. Especially, cells of the myeloid lineage demonstrate a high level of functional plasticity (Buchmann, 2014; Xue *et al.*, 2014; Guilliams *et al.*, 2014; Lavin *et al.*, 2014; Ginhoux *et al.*, 2016; Schultze *et al.*, 2019).

Taken together, cellular heterogeneity and functional plasticity are essential principles of the immune system, but the direct consequences and dynamics of cellular diversity are not yet fully understood. Advances in identifying these mechanisms will improve our knowledge of the regulation of such processes and will significantly benefit our ability to understand and treat diseases.

Of note, recent technological discoveries have paved the way to map cellular diversity and its consequences in a comprehensive fashion. To summarize the latest technological developments that have increased our knowledge about the cellular heterogeneity in myeloid cells, Joachim L. Schultze and I published a review article in *Frontiers in Immunology* (Günther and Schultze, 2019). A copy of the published manuscript is reprinted in the appendix.

## 1.2 Publication

Günther, P. and Schultze, J.L. (2019) ‘Mind the Map: Technology Shapes the Myeloid Cell Space’. *Frontiers in Immunology*, 10, p. 2287. DOI: 10.3389/FIMMU.2019.02287.

Publication I: The manuscript is reprinted in the appendix with permissions from Frontiers Media SA.

## 1.3 Objectives of the thesis

This thesis aimed to investigate the consequences of functional and cellular plasticity in the myeloid cell system. It is crucial to understand the exact mechanisms of myeloid cell plasticity since this cell compartment is involved in almost all major causes of death including cardiovascular diseases (CVDs), stroke, chronic obstructive pulmonary disease (COPD), different types of cancer, diabetes, Alzheimer's disease and infectious diseases. The studies reported in this thesis aim to investigate the cellular composition, mechanisms of cellular differentiation, and signal integration of the myeloid cell compartment.

In this regard, the aims of the studies represented here were:

- to evaluate and improve current strategies for cell-type classification
- to investigate the consequences of Western diet regarding myeloid cells and their hematopoietic precursors
- to define the heterogeneity of the dendritic cell compartment in human peripheral blood
- to improve the current classification of human mononuclear myeloid cells in peripheral blood
- to develop a universal approach to establish consensus maps from scRNA-seq datasets
- to create a model for the early generation of tissue-resident macrophages from precursor states.

## Chapter 2 Methods

In the following chapter, I would like to summarize the critical technologies that have been utilized in this thesis.

A series of discoveries in the areas of microfluidics, flow cytometry and the application of fluorescent dyes or fluorescent-labeled antibodies for biological research (Coulter, 1949; Coons and Kaplan, 1949; Parker and Horst, 1959; Kamentsky *et al.*, 1965; Van Dilla *et al.*, 1969; Hulett *et al.*, 1969; Salzman *et al.*, 1975) have all paved the way for modern cell type discovery. These technologies were further advanced by the ability to sort cells according to their phenotypical characteristics as introduced by the development of the first fluorescent activated cell sorter (FACS) (Hulett *et al.*, 1969; Hulett *et al.*, 1973). Flow cytometry (FC) is used to detect the expression of specific cell type marker genes allowing for classification of heterogeneous cell populations. A primary principle of this is the combination of positive and negative protein markers that creates a well-defined profile of a cell type. However, the specific markers for cell type detection must be defined before the analysis of the sample, which may reduce the potential for finding new cell populations. One of the main advantages of flow cytometry is that the process itself does not require fixation of the cells, therefore enabling the analysis of living specimens. In addition, it permits the sorting of cell types and the further use in experiments to examine the specific functions of these cell types including transcriptional and epigenetic profiling.

The ability to sequence the whole messenger RNA (mRNA) content of a sample by high-throughput RNA-seq represents a key technology that is used in all areas of life sciences (Nagalakshmi *et al.*, 2008; Wang *et al.*, 2009). The process of RNA-seq is initiated by the generation of cDNA from (messenger) RNA, followed by the ligation of DNA sequencing adaptors, and the amplification of the construct by polymerase chain reaction (PCR). The resulting cDNA library is sequenced by next-generation sequencing (NGS) to provide information about the abundance and sequence of RNA species (Ozsolak and Milos, 2011). While the applications of this method are manifold, RNA-seq is primarily used to examine transcriptional changes in a system as a consequence of perturbations (Han *et al.*, 2015). The

strength of transcriptional profiling is the unbiased measurement of thousands of genes within a sample. This does not only allow stratification of cell types by expression of signature genes, but it enables the understanding of deep regulatory circuits within cells. Hence, RNA-seq data is used to understand the transcriptional changes during processes including cell activation, cell differentiation, and cell communication. Notably, the power to model the dynamics of the transcriptional networks and to understand the causal effects of stimulations has made a significant impact on modern life science. However, population-based RNA-seq is limited considering the detection of cellular heterogeneity, since the quantification of population-based RNA represents an average of all measured cells, which may hide bimodal expression of genes.

RNA-seq of individual cells rather than a population of cells was first reported in 2009. Here, RNA-Seq of a small number of cells yielded the first sparse transcriptomes of individual cells. Since then, a multitude of scRNA-seq methods have been reported, which enable parallel profiling of thousands of cells with high sensitivity and low costs. According to the cell isolation strategy, two major single-cell technologies are frequently used including well/ microwell-based and droplet-based methods.

One of the first technologies that incorporated unique molecular identifier (UMI) to increase the precision of RNA molecule quantification allows massively parallel single-cell RNA-sequencing (MARS-Seq) (Jaitin *et al.*, 2014; Jaitin *et al.*, 2015). MARS-seq enables profiling FACS sorted single-cells by using a robotics-assisted protocol. Another well-established protocol is called SMART-Seq2, which creates single-cell libraries by using a template-switching mechanism leading to a robust method with high sensitivity (Picelli *et al.*, 2013; Picelli *et al.*, 2014; Ziegenhain *et al.*, 2017). Since MARS-Seq and SMART-Seq2 are plate-based technologies, their setup allows the combination of protein information recorded during cell sorting with unbiased transcriptome analysis of individual cells, which qualifies them for cell type discovery approaches (Jaitin *et al.*, 2014; Paul *et al.*, 2015; Mass *et al.*, 2016; See *et al.*, 2017; Giladi *et al.*, 2018; Montoro *et al.*, 2018).

One of the biggest challenges of scRNA-seq experiments is the analysis of the sparse and noisy datasets (Ning *et al.*, 2014; Macaulay and Voet, 2014; Stegle *et al.*, 2015; Bacher and Kendziorski, 2016; Vallejos *et al.*, 2017). Although generally speaking the scientific questions may differ, the general outline of data analysis is similar (Baßler *et al.*, 2019). Pre-processing pipelines are designed to perform quality control, read alignment and barcode correction resulting in the compilation of count matrices that contain the number of gene-specific reads or UMIs for every cell respectively (Tian *et al.*, 2018). Dimensionality reduction of high-dimensional data by applying a combination of linear- and non-linear dimensionality reduction methods is a prerequisite for representing the data in a low-dimensional space (Becht *et al.*, 2019). A primary goal of data analysis is the clustering of cells into groups of transcriptionally similar cells, which is followed by the detection of cluster-specific genes by differential expression analysis.

One of the current challenges of scRNA-seq data analysis is the annotation of cell clusters (Stegle *et al.*, 2015; Ilicic *et al.*, 2016). One common approach is to check the expression of specific cell type markers. However, the sparsity of single-cell data may lead to unsatisfying results. Hence, the employment of gene signature scores is a valid tool for cell-type classification (Mass *et al.*, 2016; See *et al.*, 2017; Aran *et al.*, 2019).

The advent of scRNA-seq has led to the release of a multitude of single-cell transcriptome datasets creating detailed maps of cell types from complex tissues. Also, large international consortia aim to produce comprehensive datasets of all tissues of the human body (Rozenblatt-Rosen *et al.*, 2017). To benefit from such efforts, it will be crucial to combine diverse datasets establishing unified consensus maps of cell types.

# Chapter 3 Western diet Triggers NLRP3-Dependent Innate Immune Reprogramming

## 3.1 Introduction

Significant breakthroughs in modern medicine including the use of antibiotics and vaccines, lead to a substantial decrease in deaths by infectious diseases in western societies. However, changes in our lifestyle favored the increase in deaths caused by non-communicable diseases (NCD). Nutrient-dense and high caloric Western diets (WD) promote obesity, which is linked to increased risk for several diseases including type 2 diabetes, CVDs like atherosclerosis and cardiac infarction.

Like most CVDs, the pathogenesis of atherosclerosis is associated with the formation of an inflammatory environment, which is followed by a massive infiltration of monocytes into the tunica intima of arterial walls. Subsequent differentiation of monocytes is a prerequisite for the generation of arterial plaques, a major hallmark of atherosclerosis. An inflammatory environment induces transcriptional and epigenetic reprogramming of monocytes, which facilitates their differentiation into foam cells, further promoting plaque generation. While the direct action of Western diet on these terminal cell types is well studied, it remains elusive whether the effects of WD on monocytes are already introduced earlier during hematopoiesis, e.g. in the progenitors of monocytes termed granulocyte-monocyte progenitor (GMP).

Given the regular challenge by nutritional ingredients, the inflammatory response of monocytes during the pathogenesis of atherosclerosis might be explained by the concept of trained immunity. Here, innate immune cells like monocytes have been shown to react to lipopolysaccharides (LPS) stronger, if they already have had contact with a particular group of stimuli (e.g. Bacille Calmette–Guérin (BCG) vaccine or β-glucan), leading to the formation of an innate immune memory (Netea *et al.*, 2016; Netea *et al.*, 2019). In the outlined context, it can be assumed that innate immune training of GMPs contributes to transcriptional and epigenetic reprogramming of monocytes and hence may enhance the progression of atherosclerosis.

In cooperation with Dr. Anette Christ and Prof. Eicke Latz located at the Institute of Innate Immunity at the University of Bonn, we investigated the consequences of a WD on GMPs and examined whether it can induce innate immune training in GMPs and monocytes.

The results of this study were published in Cell by Christ and Günther et al. (Christ *et al.*, 2018). A copy of the published article is reprinted in the appendix.

## 3.2 Publication

Christ, A.*, **Günther, P.***, Lauterbach, M.A.R., Duewell, P., Biswas, D., Pelka, K., Scholz, C.J., Oosting, M., Haendler, K., Baßler, K., Klee, K., Schulte-Schrepping, J., Ulas, T., Moorlag, S.J.C.F.M., Kumar, V., Park, M.H., Joosten, L.A.B., Groh, L.A., Riksen, N.P., Espevik, T., Schlitzer, A., Li, Y., Fitzgerald, M.L., Netea, M.G., Schultze, J.L. and Latz, E. (2018) 'Western Diet Triggers NLRP3-Dependent Innate Immune Reprogramming.' *Cell*, 172(1–2), pp. 162-175.e14. DOI: 10.1016/j.cell.2017.12.013.

* Shared first authorship

Publication II: The manuscript is reprinted in the appendix with permissions from Cell Press.

## 3.3 Summary

In this study, we investigated the consequences of a Western diet on the hematopoietic system in a murine atherosclerosis model (*Ldlr*$^{-/-}$ mice).

WD induces both systemic inflammation and peripheral priming of specific myeloid cell subsets, which is characterized by increased responsiveness to certain inflammatory stimuli. Remarkably, after setting the mice back on a chow diet (CD), myeloid cells still retained a primed state, clearly suggesting a WD triggered long-term reprogramming of myeloid cells. In addition to a long-lasting reprogramming of myeloid cells, we also observed quantitative changes in the composition of peripheral blood including an increase of red blood cells, monocytes, and granulocytes indicating a stimulation of the myeloid precursors residing in the bone marrow.

By performing high-depth transcriptomics of GMPs, we identified a unique gene signature induced by WD, which is highly enriched for monocyte and granulocyte markers. Indeed, further analysis showed that WD led to a transcriptional lineage bias of GMPs towards monocytes at the expense of granulocytes.

When assessing the activation status of monocytes and granulocytes after a systemic LPS challenge, it became clear that WD causes priming of myeloid cells resulting in a stronger activation phenotype. Similarly, LPS induced a stronger inflammatory signature in bone marrow precursor cells isolated from WD-fed mice, suggesting a WD induced functional reprogramming of myeloid cells and precursor cells.

Strikingly, the analysis of GMPs from mice fed with WD and subsequent CD showed that the response to LPS was still strongly affected. By using gene co-expression networks and transcription factor predictions, we showed that the LPS response in WD/ CD GMPs is more similar to the LPS response of WD GMPs and moreover displays a unique signature. Collectively, these results imply that WD triggers the formation of innate immune memory by long-lasting reprogramming of myeloid precursor cells. To understand the mechanism of this phenomenon, an assay for transposase-accessible chromatin using sequencing

(ATAC-seq) was performed on GMPs after WD, which demonstrated substantial and long-term epigenetic reprogramming of GMPs.

Next, we performed a functional trained immunity (quantitative trait locus) QTL study in human monocytes, which indicated that members of the IL1 pathway are linked to oxidized LDL (ox-LDL) induce trained immunity. To test this hypothesis, we analyzed the response to a WD of the bone marrow compartment from mice lacking NLRP3 (*Ldlr*$^{-/-}$*Nlrp3*$^{-/-}$). In line with the results from the QTL study, NLRP3-deficient mice did not show similar reprogramming towards WD, suggesting NLRP3 as a critical player in the sensing of Western diet-induced sterile inflammation.

# Chapter 4 Mapping the human DC lineage through the integration of high-dimensional techniques

## 4.1 Introduction

In the previous chapter, we showed that myeloid cells show a substantial ability to respond to a variety of stimuli highlighting the extensive plasticity of these cells. One of the main strategies to facilitate this plasticity is the functional compartmentalization into different types of immune cells, including monocytes and DCs.

Initially identified by Ralph Steinmann in the early 1970s, DCs are equipped to be very efficient in antigen presentation to naïve lymphocytes (Steinman and Cohn, 1973; Steinman, 1991; Banchereau and Steinman, 1998). Decades of focused research have shown that DCs are not a homogenous population, but rather consist of different functional entities resulting from variability in their phenotype, tissue localization, and ontogeny (Thomas *et al.*, 1993; O'Doherty *et al.*, 1994; Fanger *et al.*, 1997; Anderson *et al.*, 2000; Lindstedt *et al.*, 2005; Randolph *et al.*, 2008). Human DCs in the blood are identified by using a combination of lineage markers and the receptors HLA-DR and CD11c. Three distinct populations can be identified by using established flow cytometry schemes involving the cell type-specific markers CD141 for cDC1, CD1c for cDC2 and CD123 for plasmacytoid DCs (pDCs). However, flow cytometry heavily relies on already known marker genes impeding the discovery of the full heterogeneity in complex tissues.

One of the technologies that has revolutionized cell type discovery is scRNA-seq, which allows users to perform parallel transcriptomic profiling on thousands of cells. The transcriptome information is used to identify transcriptionally similar populations of cells by unsupervised machine learning and is therefore not relying on *a priori* knowledge. In order to study the human DC compartment in peripheral blood, we applied scRNA-seq to the $Lin^{-}HLA\text{-}DR^{+}CD135^{+}$ fraction of PBMCs. In collaboration with the group of Florent Ginhoux from the Singapore Immunology Network, we published the results of this study in Science (See et al., 2017). A copy of the published manuscript is reprinted in the appendix with permissions of AAAS.

## 4.2 Publication

See, P.*, Dutertre, C.A.*, Chen, J.*, **Günther, P.**, McGovern, N., Irac, S.E., Gunawan, M., Beyer, M., Händler, K., Duan, K., Bin Sumatoh, H.R., Ruffin, N., Jouve, M., Gea-Mallorquí, E., Hennekam, R.C.M., Lim, T., Yip, C.C., Wen, M., Malleret, B., Low, I., Shadan, N.B., Fen, C.F.S., Tay, A., Lum, J., Zolezzi, F., Larbi, A., Poidinger, M., Chan, J.K.Y., Chen, Q., Rénia, L., Haniffa, M., Benaroch, P., Schlitzer, A., Schultze, J.L., Newell, E.W. and Ginhoux, F. (2017) 'Mapping the Human DC Lineage through the Integration of High-Dimensional Techniques'. *Science*, 356(6342), p. eaag3009. DOI: 10.1126/science.aag3009.

* Shared first authorship

Publication III: The manuscript is reprinted in the appendix with permission from AAAS.

## 4.3 Summary

To study the heterogeneity of DCs in an unbiased fashion, peripheral blood cells were sorted into 384-well plates and analyzed by MARS-Seq. Since the number of dendritic cells in peripheral blood is low, sorted cells were enriched by sampling from the $Lin^-$(CD3/14/16/19/20/34)$CD45^+CD135^+$HLA-$DR^+CD123^+$ $CD33^+$ population containing all DC populations. Performing a detailed quality control yielded a scRNA-seq dataset containing transcriptome information of 710 cells.

To reduce the complexity of the gene space, we applied a non-linear dimensionality reduction by t-stochastic neighbor embedding (tSNE), which was used to perform a density-based spatial clustering with noise (DBSCAN) model. tSNE embedding and clustering allowed us to define five transcriptionally distinct clusters within the HLA-$DR^+CD135^+$ cell space. To annotate these clusters, we generated cell-type-specific signatures for pDCs, cDC1s, and cDC2s by analyzing previously published microarray data. Overlay of signature scores onto the tSNE map allowed to identify the three main DC populations in the scRNA-seq dataset. To understand the transcriptional relationship of cell clusters, different trajectory analysis algorithms allowed us to define distinct branches of cells suggesting developmental connections between cells on the same branch. The close relationship to the cDC clusters suggested a mixed phenotype for one of the groups which were further validated by a diverse expression of both pDC and cDC marker genes. Investigation of single-cell RNA-seq data allowed the identification of a population of cells with a close relationship to cDCs and a mixed expression of cDC and pDC marker genes, indicating a progenitor phenotype of these cells. This assumption was further supported by the analysis of the HLA-$DR^+CD135^+$ cell compartment by cytometry by time-of-flight (CyToF).

After establishing the $CD33^+CD123^+$ character of the putative progenitor population, the progenitor capacity was tested by using a stromal culture system. While the pDCs and cDCs mostly stayed within their phenotype, the putative DC progenitor cells contributed to both cDC1s and cDC2s, suggesting their role as circulating preDC. Remarkably, functional assays clearly showed that some of

the functional attributes that have been published for pDCs may be attributed to a contaminating preDC population, which underlines the importance of unbiased discovery approaches by high-dimensional technologies. While the ’pure’ pDC fraction was able to produce interferon-α (IFN-α), the production of IL-12p40 and the ability to stimulate naïve $CD4^+$ T cells in a T cell stimulation assay was clearly attributed to preDCs rather than pDCs. Taken together, these results revealed that these two populations present two transcriptionally, phenotypically and functionally distinct subsets.

High-content scRNA-seq data was generated from human peripheral blood preDCs to investigate their transcriptional heterogeneity. Trajectory analysis and signature score calculation could clearly show a transcriptional stratification of preDCs with cDC1-like and cDC2-like gene expression signatures and a population of unprimed preDCs. With the phenotypic characterization at hand, all preDC and other DC subsets were sorted and analyzed by population-based RNA-seq, which allowed deep transcriptional profiling of these subsets. Bioinformatic analysis revealed DC subset specific signatures for all major subpopulations and besides allowed to visualize the similarity between committed preDC subsets and their progeny. Furthermore, analysis of transcription factors (TF) exposed discrete patterns of cell type-specific TF and their dynamic expression during preDC development.

To understand the ontogeny of these blood-derived preDC populations, human blood, and bone marrow cells were analyzed by CyTOF. Unsupervised dimensionality reduction clearly distinguished between CDPs, pDC, and preDCs in the bone marrow. Interestingly, this analysis suggested that polarization of preDCs into the committed preDC subsets is already established in the bone marrow.

# Chapter 5 Inference of high-resolution single-cell transcriptomic consensus maps illustrated for HLA-DR$^+$ and CD34$^+$ cells from human peripheral and cord blood

## 5.1 Introduction

Single-cell RNA-seq is a powerful technology that has paved the way for categorizing the cellular space in humans and other organisms. Creating a map of all human cell types will generate a reference to detect and fully understand how imbalances of this system can build the basis for disease pathogenesis (Regev *et al.*, 2017). However, to develop unified consensus maps it is crucial to design strategies that allow combining datasets across different individuals, studies, or even different tissues. Furthermore, the wealth of single-cell omics technologies will create diverse sets of data including epigenetic, transcriptomic and protein expression data in part enriched by spatial and temporal information (Neu *et al.*, 2017; Marioni and Arendt, 2017; Packer and Trapnell, 2018). It will be crucial to design strategies to channel the knowledge of these technologies towards creating meaningful, interpretable and comprehensive consensus maps.

Only recently, single-cell technologies including mass cytometry and scRNA-seq, have described an unprecedented heterogeneity in the monocyte/ DC compartment (Breton *et al.*, 2016; Villani *et al.*, 2017; See *et al.*, 2017; Alcántara-Hernández *et al.*, 2017). In our study, we have extended these findings to create the next iteration of the monocyte/ DC space by combining unbiased cellular characterization using scRNA-seq with the assessment of crucial protein markers by flow cytometry assisted index-sorting of major cell populations. The combination of both allows us to connect the results from an unbiased characterization of this cell space to the enormous body of knowledge that has been created in the past decades. To consolidate the monocyte/ DC compartment we utilized a strategy to unambiguously classify these cell types and their subsets. Importantly, we also show that a similar approach can be used to detect rare cell types in large top-down scRNA datasets.

In cooperation with Dr. Branko Cirovic, Kevin Baßler and Prof. Dr. Andreas Schlitzer, we created a new and improved consensus map of the human monocyte and DC cell space and developed a widely applicable strategy to study rare cell types in top-level scRNA-seq datasets. While the paper is under review, we have published the results on the preprint server biorxiv (Günther *et al.*, 2019). A copy of the published article is reprinted in the appendix.

## 5.2 Publication

**Günther, P.*,** Cirovic, B.*, Baßler, K.*, Händler, K., Becker, M., Dutertre, C.A.A., Bigley, V., Newell, E., Collin, M., Ginhoux, F., Schlitzer, A. and Schultze, J.L. (2019) ‘A Rule-Based Data-Informed Cellular Consensus Map of the Human Mononuclear Phagocyte Cell Space’. *BioRxiv*, p. 658179. DOI: 10.1101/658179.

* Shared first authorship

Publication IV: The manuscript is reprinted in the appendix.

## 5.3 Summary

To improve our knowledge on the heterogeneity of human monocytes and DC, we generated a consensus dataset by index-sorting and high-content scRNA-seq. This approach allowed us to compare recently published legacy maps and to form a consensus dataset of the HLA-DR$^+$ cell compartment in human peripheral blood, mainly containing monocyte and DC subpopulations. We showed that the human DC compartment in peripheral blood consists of one type of cDC1, two transcriptionally distinct types of cDC2s (DC2 and DC3), one pDC population, and an AXL$^+$SIGLEC6$^+$ preDC population. To unambiguously define the human DC map, we identified the nature of the newly suggested DC4 as a mixed population of mainly preDCs and CD16+ monocytes. Next, we re-established the structure of three main populations of human peripheral blood monocytes, by identifying a contaminating CD56$^{dim}$ natural killer (NK) cell population as these cells had been suggested as a fourth monocyte subset (Villani *et al.*, 2017).

To show the wide applicability of our approach, we successfully applied the developed strategy to identify rare cell populations in top-level datasets. This enabled us to stratify human peripheral blood CD34$^+$ hematopoietic stem cells (HSC). As a result, while CD34$^+$ cells in adult peripheral blood revealed strong erythroid-lineage priming, our data suggest that CD34$^+$ cells in cord blood exhibit a multi-lineage primed state.

Taken together, we have developed a versatile approach to study and improve cellular maps that can be applied to other cell types and tissues. By employing this approach, we could not only expand the human monocyte and DC cell atlas but also define the lineage priming of human CD34$^+$ cells in adult peripheral and cord blood.

## Chapter 6 Overall Discussion

Cells of the myeloid cell lineage have crucial roles in regulating physiological processes, keeping tissue homeostasis, and representing the first line of protection against a multitude of dangers. To accomplish this high degree of functional diversity myeloid cells are characterized by a high level of plasticity as a consequence of a highly regulated integration of diverse signals (see Figure 1). Myeloid cells have evolved into a complex system of different immune cell types, each specialized in a range of tasks. This diversity of myeloid cells is extended by the ability to quickly respond to a wide range of signals, further increasing their functional plasticity. Both cellular and functional plasticity of myeloid cells are a strong focus of current research efforts in immunology because there is still a significant need to understand the mechanisms and effects of myeloid cell heterogeneity. The specification of cellular and functional plasticity is the result of a multitude of molecular decisions. Here, every signal a cell perceives is integrated with the current state of interconnected networks like the transcriptome, epigenome, metabolome, and proteome (Figure 1). Of note, these interconnected networks build an intrinsic regulatory memory as they reflect the result of earlier molecular decisions.

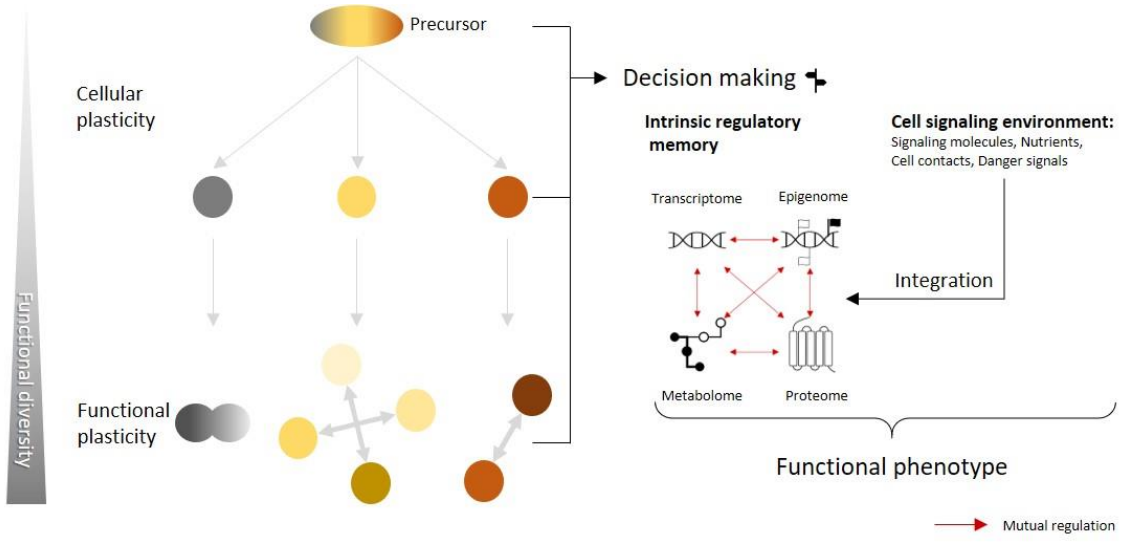



*Figure 1: Strategies of functional diversification and the regulatory decision making.*

The work that has been performed in the scope of this dissertation aimed to advance our understanding of causes, regulation, and consequences of context-specific plasticity of myeloid cells. In this regard, the response of the mononuclear myeloid cell compartment to a Western diet was studied in an atherosclerosis mouse model. Both peripheral and splenic monocytes showed evident signs of inflammation when isolated from WD-fed animals (Christ *et al.*, 2018). Furthermore, WD induced a substantial increase of total erythrocyte, granulocyte, and monocyte numbers in peripheral blood, suggesting the involvement of myeloid cell precursors. Similar to this study, Béliard et al. reported that WD or its nutritional components induced myelopoiesis in mice and led to an activated phenotype of peripheral blood monocytes (Béliard *et al.*, 2017). Of note, in humans there is a clear correlation of body fat with the number of $CD16^{+}$ nonclassical monocytes. Furthermore, the decrease of intermediate monocytes is associated with a reduction of subclinical atherosclerosis (Poitou *et al.*, 2011). Clearly, the induction of myelopoiesis suggested an involvement of the hematopoietic compartment (Singer *et al.*, 2014).

The effect of WD can be partially explained by the results of a transcriptome-wide analysis of GMPs showing a strong inflammatory activation signature of these myeloid cell precursors (Christ *et al.*, 2018). Strikingly, this inflammatory priming of GMPs persisted after resting the mice on CD, suggesting a long-term reprogramming of the hematopoietic progenitor cells. Here, WD leads to the formation of a lasting innate immune memory (Netea *et al.*, 2016; Netea *et al.*, 2019), which might be the reason for an enhanced inflammatory response of peripheral monocytes in atherosclerosis and other pathologies. Recently, several studies demonstrated the relationship between diet-induced myelopoiesis and epigenetic regulation of peripheral monocytes and hematopoietic progenitors (van Kampen *et al.*, 2014; Short *et al.*, 2017). However, these studies were limited to the epigenetic analysis of selected TFs. To investigate the global changes in the epigenome, we performed a genome-wide assessment of open chromatin by ATAC-seq. WD induced global changes in the chromatin accessibility landscape leading to lasting transcriptional consequences for GMPs. Notably, one of the reported genes made accessible by WD is the dioxygenase *Tet2*, recently directly

associated with an anti-atherogenic effect (Jaiswal *et al.*, 2017). It might be part of prospective studies to clarify whether the increased accessibility negatively regulates the expression of TET2. Besides, *Tet2* is a very well-studied tumor suppressor gene as somatic mutations are frequently observed in myeloid neoplasms including MDS, CMML, and AML (Ko *et al.*, 2010) and mutations are associated with poor prognosis (Metzeler *et al.*, 2011). This is very interesting in regard to the reported association of overweight and obesity with an increased incidence of AML (Li *et al.*, 2017) and poor prognosis for obese patients (Orgel *et al.*, 2016), which is in line with the role of chronic inflammation in tumorigenesis (Elinav *et al.*, 2013). Taken together, these results show that WD leads to an inflammatory environment, which has direct effects on peripheral blood cells but also leads to a long-term transcriptional and epigenetic reprogramming of myeloid precursor cells.

Chronic low-grade inflammation induced by Western diet is a significant risk factor for several diseases, including obesity, type 2 diabetes, chronic liver disease, cardiovascular diseases (Christ and Latz, 2019). Our study shows that WD leads to lasting changes in the myeloid cell precursor compartment in the bone marrow, which may provide additional means for therapeutic intervention in different pathologies. Importantly, we could identify NLRP3 as a critical sensor of WD induced inflammation (Christ *et al.*, 2018), which offers a potential therapeutic target for chronic low-grade inflammation. To design effective therapeutics, a better understanding of the direct mechanisms of how WD induces this low-grade inflammation is needed. Here, it will be crucial to clarify the exact stimulus of NLRP3 and weather NLRP3 senses this inflammation on the level of the bone marrow compartment or already in the gastrointestinal tract. Remarkably, several studies suggest that WD triggers changes of the gut microbiome leading to the influx and activation of myeloid cells (David *et al.*, 2014; Wilson *et al.*, 2018). WD induced trained immunity in the bone marrow compartment might contribute to this strong inflammatory response.

The lineage-priming of GMPs is heavily influenced by WD as shown by the transcriptional priming towards monocytes at the expense of the granulocyte

signature (Christ *et al.*, 2018). Lifestyle-induced changes of the cellular composition in peripheral blood in mice and humans are clearly associated with onset, progression and poor therapeutic success of various diseases including cancer, type 2 diabetes and atherosclerosis (Jaiswal *et al.*, 2017; Short *et al.*, 2017; Li *et al.*, 2017; Christ and Latz, 2019). Hence, it is crucial to understand the dynamics of changes in the peripheral blood composition in order to design effective therapeutic strategies.

To investigate the cellular heterogeneity of myeloid cells belonging to the dendritic cell compartment in human peripheral blood, high-throughput single-cell RNA-seq of sorted DCs from a healthy donor was performed (See *et al.*, 2017). In addition to the expected major DC populations, including pDCs, cDC1, and cDC2, we also detected another group of cells, namely predendritic cells (preDC), sharing transcriptional similarities to both pDC and cDCs underlining the necessity of single-cell RNA-seq for unbiased cell-type discovery.

In differentiation assays, preDCs generate cells of the cDC subsets but fail to produce pDCs, which suggests that this cell type is a progenitor of classical dendritic cells, previously identified in mice (Naik *et al.*, 2007). By using a combined strategy of single-cell RNA-seq and CyToF, transcriptional and phenotypic profiles of preDCs in human peripheral blood were established leading to a new sorting strategy for human peripheral blood DCs (See *et al.*, 2017). Updating the way of how DCs are defined is crucial, as functional assays showed that several functions previously attributed to pDCs, e.g. IL12 production or T cell stimulatory capacity (Cella *et al.*, 1999a; Matsui *et al.*, 2009), are a result of contaminating preDCs in the pDC fraction. The lineage-priming of preDCs towards both classical DC subsets is already eminent in the bone marrow extending previously reported lineage-priming in mice and humans (Breton *et al.*, 2015; Schlitzer *et al.*, 2015; Breton *et al.*, 2016). Collectively, a high-resolution map of the dendritic cell space in human peripheral blood was established, creating a reference for further investigation of changes in DC subset composition in diseases.

Our study emphasizes the importance of unbiased high-dimensional approaches for cell type discovery. Interestingly, Villani and colleagues (2017) performed a similar study investigating the heterogeneity of the monocyte and DC compartment in human peripheral blood by single-cell RNA-seq. When comparing both studies it became apparent that the suggested maps of human DCs differed significantly in the detected types of DC. Therefore, we developed a strategy to unify the results of both studies and created a universal approach to harmonize single-cell RNA-seq datasets. Application of this strategy led to the improvement of the human consensus map of the DC and monocyte cell space. Our results clarified (i) the nature and phenotype of a suggested DC4 population (Villani *et al.*, 2017) as mixture of non-classical monocytes and preDCs, (ii) identified the Mono4 (Villani *et al.*, 2017) population as contaminating $CD56^{dim}$ NK cells and (iii) showed the transcriptional and phenotypical similarity of DC5/ AS-DCs and preDCs, which collectively cumulated in the harmonization of the human DC and monocyte space (Günther *et al.*, 2019).

Next, we applied the same approach to different datasets showing the broad applicability of our strategy. For instance, integration of a high-content map and a top-level map from human blood $CD34^+$ cells were used to stratify the human peripheral blood HSC continuum. While $CD34^+$ cells from adult human peripheral blood primarily showed a lineage-priming for the erythroid lineage, $CD34^+$ cells from human cord blood showed a more unprimed signature (Günther *et al.*, 2019). Our approach may represent an important strategy to consolidate cellular maps generated by large international consortia aiming to create cell atlases of different tissues during health and disease (Rozenblatt-Rosen *et al.*, 2017).

Our efforts to map the human DC and monocyte cell space (See *et al.*, 2017; Günther *et al.*, 2019) will serve as a reference to detect imbalances of mononuclear myeloid cells in several diseases including atherosclerosis and chronic inflammatory diseases. However, further efforts are required to identify the relationship of blood monocytes and DCs with their counterparts in solid tissues including skin, lymph nodes, lung, spleen, etc. Furthermore, future studies need to investigate the assignment of pDCs to the DC lineage. Their mixed

phenotype, but close relationship to DCs impeded an assignment to a cell lineage (Lennert and Remmele, 1958; Facchetti *et al.*, 1988; Chehimi *et al.*, 1989; Siegal *et al.*, 1999; Cella *et al.*, 1999b). Although pDCs display an inefficient antigen presentation capacity, they were classified as dendritic cells (Banchereau *et al.*, 2000). A new study in mice by Dress *et al.* (2019) reported that priming of progenitors towards the pDCs lineage is already evident in common lymphoid progenitors (CLP), suggesting an early separation from the cDC lineage. However, further information is required to unambiguously identify the cell lineage of pDCs (Günther and Schultze, 2019).

Another important observation from our study is the mixed potential of preDCs that is already acquired in the bone marrow compartment and still present in peripheral blood (See *et al.*, 2017; Günther *et al.*, 2019). This transcriptional priming implies that a part of the preDCs is already committed to either the cDC1 or cDC2 lineage when entering the tissue. This phenomenon is in line with a previous report on the limited effects of the tissue microenvironment on reprogramming of DCs, suggesting a minor influence of signals from the local environment (Heidkamp *et al.*, 2016). In contrast to DCs, monocytes and macrophages are more strongly influenced by the microenvironment as outlined in various studies (Xue *et al.*, 2014; Lavin *et al.*, 2014; Schmidt *et al.*, 2016). This implication is also observable in the ontogeny of tissue-resident macrophages. Single-cell RNA-seq was used to investigate the generation of tissue-resident macrophages from erythroid-myeloid progenitors (EMP) in early murine development (Mass *et al.*, 2016). The analysis of scRNA-seq data contributed to formulating a model of specification of tissue-resident macrophages. Here, during early development EMPs are generated in the yolk sac (YS) progressively differentiating towards tissue-resident macrophages. Following, an intermediate population termed preMac is committed to the macrophage lineage as these cells already acquire a key macrophage core signature and migrate to the developing tissues in a CXCR3-dependent fashion (Mass *et al.*, 2016). Once, the preMacs have reached the organs; the tissue microenvironment favors maturation of preMacs into tissue-resident macrophages, equipped to perform tissue-specific functions (Paolicelli *et al.*, 2011; Haldar *et al.*, 2014; Okabe and Medzhitov, 2014).

In conclusion, the work that has been carried out in the scope of this thesis has significantly enhanced our understanding of functional heterogeneity in the myeloid cell system. The mechanisms reported in the studies included in this thesis will help further investigating the roles of myeloid cells in various disease, including obesity, CVDs, cancer, and autoimmune diseases.

## References

References contain publications cited in the main text of the thesis. Citations used within the re-printed publications are listed in the respective manuscripts.

Alcántara-Hernández, M., Leylek, R., Wagar, L.E., Engleman, E.G., Keler, T., Marinkovich, M.P., Davis, M.M., Nolan, G.P. and Idoyaga, J. (2017) 'High-Dimensional Phenotypic Mapping of Human Dendritic Cells Reveals Interindividual Variation and Tissue Specialization.' *Immunity*, 47(6), pp. 1037-1050.e6. DOI: 10.1016/j.immuni.2017.11.001.

Anderson, K.L., Perkin, H., Surh, C.D., Venturini, S., Maki, R.A. and Torbett, B.E. (2000) 'Transcription Factor PU.1 Is Necessary for Development of Thymic and Myeloid Progenitor-Derived Dendritic Cells'. *The Journal of Immunology*, 164(4), pp. 1855–1861. DOI: 10.4049/jimmunol.164.4.1855.

Aran, D., Looney, A.P.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P.P., Wolters, P.J.J., Abate, A.R.R., Butte, A.J.J. and Bhattacharya, M. (2019) 'Reference-Based Analysis of Lung Single-Cell Sequencing Reveals a Transitional Profibrotic Macrophage'. *Nature Immunology*, 20(2), pp. 163–172. DOI: 10.1038/s41590-018-0276-y.

Arendt, D. (2008) 'The Evolution of Cell Types in Animals: Emerging Principles from Molecular Studies'. *Nature Reviews Genetics*, 9(11), pp. 868–882. DOI: 10.1038/nrg2416.

Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D. and Wagner, G.P. (2016) 'The Origin and Evolution of Cell Types'. *Nature Reviews Genetics*, 17(12), pp. 744–757. DOI: 10.1038/nrg.2016.127.

Bacher, R. and Kendziorski, C. (2016) 'Design and Computational Analysis of Single-Cell RNA-Sequencing Experiments'. *Genome Biology*, 17(1), p. 63. DOI: 10.1186/s13059-016-0927-y.

Banchereau, J., Briere, F., Caux, C., Davoust, J., Lebecque, S., Liu, Y.-J.,

Pulendran, B. and Palucka, K. (2000) ‘Immunobiology of Dendritic Cells’. *Annual Review of Immunology*, 18(1), pp. 767–811. DOI: 10.1146/annurev.immunol.18.1.767.

Banchereau, J. and Steinman, R.M. (1998) ‘Dendritic Cells and the Control of Immunity.’ *Nature*, 392(6673), pp. 245–252.

Baßler, K., Günther, P., Schulte-Schrepping, J., Becker, M. and Biernat, P. (2019) ‘A Bioinformatic Toolkit for Single-Cell MRNA Analysis’. In Proserpio, V. (ed.) *Single Cell Methods: Sequencing and Proteomics, Methods in Molecular Biology*. Springer Nature, New York, NY, pp. 433–455. DOI: 10.1007/978-1-4939-9240-9_26.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F. and Newell, E.W. (2019) ‘Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP’. *Nature Biotechnology*, 37(1), pp. 38–44. DOI: 10.1038/nbt.4314.

Béliard, S., Le Goff, W., Saint-Charles, F., Poupel, L., Deswaerte, V., Bouchareychas, L., Huby, T. and Lesnik, P. (2017) ‘Modulation of Gr1low Monocyte Subset Impacts Insulin Sensitivity and Weight Gain upon High-Fat Diet in Female Mice’. *International Journal of Obesity*, 41(12), pp. 1805–1814. DOI: 10.1038/ijo.2017.179.

Breton, G., Lee, J., Zhou, Y.J., Schreiber, J.J., Keler, T., Puhr, S., Anandasabapathy, N., Schlesinger, S., Caskey, M., Liu, K. and Nussenzweig, M.C. (2015) ‘Circulating Precursors of Human CD1c+ and CD141+ Dendritic Cells.’ *The Journal of Experimental Medicine*, 212(3), pp. 401–13. DOI: 10.1084/jem.20141441.

Breton, G., Zheng, S., Valieris, R., Tojal da Silva, I., Satija, R. and Nussenzweig, M.C. (2016) ‘Human Dendritic Cells (DCs) Are Derived from Distinct Circulating Precursors That Are Precommitted to Become CD1c+ or CD141+ DCs.’ *The Journal of Experimental Medicine*, 213(13), pp. 2861–2870. DOI: 10.1084/jem.20161135.

Buchmann, K. (2014) ‘Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals.’ *Frontiers in Immunology*, 5, p. 459. DOI: 10.3389/fimmu.2014.00459.

Cella, M., Jarrossay, D., Facchetti, F., Alebardi, O., Nakajima, H., Lanzavecchia, A. and Colonna, M. (1999a) ‘Plasmacytoid Monocytes Migrate to Inflamed Lymph Nodes and Produce Large Amounts of Type I Interferon’. *Nature Medicine*, 5(8), pp. 919–923. DOI: 10.1038/11360.

Cella, M., Jarrossay, D., Facchetti, F., Alebardi, O., Nakajima, H., Lanzavecchia, A. and Colonna, M. (1999b) ‘Plasmacytoid Monocytes Migrate to Inflamed Lymph Nodes and Produce Large Amounts of Type I Interferon’. *Nature Medicine*, 5(8), pp. 919–923. DOI: 10.1038/11360.

Chehimi, J., Starr, S.E., Kawashima, H., Miller, D.S., Trinchieri, G., Perussia, B. and Bandyopadhyay, S. (1989) ‘Dendritic Cells and IFN-Alpha-Producing Cells Are Two Functionally Distinct Non-B, Non-Monocytic HLA-DR+ Cell Subsets in Human Peripheral Blood.’ *Immunology*, 68(4), pp. 486–90.

Christ, A., Günther, P., Lauterbach, M.A.R., Duewell, P., Biswas, D., Pelka, K., Scholz, C.J., Oosting, M., Haendler, K., Baßler, K., Klee, K., Schulte-Schrepping, J., Ulas, T., Moorlag, S.J.C.F.M., Kumar, V., Park, M.H., Joosten, L.A.B., Groh, L.A., Riksen, N.P., Espevik, T., Schlitzer, A., Li, Y., Fitzgerald, M.L., Netea, M.G., Schultze, J.L. and Latz, E. (2018) ‘Western Diet Triggers NLRP3-Dependent Innate Immune Reprogramming.’ *Cell*, 172(1–2), pp. 162-175.e14. DOI: 10.1016/j.cell.2017.12.013.

Christ, A. and Latz, E. (2019) ‘The Western Lifestyle Has Lasting Effects on Metaflammation’. *Nature Reviews Immunology*, 19(5), pp. 267–268. DOI: 10.1038/s41577-019-0156-1.

Coons, A.H. and Kaplan, M.H. (1949) ‘Localization of Antigenin Tissue Cells: II. Improvements in a Method for the Detection of Antigen by Means of Fluorescent Antibody’. *Journal of Experimental Medicine*, 91(1), pp. 1–13. DOI: 10.1084/jem.91.1.1.

Coulter, W.H. (1949) (2,656,508) ‘Means for Counting Particles Suspended in a Fluid.’ Available at: https://patents.google.com/patent/US2656508A/en (Accessed: 13 September 2019).

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A. V., Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J. and Turnbaugh, P.J. (2014) ‘Diet Rapidly and Reproducibly Alters the Human Gut Microbiome’. *Nature*, 505(7484), pp. 559–563. DOI: 10.1038/nature12820.

Van Dilla, M.A., Truiullo, T.T., Mullaney, P.F. and Coultex, J.R. (1969) ‘Cell Microfluorometry: A Method for Rapid Fluorescence Measurement’. *Science*, 163(3872), pp. 1213–1214. DOI: 10.1126/science.163.3872.1213.

Dress, R.J., Dutertre, C.-A., Giladi, A., Schlitzer, A., Low, I., Shadan, N.B., Tay, A., Lum, J., Kairi, M.F.B.M., Hwang, Y.Y., Becht, E., Cheng, Y., Chevrier, M., Larbi, A., Newell, E.W., Amit, I., Chen, J. and Ginhoux, F. (2019) ‘Plasmacytoid Dendritic Cells Develop from Ly6D+ Lymphoid Progenitors Distinct from the Myeloid Lineage’. *Nature Immunology*, 20(7), pp. 852–864. DOI: 10.1038/s41590-019-0420-3.

Elinav, E., Nowarski, R., Thaiss, C.A., Hu, B., Jin, C. and Flavell, R.A. (2013) ‘Inflammation-Induced Cancer: Crosstalk between Tumours, Immune Cells and Microorganisms’. *Nature Reviews Cancer*, 13(11), pp. 759–771. DOI: 10.1038/nrc3611.

Facchetti, F., de Wolf-Peeters, C., Mason, D.Y., Pulford, K., van den Oord, J.J. and Desmet, V.J. (1988) ‘Plasmacytoid T Cells. Immunohistochemical Evidence for Their Monocyte/Macrophage Origin.’ *The American Journal of Pathology*, 133(1), pp. 15–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3052093 (Accessed: 18 September 2019).

Fanger, N.A., Voigtlaender, D., Liu, C., Swink, S., Wardwell, K., Fisher, J., Graziano, R.F., Pfefferkorn, L.C., Guyre, P.M., Valladeau, J., Saeland, S., Fukuhara, S. and Ikehara, S. (1997) ‘Characterization of Expression, Cytokine Regulation, and Effector Function of the High Affinity IgG Receptor Fc Gamma

RI (CD64) Expressed on Human Blood Dendritic Cells.' *Journal of Immunology (Baltimore, Md.: 1950)*, 158(7), pp. 3090–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9120261 (Accessed: 29 June 2019).

Giladi, A., Paul, F., Herzog, Y., Lubling, Y., Weiner, A., Yofe, I., Jaitin, D., Cabezas-Wallscheid, N., Dress, R., Ginhoux, F., Trumpp, A., Tanay, A. and Amit, I. (2018) 'Single-Cell Characterization of Haematopoietic Progenitors and Their Trajectories in Homeostasis and Perturbed Haematopoiesis'. *Nature Cell Biology*, 20(7), pp. 836–846. DOI: 10.1038/s41556-018-0121-4.

Ginhoux, F., Schultze, J.L., Murray, P.J., Ochando, J. and Biswas, S.K. (2016) 'New Insights into the Multidimensional Concept of Macrophage Ontogeny, Activation and Function'. *Nature Immunology*, 17(1), pp. 34–40. DOI: 10.1038/ni.3324.

Guilliams, M., Ginhoux, F., Jakubzick, C., Naik, S.H., Onai, N., Schraml, B.U., Segura, E., Tussiwand, R. and Yona, S. (2014) 'Dendritic Cells, Monocytes and Macrophages: A Unified Nomenclature Based on Ontogeny.' *Nature Reviews. Immunology*, 14(8), pp. 571–8. DOI: 10.1038/nri3712.

Günther, P., Cirovic, B., Baßler, K., Händler, K., Becker, M., Dutertre, C.A.A., Bigley, V., Newell, E., Collin, M., Ginhoux, F., Schlitzer, A. and Schultze, J.L. (2019) 'A Rule-Based Data-Informed Cellular Consensus Map of the Human Mononuclear Phagocyte Cell Space'. *BioRxiv*, p. 658179. DOI: 10.1101/658179.

Günther, P. and Schultze, J.L. (2019) 'Mind the Map: Technology Shapes the Myeloid Cell Space'. *Frontiers in Immunology*, 10, p. 2287. DOI: 10.3389/FIMMU.2019.02287.

Haldar, M., Kohyama, M., So, A.Y.-L., Kc, W., Wu, X., Briseño, C.G., Satpathy, A.T., Kretzer, N.M., Arase, H., Rajasekaran, N.S., Wang, L., Egawa, T., Igarashi, K., Baltimore, D., Murphy, T.L. and Murphy, K.M. (2014) 'Heme-Mediated SPI-C Induction Promotes Monocyte Differentiation into Iron-Recycling Macrophages.' *Cell*, 156(6), pp. 1223–1234. DOI: 10.1016/j.cell.2014.01.069.

Han, Y., Gao, S., Muegge, K., Zhang, W. and Zhou, B. (2015) 'Advanced

Applications of RNA Sequencing and Challenges'. *Bioinformatics and Biology Insights*, 9s1, p. BBI.S28991. DOI: 10.4137/BBI.S28991.

Heidkamp, G.F., Sander, J., Lehmann, C.H.K., Heger, L., Eissing, N., Baranska, A., Lu hr, J.J., Hoffmann, A., Reimer, K.C., Lux, A., So der, S., Hartmann, A., Zenk, J., Ulas, T., McGovern, N., Alexiou, C., Spriewald, B., Mackensen, A., Schuler, G., Schauf, B., Forster, A., Repp, R., Fasching, P.A., Purbojo, A., Cesnjevar, R., Ullrich, E., Ginhoux, F., Schlitzer, A., Nimmerjahn, F., Schultze, J.L. and Dudziak, D. (2016) 'Human Lymphoid Organ Dendritic Cell Identity Is Predominantly Dictated by Ontogeny, Not Tissue Microenvironment'. *Science Immunology*, 1(6), pp. eaai7677–eaai7677. DOI: 10.1126/sciimmunol.aai7677.

Hulett, H.R., Bonner, W.A., Barrett, J. and Herzenberg, L.A. (1969) 'Cell Sorting: Automated Separation of Mammalian Cells as a Function of Intracellular Fluorescence'. *Science*, 166(3906), pp. 747–749. DOI: 10.1126/science.166.3906.747.

Hulett, H.R., Bonner, W.A., Sweet, R.G. and Herzenberg, L.A. (1973) 'Development and Application of a Rapid Cell Sorter'. *Clinical Chemistry*, 19(8). Available at: http://clinchem.aaccjnls.org/content/19/8/813 (Accessed: 13 September 2019).

Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C. and Teichmann, S.A. (2016) 'Classification of Low Quality Cells from Single-Cell RNA-Seq Data'. *Genome Biology*, 17(1), p. 29. DOI: 10.1186/s13059-016-0888-1.

Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., Baber, U., Mehran, R., Fuster, V., Danesh, J., Frossard, P., Saleheen, D., Melander, O., Sukhova, G.K., Neuberg, D., Libby, P., Kathiresan, S. and Ebert, B.L. (2017) 'Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease'. *New England Journal of Medicine*, 377(2), pp. 111–121. DOI: 10.1056/NEJMoa1701719.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I.,

Mildner, A., Cohen, N., Jung, S., Tanay, A. and Amit, I. (2014) 'Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types'. *Science*, 343(6172), pp. 776–9. DOI: 10.1126/science.1247651.

Jaitin, D.A., Keren-Shaul, H., Elefant, N. and Amit, I. (2015) 'Each Cell Counts: Hematopoiesis and Immunity Research in the Era of Single Cell Genomics'. *Seminars in Immunology*, 27(1), pp. 67–71. DOI: 10.1016/j.smim.2015.01.002.

Kamentsky, L.A., Melamed, M.R. and Derman, H. (1965) 'Spectrophotometer: New Instrument for Ultrarapid Cell Analysis'. *Science*, 150(3696), pp. 630–631. DOI: 10.1126/science.150.3696.630.

van Kampen, E., Jaminon, A., van Berkel, T.J.C. and Van Eck, M. (2014) 'Diet-Induced (Epigenetic) Changes in Bone Marrow Augment Atherosclerosis'. *Journal of Leukocyte Biology*, 96(5), pp. 833–841. DOI: 10.1189/jlb.1A0114-017R.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., Liu, X.S., Aravind, L., Agarwal, S., Maciejewski, J.P. and Rao, A. (2010) 'Impaired Hydroxylation of 5-Methylcytosine in Myeloid Cancers with Mutant TET2'. *Nature*, 468(7325), pp. 839–843. DOI: 10.1038/nature09586.

Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S. and Amit, I. (2014) 'Tissue-Resident Macrophage Enhancer Landscapes Are Shaped by the Local Microenvironment.' *Cell*, 159(6), pp. 1312–26. DOI: 10.1016/j.cell.2014.11.018.

Lennert, K. and Remmele, W. (1958) 'Karyometrische Untersuchungen an Lymphknotenzellen Des Menschen'. *Acta Haematologica*, 19(2), pp. 99–113. DOI: 10.1159/000205419.

Li, S., Chen, L., Jin, W., Ma, X., Ma, Y., Dong, F., Zhu, H., Li, J. and Wang, K. (2017) 'Influence of Body Mass Index on Incidence and Prognosis of Acute Myeloid Leukemia and Acute Promyelocytic Leukemia: A Meta-Analysis'. *Scientific Reports*, 7(1), p. 17998. DOI: 10.1038/s41598-017-18278-x.

Lindstedt, M., Lundberg, K. and Borrebaeck, C.A.K. (2005) ‘Gene Family Clustering Identifies Functionally Associated Subsets of Human in Vivo Blood and Tonsillar Dendritic Cells.’ *Journal of Immunology (Baltimore, Md. : 1950)*, 175(8), pp. 4839–46. DOI: 10.4049/jimmunol.175.8.4839.

Macaulay, I.C. and Voet, T. (2014) ‘Single Cell Genomics: Advances and Future Perspectives.’ *PLoS Genetics*, 10(1), p. e1004126. DOI: 10.1371/journal.pgen.1004126.

Marioni, J.C. and Arendt, D. (2017) ‘How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology’. *Annual Review of Cell and Developmental Biology*, 33(1), pp. 537–553. DOI: 10.1146/annurev-cellbio-100616-060818.

Mass, E., Ballesteros, I., Farlik, M., Halbritter, F., Günther, P., Crozet, L., Jacome-Galarza, C.E., Händler, K., Klughammer, J., Kobayashi, Y., Gomez-Perdiguero, E., Schultze, J.L., Beyer, M., Bock, C. and Geissmann, F. (2016) ‘Specification of Tissue-Resident Macrophages during Organogenesis.’ *Science (New York, N.Y.)*, 353(6304), p. aaf4238. DOI: 10.1126/science.aaf4238.

Matsui, T., Connolly, J.E., Michnevitz, M., Chaussabel, D., Yu, C.-I., Glaser, C., Tindle, S., Pypaert, M., Freitas, H., Piqueras, B., Banchereau, J. and Palucka, A.K. (2009) ‘CD2 Distinguishes Two Subsets of Human Plasmacytoid Dendritic Cells with Distinct Phenotype and Functions.’ *Journal of Immunology (Baltimore, Md. : 1950)*, 182(11), pp. 6815–23. DOI: 10.4049/jimmunol.0802008.

Metzeler, K.H., Maharry, K., Radmacher, M.D., Mrózek, K., Margeson, D., Becker, H., Curfman, J., Holland, K.B., Schwind, S., Whitman, S.P., Wu, Y.-Z., Blum, W., Powell, B.L., Carter, T.H., Wetzler, M., Moore, J.O., Kolitz, J.E., Baer, M.R., Carroll, A.J., Larson, R.A., Caligiuri, M.A., Marcucci, G. and Bloomfield, C.D. (2011) ‘TET2 Mutations Improve the New European LeukemiaNet Risk Classification of Acute Myeloid Leukemia: A Cancer and Leukemia Group B Study.’ *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, 29(10), pp. 1373–81. DOI: 10.1200/JCO.2010.32.7742.

Montoro, D.T., Haber, A.L., Biton, M., Vinarsky, V., Lin, B., Birket, S.E., Yuan, F.,

Chen, S., Leung, H.M., Villoria, J., Rogel, N., Burgin, G., Tsankov, A.M., Waghray, A., Slyper, M., Waldman, J., Nguyen, L., Dionne, D., Rozenblatt-Rosen, O., Tata, P.R., Mou, H., Shivaraju, M., Bihler, H., Mense, M., Tearney, G.J., Rowe, S.M., Engelhardt, J.F., Regev, A. and Rajagopal, J. (2018) 'A Revised Airway Epithelial Hierarchy Includes CFTR-Expressing Ionocytes'. *Nature*, 560(7718), pp. 319–324. DOI: 10.1038/s41586-018-0393-7.

Müller, V., de Boer, R.J., Bonhoeffer, S. and Szathmáry, E. (2018) 'An Evolutionary Perspective on the Systems of Adaptive Immunity'. *Biological Reviews*, 93(1), pp. 505–528. DOI: 10.1111/brv.12355.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008) 'The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing'. *Science* , 320(5881), pp. 1344–1349. DOI: 10.1126/science.1158441.

Naik, S.H., Sathe, P., Park, H.-Y., Metcalf, D., Proietto, A.I., Dakic, A., Carotta, S., O'Keeffe, M., Bahlo, M., Papenfuss, A., Kwak, J.-Y., Wu, L. and Shortman, K. (2007) 'Development of Plasmacytoid and Conventional Dendritic Cell Subtypes from Single Precursor Cells Derived in Vitro and in Vivo.' *Nature Immunology*, 8(11), pp. 1217–26. DOI: 10.1038/ni1522.

Netea, M.G., Joosten, L.A.B., Latz, E., Mills, K.H.G., Natoli, G., Stunnenberg, H.G., ONeill, L.A.J. and Xavier, R.J. (2016) 'Trained Immunity: A Program of Innate Immune Memory in Health and Disease'. *Science*, 352(6284), pp. aaf1098–aaf1098. DOI: 10.1126/science.aaf1098.

Netea, M.G., Schlitzer, A., Placek, K., Joosten, L.A.B. and Schultze, J.L. (2019) 'Innate and Adaptive Immune Memory: An Evolutionary Continuum in the Host's Response to Pathogens.' *Cell Host & Microbe*, 25(1), pp. 13–26. DOI: 10.1016/j.chom.2018.12.006.

Neu, K.E.E., Tang, Q., Wilson, P.C.C. and Khan, A.A.A. (2017) 'Single-Cell Genomics: Approaches and Utility in Immunology'. *Trends in Immunology*, 38(2), pp. 140–149. DOI: 10.1016/j.it.2016.12.001.

Ning, L., Liu, G., Li, G., Hou, Y., Tong, Y. and He, J. (2014) 'Current Challenges in the Bioinformatics of Single Cell Genomics'. *Front Oncol*, 4(January), p. 7. DOI: 10.3389/fonc.2014.00007.

O'Doherty, U., Peng, M., Gezelter, S., Swiggard, W.J., Betjes, M., Bhardwaj, N. and Steinman, R.M. (1994) 'Human Blood Contains Two Subsets of Dendritic Cells, One Immunologically Mature and the Other Immature.' *Immunology*, 82(3), pp. 487–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7525461 (Accessed: 27 March 2019).

Okabe, Y. and Medzhitov, R. (2014) 'Tissue-Specific Signals Control Reversible Program of Localization and Functional Polarization of Macrophages'. *Cell*, 157(4), pp. 832–844. DOI: 10.1016/j.cell.2014.04.016.

Orgel, E., Genkinger, J.M., Aggarwal, D., Sung, L., Nieder, M. and Ladas, E.J. (2016) 'Association of Body Mass Index and Survival in Pediatric Leukemia: A Meta-Analysis'. *The American Journal of Clinical Nutrition*, 103(3), pp. 808–817. DOI: 10.3945/ajcn.115.124586.

Ozsolak, F. and Milos, P.M. (2011) 'RNA Sequencing: Advances, Challenges and Opportunities'. *Nature Reviews Genetics*, 12(2), pp. 87–98. DOI: 10.1038/nrg2934.

Packer, J. and Trapnell, C. (2018) 'Single-Cell Multi-Omics: An Engine for New Quantitative Models of Gene Regulation.' *Trends in Genetics : TIG*, 0(0). DOI: 10.1016/j.tig.2018.06.001.

Paolicelli, R.C., Bolasco, G., Pagani, F., Maggi, L., Scianni, M., Panzanelli, P., Giustetto, M., Ferreira, T.A., Guiducci, E., Dumas, L., Ragozzino, D. and Gross, C.T. (2011) 'Synaptic Pruning by Microglia Is Necessary for Normal Brain Development'. *Science*, 333(6048), pp. 1456–1458. DOI: 10.1126/science.1202529.

Parker, J.C. and Horst, W.R. (1959) (2,875,666) 'Method of Simultaneously Counting Red and White Blood Cells'. Available at: https://patents.google.com/patent/US2875666 (Accessed: 13 September 2019).

Parkin, J. and Cohen, B. (2001) ‘An Overview of the Immune System’. *The Lancet*, 357(9270), pp. 1777–1789. DOI: 10.1016/S0140-6736(00)04904-7.

Paul, F., Arkin, A., Giladi, A., Torben Porse, B., Tanay, A. and Amit, I. (2015) ‘Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors’. *Cell*, 163, pp. 1663–1677. DOI: 10.1016/j.cell.2015.11.013.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G. and Sandberg, R. (2013) ‘Smart-Seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells.’ *Nature Methods*, 10, pp. 1096–8. DOI: 10.1038/nmeth.2639.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S. and Sandberg, R. (2014) ‘Full-Length RNA-Seq from Single Cells Using Smart-Seq2.’ *Nature Protocols*, 9(1), pp. 171–81. DOI: 10.1038/nprot.2014.006.

Poitou, C., Dalmas, E., Renovato, M., Benhamo, V., Hajduch, F., Abdennour, M., Kahn, J.-F., Veyrie, N., Rizkalla, S., Fridman, W.-H., Sautès-Fridman, C., Clément, K. and Cremer, I. (2011) ‘CD14dimCD16+ and CD14+CD16+ Monocytes in Obesity and during Weight Loss: Relationships with Fat Mass and Subclinical Atherosclerosis.’ *Arteriosclerosis, Thrombosis, and Vascular Biology*, 31(10), pp. 2322–30. DOI: 10.1161/ATVBAHA.111.230979.

Randolph, G.J., Ochando, J. and Partida-Sánchez, S. (2008) ‘Migration of Dendritic Cell Subsets and Their Precursors’. *Annual Review of Immunology*, 26(1), pp. 293–316. DOI: 10.1146/annurev.immunol.26.021607.090254.

Regev, A., Rozenblatt-Rosen, O. and Teichmann, S.A. (2017) *THE HUMAN CELL ATLAS*. Available at: https://arxiv.org/ftp/arxiv/papers/1810/1810.05192.pdf (Accessed: 29 March 2019).

Rozenblatt-Rosen, O., Stubbington, M.J.T., Regev, A. and Teichmann, S.A. (2017) ‘The Human Cell Atlas: From Vision to Reality’. *Nature*, 550(7677), pp. 451–453. DOI: 10.1038/550451a.

Salzman, G.C., Crowell, J.M., Martin, J.C., Trujillo, T.T., Romero, A., Mullaney,

P.F. and LaBauve, P.M. (1975) ‘Cell Classification by Laser Light Scattering: Identification and Separation of Unstained Leukocytes.’ *Acta Cytologica*, 19(4), pp. 374–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/808927 (Accessed: 13 September 2019).

Schlitzer, A., Sivakamasundari, V., Chen, J., Sumatoh, H.R. Bin., Schreuder, J., Lum, J., Malleret, B., Zhang, S., Larbi, A., Zolezzi, F., Renia, L., Poidinger, M., Naik, S., Newell, E.W., Robson, P. and Ginhoux, F. (2015) ‘Identification of CDC1- and CDC2-Committed DC Progenitors Reveals Early Lineage Priming at the Common DC Progenitor Stage in the Bone Marrow’. *Nature Immunology*, 16(7), pp. 718–728. DOI: 10.1038/ni.3200.

Schmidt, S. V., Krebs, W., Ulas, T., Xue, J., Baßler, K., Günther, P., Hardt, A.-L., Schultze, H., Sander, J., Klee, K., Theis, H., Kraut, M., Beyer, M. and Schultze, J.L. (2016) ‘The Transcriptional Regulator Network of Human Inflammatory Macrophages Is Defined by Open Chromatin’. *Cell Research*, 26(2), pp. 151–170. DOI: 10.1038/cr.2016.1.

Schultze, J.L., Mass, E. and Schlitzer, A. (2019) ‘Emerging Principles in Myelopoiesis at Homeostasis and during Infection and Inflammation.’ *Immunity*, 50(2), pp. 288–301. DOI: 10.1016/j.immuni.2019.01.019.

See, P., Dutertre, C.A., Chen, J., Günther, P., McGovern, N., Irac, S.E., Gunawan, M., Beyer, M., Händler, K., Duan, K., Bin Sumatoh, H.R., Ruffin, N., Jouve, M., Gea-Mallorquí, E., Hennekam, R.C.M., Lim, T., Yip, C.C., Wen, M., Malleret, B., Low, I., Shadan, N.B., Fen, C.F.S., Tay, A., Lum, J., Zolezzi, F., Larbi, A., Poidinger, M., Chan, J.K.Y., Chen, Q., Rénia, L., Haniffa, M., Benaroch, P., Schlitzer, A., Schultze, J.L., Newell, E.W. and Ginhoux, F. (2017) ‘Mapping the Human DC Lineage through the Integration of High-Dimensional Techniques’. *Science*, 356(6342), p. eaag3009. DOI: 10.1126/science.aag3009.

Short, J.D., Tavakoli, S., Nguyen, H.N., Carrera, A., Farnen, C., Cox, L.A. and Asmis, R. (2017) ‘Dyslipidemic Diet-Induced Monocyte “Priming” and Dysfunction in Non-Human Primates Is Triggered by Elevated Plasma Cholesterol and Accompanied by Altered Histone Acetylation’. *Frontiers in Immunology*, 8, p. 958.

DOI: 10.3389/fimmu.2017.00958.

Siegal, F.P.P., Kadowaki, N., Shodell, M., Fitzgerald-Bocarsly, P.A.A., Shah, K., Ho, S., Antonenko, S. and Liu, Y.-J. (1999) 'The Nature of the Principal Type 1 Interferon-Producing Cells in Human Blood'. *Science*, 284(5421), pp. 1835–1837. DOI: 10.1126/SCIENCE.284.5421.1835.

Singer, K., DelProposto, J., Morris, D.L., Zamarron, B., Mergian, T., Maley, N., Cho, K.W., Geletka, L., Subbaiah, P., Muir, L., Martinez-Santibanez, G. and Lumeng, C.N.-K. (2014) 'Diet-Induced Obesity Promotes Myelopoiesis in Hematopoietic Stem Cells'. *Molecular Metabolism*, 3(6), p. 664. DOI: 10.1016/J.MOLMET.2014.06.005.

Stegle, O., Teichmann, S.A. and Marioni, J.C. (2015) 'Computational and Analytical Challenges in Single-Cell Transcriptomics'. *Nature Reviews Genetics*, 16(3), pp. 133–145. DOI: 10.1038/nrg3833.

Steinman, R.M. (1991) 'The Dendritic Cell System and Its Role in Immunogenicity'. *Annual Review of Immunology*, 9(1), pp. 271–296. DOI: 10.1146/annurev.iy.09.040191.001415.

Steinman, R.M. and Cohn, Z.A. (1973) 'Identification of a Novel Cell Type in Peripheral Lymphoid Organs of Mice. I. Morphology, Quantitation, Tissue Distribution.' *The Journal of Experimental Medicine*, 137(5), pp. 1142–62. DOI: 10.1084/JEM.137.5.1142.

Thomas, R., Davis, L.S.S. and Lipsky, P.E.E. (1993) 'Isolation and Characterization of Human Peripheral Blood Dendritic Cells.' *Journal of Immunology (Baltimore, Md.: 1950)*, 150(3), pp. 821–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7678623 (Accessed: 28 June 2019).

Tian, L., Su, S., Dong, X., Amann-Zalcenstein, D., Biben, C., Seidi, A., Hilton, D.J., Naik, S.H. and Ritchie, M.E. (2018) 'ScPipe: A Flexible R/Bioconductor Preprocessing Pipeline for Single-Cell RNA-Sequencing Data' Pertea, M. (ed.). *PLOS Computational Biology*, 14(8), p. e1006361. DOI: 10.1371/journal.pcbi.1006361.

Vallejos, C.A., Risso, D., Scialdone, A., Dudoit, S. and Marioni, J.C. (2017) 'Normalizing Single-Cell RNA Sequencing Data: Challenges and Opportunities.' *Nature Methods*, 14(6), pp. 565–571. DOI: 10.1038/nmeth.4292.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., Jardine, L., Dixon, D., Stephenson, E., Nilsson, E., Grundberg, I., McDonald, D., Filby, A., Li, W., De Jager, P.L., Rozenblatt-Rosen, O., Lane, A.A., Haniffa, M., Regev, A. and Hacohen, N. (2017) 'Single-Cell RNA-Seq Reveals New Types of Human Blood Dendritic Cells, Monocytes, and Progenitors.' *Science (New York, N.Y.)*, 356(6335), p. eaah4573. DOI: 10.1126/science.aah4573.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: A Revolutionary Tool for Transcriptomics'. *Nat Rev Genet*, 10(1), pp. 57–63. DOI: 10.1038/nrg2484.

Wilson, Q.N., Wells, M., Davis, A.T., Sherrill, C., Tsilimigras, M.C.B., Jones, R.B., Fodor, A.A. and Kavanagh, K. (2018) 'Greater Microbial Translocation and Vulnerability to Metabolic Disease in Healthy Aged Female Monkeys'. *Scientific Reports*, 8(1), p. 11373. DOI: 10.1038/s41598-018-29473-9.

Xue, J., Schmidt, S. V., Sander, J., Draffehn, A., Krebs, W., Quester, I., De Nardo, D., Gohel, T.D., Emde, M., Schmidleithner, L., Ganesan, H., Nino-Castro, A., Mallmann, M.R., Labzin, L., Theis, H., Kraut, M., Beyer, M., Latz, E., Freeman, T.C., Ulas, T. and Schultze, J.L. (2014) 'Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation.' *Immunity*, 40(2), pp. 274–88. DOI: 10.1016/j.immuni.2014.01.006.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I. and Enard, W. (2017) 'Comparative Analysis of Single-Cell RNA Sequencing Methods.' *Molecular Cell*, 65(4), pp. 631-643.e4. DOI: 10.1016/j.molcel.2017.01.023.

Zou, J., Castro, R. and Tafalla, C. (2016) 'Antiviral Immunity: Origin and Evolution in Vertebrates'. *The Evolution of the Immune System*, pp. 173–204. DOI: 10.1016/B978-0-12-801975-7.00007-4.

# Appendix

Publication I

Günther, P. and Schultze, J.L. (2019) ‘Mind the Map: Technology Shapes the Myeloid Cell Space’. *Frontiers in Immunology*, 10, p. 2287. DOI: 10.3389/FIMMU.2019.02287.

The manuscript is reprinted with permission of *Frontiers in Immunology*.

My contribution to this publication includes: Conceptualization and writing of the manuscript.

frontiers in Immunology

REVIEW
published: 04 October 2019
doi: 10.3389/fimmu.2019.02287

# Mind the Map: Technology Shapes the Myeloid Cell Space

*Patrick Günther*[1,2] *and Joachim L. Schultze*[1,2]*

[1] Genomics and Immunoregulation, Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany, [2] Platform for Single Cell Genomics and Epigenomics, German Center for Neurodegenerative Diseases and University of Bonn, Bonn, Germany

The myeloid cell system shows very high plasticity, which is crucial to quickly adapt to changes during an immune response. From the beginning, this high plasticity has made cell type classification within the myeloid cell system difficult. Not surprising, naming schemes have been frequently changed. Recent advancements in multidimensional technologies, including mass cytometry and single-cell RNA sequencing, are challenging our current understanding of cell types, cell subsets, and functional states of cells. Despite the power of these technologies to create new reference maps for the myeloid cell system, it is essential to put these new results into context with previous knowledge that was established over decades. Here we report on earlier attempts of cell type classification in the myeloid cell system, discuss current approaches and their pros and cons, and propose future strategies for cell type classification within the myeloid cell system that can be easily extended to other cell types.

**Keywords: monocytes, dendritic cells, human peripheral blood, multidimensional, single-cell RNA sequencing, mass cytometry**

OPEN ACCESS

***Edited by:***
*Loems Ziegler-Heitbrock, Independent Researcher, Munich, Germany*

***Reviewed by:***
*Marc Dalod, Centre National de la Recherche Scientifique (CNRS), France*
*Alexander Mildner, Max Delbrück Center for Molecular Medicine, Helmholtz Association of German Research Centers (HZ), Germany*

***Correspondence:**
*Joachim L. Schultze*
*j.schultze@uni-bonn.de*

***Specialty section:***
*This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology*

***Received:*** *01 July 2019*
***Accepted:*** *10 September 2019*
***Published:*** *04 October 2019*

***Citation:***
*Günther P and Schultze JL (2019) Mind the Map: Technology Shapes the Myeloid Cell Space. Front. Immunol. 10:2287. doi: 10.3389/fimmu.2019.02287*

## INTRODUCTION

Cell-type identification is an integral part of current immunology (1–5). The immune system as an organ is an assembly of an incredibly complex network of different types of immune cells including T and B lymphocytes, NK cells, innate lymphoid cells, monocytes, macrophages, and dendritic cells (DC), granulocytes including neutrophils, basophils and eosinophils, and mast cells (6). These cell types have specialized roles during homeostasis and infection. Moreover, it became clear that each of these significant immune cell types consists of cell type-specific cell subsets, for example, three monocyte subsets have been described in human peripheral blood, the so-called classical, intermediate, and non-classical monocyte (7). To understand the individual role of each of these subsets, it is crucial to understand the full heterogeneity of these cell types and their subsets to pinpoint the dedicated functions (8). This also needs to be considered in a spatiotemporal fashion, since immune cells are influenced in their function by their respective microenvironment as well as over time (9–11). For example, monocytes accumulate in peripheral reservoirs under homeostatic conditions, but during inflammation, they exert primarily pro-inflammatory effector functions (11–13). At a later time point during the repair phase of an inflammatory response, monocytes are characterized by regulatory properties necessary for tissue repair (14). During the last decade, technological advancements have been used to further refine our understanding of the diversity of cell types and subgroups within the immune system (15). These novel technologies must be put into context with the traditional way of defining cell types mostly relying on low-dimensional data

including microscopy, functional assays, and expression of single marker genes. In the first part of the review, we discuss the current principles and strategies of defining cell types and subsets, while highlighting the different aspects of resolving cellular heterogeneity. Here we want to outline how these principles have been applied to the DC/monocyte cell space. Moreover, we will provide a framework for the integration of these recent technological advances to define cell types, subsets, but also functional states of these subsets in an iterative process.

## THE MONONUCLEAR MYELOID CELL SPACE AS AN EXAMPLE FOR CELL TYPE DEFINITION

Monocytes and DC arise from the myeloid lineage of the hematopoietic system and makeup about 11% of human blood leukocytes (monocytes ∼10%, DC ∼1%). In humans, monocytes and DC are defined as $MHCII^{+}CSF\text{-}1R^{+}$ cells, mostly generated through a cascade of continuously differentiating progenitors in the bone marrow. The last shared intermediate is the monocyte-DC progenitor, MDP, which is characterized as a $CD45RA^{+}CD123^{int}CD115^{+}$ fraction of a heterogeneous granulocyte-monocyte progenitor (GMP) population (16). Using CLEC12A and CD64 expression, a focused monocyte progenitor without DC potential, the common monocyte progenitor (cMoP), was described recently (17). This restricted precursor differentiates via pre-monocytes to monocytes, which in mice egress the bone marrow in a CCR2-dependent fashion (18).

Monopoiesis is highly dependent on the hematopoietic growth factor receptor CSF-1R and is enhanced, especially during infection or "sterile" inflammation (19–22). This phenomenon highlights the function of blood monocytes, which mainly serve as a reservoir for tissue-residing monocyte-derived macrophages and monocyte-derived DCs, especially during inflammation. Under homeostatic conditions, the majority of monocytes are weak phagocytic cells and are less efficient in antigen presentation when compared to DCs and macrophages (14, 23).

Initially described by Steinman and Cohn in the early 1970s DC have been extensively studied in recent decades (24, 25). Nevertheless, the high variability regarding ontogeny, phenotype, tissue localization, and function has hampered to find a comprehensive description of this cell type for a long time. On a functional level, DC are very efficient in phagocytosis and antigen presentation and are therefore crucial for the initiation of an adaptive immune response (23). DC are generated from MDPs giving rise to DC-committed precursor cells called common DC progenitors (CDP) which serve as precursor for plasmacytoid DCs and the two classical DC subtypes cDC1 and cDC2 (26, 27). Recently, a cDC-restricted progenitor cell, the pre-cDC, has been described in mouse and human (5, 28–30). Concerning pDCs, a new model has been recently suggested (1, 31). In fact, 70–90% of pDCs seem to be IRF8-dependent and derive from a different pre-pDC precursor. These cells actively produce type I interferons and do not present antigen very well. Further studies are required to corroborate these recent findings.

## WHICH ASPECTS DEFINE CELLULAR IDENTITY?

### The Traditional Approach: Morphology, Phenotype, and Function

Several characteristics have been used to describe and define cell types and subsets. Initially, morphological characterization by early microscopy and functional observations laid the ground for the idea of different categories of cells. Primarily, features like size, shape of the cell, and/or nucleus, density, and staining behavior for specific dyes were used to separate immune cells into several cell types and subsets (24, 32–37).

Collectively described as mononuclear phagocytic cells, macrophages and monocytes were defined by their unique morphology and ability to take up pathogens and debris (32, 33, 38, 39). Several experiments suggested that blood-derived monocytes will give rise to different types of tissue-resident macrophages, which was comprehended by van Furth and Cohn as the "mononuclear phagocyte system" (MPS) (40). Later, Ralph Steinmann described cells that display a characteristic morphology when cultured on glass surfaces (24). Due to their morphology, he termed them dendritic cells. These DCs were quickly found to be professional antigen presenting phagocytes and were incorporated into the definition of the MPS (25, 41, 42).

The MPS has been defined based on morphology and shared functionality of monocytes, DCs, and macrophages as a broader framework to describe the role of these cell types during homeostasis and immunity. However, the original definition of the MPS cannot adequately explain the heterogeneity of these cell types concerning their origin, tissue localization, disease association, regulation, and function. For example, contrary to the original ideas, blood monocytes are not the only reservoir for tissue-resident macrophages. An enormous body of research established that tissue-resident macrophages are mostly generated by early progenitors during embryogenesis and exhibit to a limited extend the partial ability for self-renewal (43–47). Only some tissues of barrier organs like the intestine rely on the replenishment of tissue-resident macrophages by differentiation of monocytes during adult life, especially during infection or inflammatory conditions (48). Nevertheless, when looking at monocyte-derived and tissue-resident macrophages, we must acknowledge that these cells have a high phenotypic and functional similarity. This redundancy is essential for the (functional) replacement of yolk-sac derived tissue-resident macrophages in some tissues but makes it difficult to find a unified classification.

The use of surface marker detection by monoclonal antibodies and flow cytometry has revolutionized the way of cell type definition throughout immunology. While a functional heterogeneity of monocytes was suggested by several earlier studies (34–37, 49), it was two-color flow cytometry that provided a tool to clearly define two major monocyte subsets by their expression of CD14 and CD16 (50, 51). About 80 to 90% percent of peripheral blood monocytes express CD14 but lack the expression of the Fcγ-receptor III (FcγRIII/CD16). This subset is characterized by a higher phagocytic activity compared to the minor subset expressing CD16 and intermediate levels

of CD14. Also, $CD16^+$ monocytes can be further separated based on their expression of CD14 into $CD14^{dim}$ $CD16^+$ population and a less frequent $CD14^+$ $CD16^+$ subset (52, 53). The $CD14^+CD16^-$ subset of monocytes is referred to as classical monocytes, monocytes expressing CD14 and CD16 as intermediate monocytes and non-classical monocytes are defined as the $CD14^{dim}CD16^+$ subset (7, 52, 53). Furthermore, during the last decade, several markers have been suggested for defining the monocyte cell heterogeneity, including Slan and CD2/Fc$\varepsilon$RI (54–56). However, these markers do not reach the specificity that would be required for an unambiguous definition of cell types or cell subsets (also see below and **Box 1**).

Like monocytes DC have been first described on the basis of their morphological and functional aspects. Here, pDCs are characterized as main type-I interferon (IFN-α/β) secreting cells with plasma cell-like morphology (57). Activation and secretion of type-I interferons are facilitated by recognition of virus-derived nucleic acids, especially by endosomal nucleic acid-sensing Toll-like receptors (TLRs) TLR7 and TLR9 (57). Initially, these cells were identified by several groups under different names, including natural interferon-producing cells, plasmacytoid monocytes, and plasmacytoid T-cells (58–61). Finally, a consensus name, the plasmacytoid DC was introduced and phenotypic markers were defined including human blood dendritic cell antigen (BDCA)-2, human IL-3Rα (CD123) and BDCA-4 (57, 62–65). However, as already mentioned before and described in more detail later, previously reported experiments suggest that this consensus is once again challenged (1, 31) strongly arguing for an iterative process of cell type definition continuously including new information.

Besides pDCs, there are two subsets of myeloid or classical DC (mDC/cDC) that can be distinguished in the Lin–MHC-II+CD11c+ fraction (66, 67) by using the non-overlapping markers CD1c (BDCA1) or CD141 (BDCA3) in flow cytometry (64, 65). These DC subsets have been termed cDC1 ($CD141^+$ DC) and cDC2 ($CD1c^+$ DCs), respectively, which have been reviewed extensively elsewhere (68–73). While these classical markers are widely used, further markers have also been suggested for subset classification of DCs (73, 74). For instance, $CD141^+$ cDC1 can be identified by using antibodies against XCR1 (75, 76), CLEC9A (77–79) and CADM1 (80). Interestingly, all DC populations vary regarding their expression of the pattern recognition receptor family toll-like receptors, which is highly correlated with the functional roles these cells play in T-cell

**BOX 1** | Proposed framework for the definition of cell types, cell subsets, and functional states of cell types and subsets.

Cell type definition based on a single parameter space (e.g. only ontogeny) will be inferior to integrated approaches utilizing additional information (ontogeny, -omics data, phenotypic, and functional data). Nevertheless, even with such a large heterogeneous parameter space at hand, cell type definition is still not trivial. We propose a framework to define cell types and their subsets that is based on knowledge from decades of developmental and cell biology, further substantiated with recent developments and results in the field of single-cell omics (165–168). Certainly, such proposal will require larger community involvement and is mentioned as a starting point for discussion. This principle can be extended to define other cell types as well.

According to this framework, "**cell types**" would be defined as follows:

"Cell types" constitute the highest category. Cell types are defined by the lack of transdifferentiation capacity in more than 95% of all physiological and non-physiological conditions. Furthermore, cell types exhibit certain phenotypic, functional and genome-wide (transcriptome, epigenome, other) characteristics that are unique to all cells of a particular type. For immune cells that are terminally differentiated, cell types would include T and B lymphocytes, NK cells, monocytes, macrophages and DC, neutrophils, basophils and eosinophils, mast cells and innate lymphoid cells. For the stem cell and precursor compartment, the hematopoietic stem cell would be one cell type, while all precursors could be another cell type. Particularly in the precursor space, more research is required to define whether—based on this definition—further cell types or only cell subsets (see below) exist. This is similarly true for cell type development during embryogenesis. However, such a framework would certainly guide future research, specifically exploiting experimental systems that would allow answering the question, whether a cell is still capable of transdifferentiating toward another cell type.

"**Cell subsets**" would be defined as follows:

"Cell subsets" are a secondary category within any given cell type. Cell subsets share certain phenotypic, genome-wide (transcriptome, epigenome) and functional features within a given cell type, but are distinct in other phenotypic, functional, or genome-wide features that are unique to them within a cell type. In an ideal setting, these features should not overlap with those features that characterize the cell type. Furthermore, the feature set characterizing a cell subset should not change if cells are analyzed from different compartments (tissues, organs) and under differing conditions (homeostasis, acute inflammation, repair conditions, etc.). Cell subsets can be further distinguished from cell types in that cell subsets can change into another subset of the same cell type to the degree that is higher than 5%. For example, it is known that classical monocytes can further differentiate into non-classical monocytes via the intermediate monocyte subset.

"**Functional states**" are defined as follows:

"Functional states" are the overall current program of any given cell. Again, "functional states" would be defined by a specific pattern of phenotypic, functional and genome-wide characteristics, which ideally would exclude features characterizing cell types or subsets. "Functional states" rely on spatiotemporal information (e.g., location, the cell's individual age, the age of the organism), the activation state (homeostasis, acute, chronic inflammation, repair phase, etc.) and any combination thereof. Clearly, "functional states" can only be defined by integrated approaches and patterns or signatures of many parameters. Single parameter definitions for functional states are very unlikely. Any given cell can be described by combinations of "functional states." In other words, "functional states" can be linked to intracellular biological modules responsible for different cellular functions. A cell could express pro-inflammatory cytokines and have elevated migratory capacity. "Functional states" can even be shared among different cell types and cell subsets. However, together with the definition of the cell type and subset, a cell can be defined unambiguously according to the three levels of cell type classification.

"Cell types," "cell subsets," and "functional states" will be governed by transcriptional programs that are linked to defined and specific networks of transcription factors (TFs) not only single TFs. Therefore, the description of such networks might be another means of defining cells accordingly.

The introduction of functional states will reduce the excessive introduction of new cell types or subsets and—in our view—also represents the well-known plasticity of the myeloid cell space better.

activation. For example, human CD141$^+$cDC1 cells express high amounts of TLR3 (81), a pattern recognition receptor highly associated with cross-presentation (82) and thus cDC1s are specialized in presenting intracellular antigens to CD8$^+$ T-cells in human and mice (83).

The most abundant subset of blood DCs are CD1c$^+$ cDC2s, which can be defined analytically by expression of CD11c, CD1c (BDCA1), and FCεRIa (54, 64, 84). Furthermore, CD1c$^+$ cDC2 express high levels of class II MHC molecules like HLA-DR, HLA-DQ, and show a high endocytic capacity, which specializes this DC type for the presentation of exogenous antigen to CD4$^+$ T cells (64, 84). As we will outline below, future work will require community efforts to integrate the differential usage of cell subset classification markers to generate consensus nomenclatures.

Collectively, the definition of cell types of the MPS and their subsets was initially based on cellular morphology, further developed by introducing immunophenotyping using antibodies against the respective cell surface markers and complemented by a functional assessment of the cell subsets identified. We spare the many controversial findings throughout this period, which only reflects the limitations of these approaches to generate a widely accepted nomenclature of cell types and subsets.

## Ontogeny as a Concept for Cell Type Definition

A group of leading experts in the field of monocyte, DC, and macrophage biology has recently proposed a nomenclature, which is based mainly on the ontogeny and tissue localization of cells (73). The proposed two-level model defines a cell type, first by its origin (level 1), which is further improved by adding a functional, phenotypic or location information (level 2) of the particular cell type. This aspect of cell type classification and the ontogeny of DCs and monocytes have been reviewed extensively (48, 72, 85).

The usage of cellular origin for cell type classification is beneficial since such approach already segregates distinct, functional units. For example, it was suggested that all phagocytes that are generated by yolk-sac derived progenitors should be referred to as macrophages and cells derived from the hematopoietic lineage as monocyte-derived cells (8, 68). A further advantage of using origin and development of immune cells as a guiding principle for cell type definition is the conservation of ontogeny across species. However, although there is a substantial overlap of ontogenies in human and murine macrophage, monocyte and DC development, there is also considerable disagreement (16, 83, 86–88). Additionally, the ontogeny of myeloid cells is difficult to study in humans, and most results are obtained by mice experiments and then projected to human myeloid cells. Clearly, the ontogeny approach is a very important aspect of cell type definition, but it needs to be combined with other characteristics of cells.

# HIGH-DIMENSIONAL APPROACHES SHAPE THE MYELOID CELL SPACE

Here, we introduce the latest technological advancements that have made substantial contributions to clarify the monocyte/DC compartment. Furthermore, we want to discuss open questions and challenges associated with these new technologies. Multi-dimensional approaches have significantly improved our understanding of the myeloid cell space by providing more features resulting in higher resolution for cell typing. To contextualize this, we want to provide examples that outline how high-dimensional methods have shaped our understanding of heterogeneity in human blood-derived monocytes and DC.

Although conventional flow cytometry has revolutionized cell type classification, it is limited in the number of parameters (markers <20) being analyzed at the same time. In the early 2000s, there were a couple of technological advancements that paved the way to the development of mass cytometry enabling parallel analysis of up to 40 parameters (89–93). This higher depth of data simultaneously enabled a multitude of possibilities for immunological and biomedical sciences, including the high-dimensional assessment of cross-patient cell type dynamics during acute myeloid leukemia (94–97). More recently, multi-color flow cytometry (MCFC) has been introduced, increasing the parameter space to a similar range, as seen in mass cytometry. However, although mass cytometry and MCFC allow high-throughput protein profiling of thousands of cells, the restriction to <40 protein markers may be underrepresenting the true number of variables that are necessary to define the heterogeneity in highly complex biological samples. Besides, these markers have to be selected *a priori*, which may put a bias on the results obtained by mass spectrometry or MCFC. Another revolution was introduced by the development of high-throughput gene expression profiling methods like microarray-based technologies and RNA-sequencing enabling to profile thousands of genes in a single sample (98, 99). This second genomic revolution enables the genome-wide assessment of gene expression, which not only allows to characterize cellular subsets but also to investigate regulatory networks (20, 100–102).

One of the first studies that performed microarray analysis of human DCs compared the transcriptomes of sorted cDC1, cDC2, and pDCs populations from peripheral blood and tonsils to deeply characterize these subsets (103). Robbins et al. performed a comparative study to put the transcriptome data of DC subsets into context of other myeloid and lymphocyte populations in blood (104), which resulted in the identification of important conserved signature genes, thereby strengthening cDC1, cDC2, and pDC as distinct DC subsets. Moreover, assessing transcriptomic data of both murine and human immune cells allowed to align DC subsets across species (104, 105). Another important study performed transcriptome profiling of human blood CD14 and CD16 monocyte populations, three DC subsets pDC, cDC1, and cDC2 as well as their skin counterparts cDC1, cDC2, and skin derived CD14$^+$ cells (80).

Notably, cell types like skin cDC1 and cDC2 grouped together with their counterparts isolated from blood, suggesting a high similarity of DC subsets independent from the microenvironment. We extended these findings to compare different DC subsets in many individuals and different tissues [lymphohematopoietic (blood, thymus, spleen) and non-lymphohematopoietic (skin, lung)] allowing to characterize the impact of the microenvironment on the identity of a cell

type (74). Integration of immune phenotyping, gene expression profiling, and bioinformatic analysis revealed that DC subsets from blood, spleen, and thymus were transcriptionally conserved, with only minor transcriptomic differences between the same DC subsets across tissues. In contrast, the transcriptomic consequence of the respective microenvironment was stronger in lung and skin subsets. This suggests a higher tissue imprinting of non-lymphohematopoietic DC subsets in barrier organs like lung and skin, when comparing to the tissue imprinting that has been reported for tissue-resident macrophage subsets (47, 100, 106, 107). However, the difference between different DC subsets (cDC1 vs. cDC2) is still larger than the differences between the same DC subset among different tissues (e.g., skin cDC1 vs. blood cDC1).

Collectively, gene expression profiling and comparative bioinformatic analysis have substantially contributed to understand the complex DC networks across species further improving current descriptions of unified and more unbiased classifications (73, 105, 108).

Early transcriptomic approaches of human and mouse monocyte subsets not only helped to deeply characterize these cell types but also presented a framework to validate high conservation of gene expression profiles between mouse and humans (104, 109). For example, a combination of well-designed functional assays and gene expression profiling helped to refine the role of non-classical monocytes as the counterpart to murine "patrolling" $Gr1^-$ monocytes (110). Other studies sharpened the definition of the intermediate and non-classical monocytes as distinct cell subset (110–112). Interestingly, these studies revealed a high similarity of non-classical and intermediate monocytes, underlining the transitional nature of these cells, as they show intermediate expression for most of the marker genes differentially expressed between classical and non-classical monocytes. Interestingly, a unique module of class-II MHC genes was highest expressed in the intermediate monocyte population (111).

Measuring RNA rather than protein levels represents one of the major limitations of gene expression profiling methods. While the overall correlation of transcriptome and proteome is relatively high (113, 114), RNA-seq and microarrays do not allow to assess post-translational modifications, which represent a central part of cellular regulation (115, 116). To overcome this limitation, mass cytometry has been utilized to profile post-translational modifications like phosphorylation, methylation, and glycosylation (117, 118). A good example of the value of methods with larger feature size compared to single or few marker studies is the definition of cells expressing the carbohydrate modification 6-Sulfo LacNAc (*Slan*) on the PSGL1 protein. Indeed, myeloid cells presenting Slan initially were termed "SlanDCs" (119–121), while others described an overlap of $Slan^+$ cells with non-classical monocytes (122, 123). However, all these studies largely rely on low-dimensional marker assessment by flow cytometry and are not always directly comparable due to differences in their choice of markers or gating strategies. To investigate this in a more unbiased fashion, Roussel et al. defined a 38-marker panel to study human myeloid cells from peripheral blood by mass cytometry (124). A semi-supervised analysis of the data resulted in the identification of distinct monocyte populations, two subsets overlap with markers from classical and intermediate monocytes while there are two subsets of monocytes that are similar to non-classical monocytes. The multi-dimensional analysis maps $Slan^+$ cells to the non-classical monocytes and does not show alignment with any DC population. In this study, Slan separates the non-classical monocytes into a $Slan^{high}$ and a $Slan^{low}$ $CD14^{dim}CD16^+$ population. However, earlier genomic comparisons of sorted $Slan^{high}$ vs. $Slan^{low}$ subsets did not reveal a significant difference between those two populations (110). More recently, by combining index sorting and high-content single-cell RNA-sequencing, we show further evidence that Slan expression does not reflect different cell subsets as the underlying overall transcriptional program is not different between $Slan^{high}$ and $Slan^{low}$ cells. Moreover, we clearly show that Slan+ cells are all non-classical monocytes (125).

Manual gating of monocytes by CD14 and CD16 is biased by the investigator, which is a disadvantage for large multi-center clinical studies. Unsupervised and semi-supervised computational analyses improve the accuracy and reproducibility of subset definitions (95, 117, 124, 126–128). However, interpretation of these results must be performed with special care, since the primary analysis is still dependent on manual parameter settings by the investigator. For example, in contrast to an earlier study utilizing mass cytometry (124) similar profiling of human mononuclear myeloid cells revealed three subsets of human monocytes in two other studies, while others report significant heterogeneity including three non-classical, one intermediate and four classical subsets (22). Interestingly, Hamers et al. identified a non-classical population, which is quite different to other non-classical populations and expresses $CD9^+$ $CD41^+$ and $CD61^+$, which may represent an eosinophil/basophil contamination (129–131). Another interesting observation is the rather low inter-individual difference of human monocyte populations during homeostasis when assessed by mass cytometry (22, 132).

High-throughput gene expression profiling by microarray or RNA-seq has paved the way to understand the regulatory networks within human monocytes and DC. These technologies are indispensable for high-depth characterization of immune cell types. Nevertheless, these population-based methods are not designed to detect further cellular heterogeneity within a sample. The gene expression measurement in a population-based RNA-seq represents an average signal of typically more than 10,000 individual cells, resulting in leveling out any further heterogeneity. Frequently, samples are generated by flow cytometry assisted cell sorting, which relies on the information of a limited set of marker genes. However, if these markers are not sufficient for detecting the full heterogeneity of the tissue, the results may be underestimating the true heterogeneity.

Transcriptional profiling of individual cells by single-cell RNA-seq has been introduced in 2009 (133, 134) and has revolutionized cell type discovery in all fields of biology (135–142), therefore it may be claimed as "third genomic revolution." Single-cell RNA sequencing approaches allow transcriptional profiling of 10,000s of individual cells. In contrast

to population-based RNA-seq, the groups of cells are not defined *a priori*, rather the cell classification is based on the similarity of gene expression profiles.

A series of studies applied single-cell RNA-seq to understand the heterogeneity of human blood DCs and DC progenitors (5, 30, 143). See et al., as well as Villani et al., detected and characterized the conventional subsets, including cDC1, cDC2, and pDC. Surprisingly, beyond these similarities the results differed significantly, strongly arguing that such high-dimensional data require particular care when assigning cell types and cell subsets. We defined cell types and subsets by a combination of function, phenotype and transcriptional profile, which lead to the identification of precursors (pre-cDCs) for the cDC1 and cDC2 subsets in addition to the three main DC subsets (5). To reconcile these two major initial reports, we developed a strategy that allows developing cell type classification consensus based on phenotypic and transcriptional features also including prior knowledge (125). This approach revealed that (1) the $AXL^{+}Siglec6^{+}$ DCs (AS-DCs) described by Villani et al. are mainly pre-cDCs as described in (5), (2) Mono4 are contaminating $CD56^{dim}$ NK cells, and (3) cells introduced as $CD16^{+}$ $CSF1\text{-}R^{+}$ $CTSS^{+}$ DCs are not belonging to the DC lineage. This general strategy is not restricted to myeloid cells but can be applied to any cell type classification problem in any species (125).

Recently, single-cell RNA-seq has also been used for improving our knowledge about the generation of DCs from bone marrow-derived progenitors. There is evidence that there is much higher flexibility in the development of DC and monocytes than already appreciated. Hematopoietic models that are not based on repeating rounds of division and differentiation (72, 144, 145) allow for incorporation of recent findings that suggest that cDCs can be generated by lymphoid progenitors (146). Also, the latest reports show important evidence that the large majority of pDCs arise from lymphoid progenitors rather than CDPs (1, 31). Probably, a community effort to clarify future naming and nomenclature of these cells is now warranted. Importantly, the recent high-dimensional characterization of pDCs (5, 125, 132, 143) and new insights into their ontogeny in mice (1) could form the basis for such new discussions.

Clearly, this is only the beginning of applying these technologies to open questions concerning the plasticity of the myeloid cell compartment. We also recognize that single-cell RNA-seq data are currently challenging our view on cell type classification and function within the myeloid cell compartment. However, in the long run, we are convinced that the higher information content per cell will give us a much better understanding of individual cells within any given tissue, organ, or inflammatory response.

## PROPOSAL OF GENERAL PRINCIPLES FOR CELL TYPE DEFINITIONS

Considering the apparent ease, with which different cell types were characterized based on morphological differences a century ago (39), our capabilities to simultaneously measure hundreds to thousands of parameters per single cell seem to decrease our ability to agree on defined cell types and cell subsets (1, 5, 31, 143). The ability to detect heterogeneity between individual cells has extended to biological differences that are not related to questions concerning cell type or cell subset. The best-characterized biological process in single cell –omics data being cell cycle in proliferating cells (147–149). Certainly, cell cycle differences should not classify two cells of the same type as different cell types or subsets. Stochastic behaviors of single cells, e.g., in transcription (150, 151) would be another biological phenomenon that should not impact on cell classification aspects. Furthermore, data sparsity, still very apparent in all sequencing-based single cell technologies, requires attention, when dealing with cell type definitions.

Similarly, important is the question, whether all biased approaches requiring feature selection (e.g., which markers to be analyzed) prior to analysis are good starting points for cell type definitions. These would include all multi-color flow cytometry and single-cell mass spectrometry approaches. Potentially a more appropriate approach would be the combination of markers (chosen by the investigator) with unbiased approaches provided by single cell sequencing-based technologies. This is crucial since it allows to link the enormous body of research that has been performed with flow cytometry-defined cell populations (e.g. ontogeny) with results obtained by analysis of high-dimensional data. For example, index sorting based on previously defined cell surface markers combined with scRNA-seq might be a better way of defining the cell population structure as well as the practicality of certain protein markers to capture the population structure (125, 152, 153). Alternative but significantly more expensive approaches are based on the combination of full transcriptome scRNA-seq and oligonucleotide-labeled antibodies (154, 155). It can be expected that these approaches require iterations of experiments until markers are identified that truly reflect the underlying population structure. In this context, it is important to note that even such large endeavors such as the Human Cell Atlas will require the integration of additional layers of information in addition to scRNA-seq data. Furthermore, we postulate that these iterations will lead to consensus maps as a basis for cell type definitions (125). Very much like the cluster of differentiation (CD) workshops for antibodies (156), a community effort will be necessary to agree on the different versions of such consensus maps of individual cell types.

However, even if the combination of truly unbiased single cell –omics approaches and antibody-based techniques leads to novel consensus maps of immune cells including the myeloid cell space, we propose that each cell type and more importantly each cell subset requires to be functionally characterized, as we have previously demonstrated for human DCs in blood (5). In other words, we strongly argue that a final definition of a cell subset should be validated on functional differences and not only on transcriptional and phenotypic differences.

Once cell types are defined under homeostatic conditions, which is a major goal of the Human Cell Atlas (157), an even more daunting task will be to define cell types and subsets under pathophysiological processes. While certain cell types will be under developmental trajectories (cell states) under physiological conditions, the space for different cell states in disease settings will further increase (158). More importantly, under these

conditions, there will be mainly changes in parameters related to biological function rather than features defining cell types or subset. A major goal for further cell type definitions will be to integrate these functional states and trajectories. In this context, we propose cell types as the highest level to distinguish cells. For example, DC, monocytes, and macrophages would qualify as individual cell types, while pDC, cDC1, and cDC2 would qualify as DC subsets (5, 125). Each of these subsets can exist in different functional states that depend on location, differentiation stage, acute or chronic activation signals, to name only a few (69, 74). Again, even for functional states, we would propose to define cells based on hundreds of parameters measured by single cell –omics technologies to be combined with classical marker strategies but finally also integrate functional readouts for these cellular states.

Even if we can agree on such an approach, the question remains, how this can be realized technically? In fact, this is not a mere technical question, as it requires to consider methods that are more independent of investigator bias. For example, we strongly suggest building approaches that will allow us to build cell type definitions based on machine learning rather than on investigator-driven and individualized analysis pipelines. Single-cell transcriptomics algorithms as they are implemented in singleR (159) or scMatch (160) are good starting points. Nevertheless, they still heavily rely on an investigator's interpretation of such high-feature data spaces. Cell type definition could be a classification problem requiring the respective machine learning as they are used for classifier generation in other areas (161, 162). We do not favor solely data-driven machine learning but would suggest the integration of prior knowledge. First attempts to develop such methods are currently underway, and we will soon know, whether the introduction of machine learning based cell class prediction will truly aid our attempts to make sense of the hundreds to thousands of parameters that we now can routinely measure from single cells.

## SUMMARY AND OUTLOOK

Since the discovery of myeloid cells more than a century ago, we have learned a lot about these important immune cells. Their enormous plasticity is fascinating and challenging at the same time. Not surprisingly, cell type definitions and nomenclature—up to the day—have been changed or updated regularly (48, 68, 108, 163, 164). A unified nomenclature is the basis for an effective communication among scientists and will accelerate discovery of novel therapeutics. Moreover, high-dimensional profiling of samples will facilitate to compare results and cell types across experiments, tissues and species. Even with the highest number of parameters known per any given cell, we still differ in our interpretations of certain cell types within the myeloid compartment. While it will be rather critical to include prior knowledge when labeling cells based on high-dimensional single cell data, we need to develop better tools based on robust mathematical rules that help us to determine cellular phenotypes and functions less ambiguously. With the emergence of powerful machine learning and AI-based methodology, the time has probably come to utilize such approaches to our benefit when describing cell types, cell subsets, and their functional states. Irrespective of the power of such approaches, we also need to accept that we are far from a complete understanding of these cells. Additional layers of information, for example, epigenetic information, will have to be included in cell type definitions as they arise. Therefore, we foresee numerous iterations of defining cell types and their functions in the decades to come. In other words, consensus maps of cell types and subsets that we agree on today will form the basis for newer maps with updated information content in the future. A potential framework for such a community-based effort has been outlined here.

## AUTHOR CONTRIBUTIONS

PG and JS conceived and wrote the manuscript.

## FUNDING

This work was supported by the German Research Foundation to JS (GRK 2168, INST 217/577-1, EXC2151/1), by the HGF grant sparse2big to JS, the FASTGenomics grant of the German Federal Ministry for Economic Affairs and Energy to JS, and the EU project SYSCID under grant number 733100.

## REFERENCES

1. Dress RJ, Dutertre C-A, Giladi A, Schlitzer A, Low I, Shadan NB, et al. Plasmacytoid dendritic cells develop from Ly6D+ lymphoid progenitors distinct from the myeloid lineage. *Nat Immunol*. (2019) 20:852–64. doi: 10.1038/s41590-019-0420-3
2. Van Galen P, Hovestadt V, Wadsworth MH, Aster JC, Lane AA, Bernstein Correspondence BE, et al. Single-Cell RNA-Seq reveals AML hierarchies relevant to disease progression and immunity. *Cell*. (2019) 176:1265–81.e24. doi: 10.1016/j.cell.2019.01.031
3. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The immune landscape of cancer. *Immunity*. (2018) 48:812–30.e14. doi: 10.1016/j.immuni.2018.03.023
4. Ibarra-Soria X, Jawaid W, Pijuan-Sala B, Ladopoulos V, Scialdone A, Jörg DJ, et al. Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat Cell Biol*. (2018) 20:127–34. doi: 10.1038/s41556-017-0013-z
5. See P, Dutertre CA, Chen J, Günther P, McGovern N, Irac SE, et al. Mapping the human DC lineage through the integration of high-dimensional techniques. *Science*. (2017) 356:eaag3009. doi: 10.1126/science.aag3009
6. Parkin J, Cohen B. An overview of the immune system. *Lancet*. (2001) 357:1777–89. doi: 10.1016/S0140-6736(00)04904-7
7. Ziegler-Heitbrock L, Ancuta P, Crowe S, Dalod M, Grau V, Hart DNDN, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood*. (2010) 116:e74–80. doi: 10.1182/blood-2010-02-258558
8. Guilliams M, Mildner A, Yona S. Developmental and functional heterogeneity of monocytes. *Immunity*. (2018) 49:595–613. doi: 10.1016/j.immuni.2018.10.005
9. Garcia-Bonilla L, Faraco G, Moore J, Murphy M, Racchumi G, Srinivasan J, et al. Spatio-temporal profile, phenotypic diversity, and fate of recruited monocytes into the post-ischemic brain. *J Neuroinflamm*. (2016) 13:285. doi: 10.1186/s12974-016-0750-0
10. Kepler TBB, Chan C. Spatiotemporal programming of a simple inflammatory process. *Immunol Rev*. (2007) 216:153–63. doi: 10.1111/j.1600-065X.2007.00500.x

11. Qi H, Kastenmüller W, Germain RNN. Spatiotemporal basis of innate and adaptive immunity in secondary lymphoid tissue. *Annu Rev Cell Dev Biol*. (2014) 30:141–67. doi: 10.1146/annurev-cellbio-100913-013254
12. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature*. (2018) 563:347–53. doi: 10.1038/s41586-018-0698-6
13. Schultze JL. Chromatin remodeling in monocyte and macrophage activation. *Adv Protein Chem Struct Biol*. (2017) 106:1–15. doi: 10.1016/bs.apcsb.2016.09.001
14. Geissmann F, Manz MG, Jung S, Sieweke MH, Merad M, Ley K. Development of monocytes, macrophages, and dendritic cells. *Science*. (2010) 327:656–61. doi: 10.1126/science.1178331
15. Neu KEE, Tang Q, Wilson PCC, Khan AAA. Single-cell genomics: approaches and utility in immunology. *Trends Immunol*. (2017) 38:140–9. doi: 10.1016/j.it.2016.12.001
16. Lee J, Breton G, Oliveira TYK, Zhou YJ, Aljoufi A, Puhr S, et al. Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow. *J Exp Med*. (2015) 212:385–99. doi: 10.1084/jem.20141442
17. Kawamura S, Onai N, Miya F, Sato T, Tsunoda T, Kurabayashi K, et al. Identification of a human clonogenic progenitor with strict monocyte differentiation potential: a counterpart of mouse cMoPs. *Immunity*. (2017) 46:835–48.e4. doi: 10.1016/j.immuni.2017.04.019
18. Jacquelin S, Licata F, Dorgham K, Hermand P, Poupel L, Guyon E, et al. CX3CR1 reduces Ly6Chigh-monocyte motility within and release from the bone marrow after chemotherapy in mice. *Blood*. (2013) 122:674–83. doi: 10.1182/blood-2013-01-480749
19. Patel AA, Zhang Y, Fullerton JN, Boelen L, Rongvaux A, Maini AA, et al. The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J Exp Med*. (2017) 214:1913–23. doi: 10.1084/jem.20170355
20. Christ A, Günther P, Lauterbach MAR, Duewell P, Biswas D, Pelka K, et al. Western diet triggers NLRP3-dependent innate immune reprogramming. *Cell*. (2018) 172:162–75.e14. doi: 10.1016/j.cell.2017.12.013
21. Nagareddy PR, Kraakman M, Masters SL, Stirzaker RA, Gorman DJ, Grant RW, et al. Adipose tissue macrophages promote myelopoiesis and monocytosis in obesity. *Cell Metab*. (2014) 19:821–35. doi: 10.1016/j.cmet.2014.03.029
22. Hamers AAJ, Dinh HQ, Thomas GD, Marcovecchio P, Blatchley A, Nakao CS, et al. Human monocyte heterogeneity as revealed by high-dimensional mass cytometry. *Arterioscler Thromb Vasc Biol*. (2019) 39:25–36 doi: 10.1161/ATVBAHA.118.311022
23. Banchereau J, Steinman RM. Dendritic cells and the control of immunity. *Nature*. (1998) 392:245–52. doi: 10.1038/32588
24. Steinman RM, Cohn ZA. Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution. *J Exp Med*. (1973) 137:1142–62. doi: 10.1084/jem.137.5.1142
25. Steinman RM, Witmer MD. Lymphoid dendritic cells are potent stimulators of the primary mixed leukocyte reaction in mice. *Proc Natl Acad Sci USA*. (1978) 75:5132–6. doi: 10.1073/pnas.75.10.5132
26. Auffray C, Sieweke MH, Geissmann F. Blood monocytes: development, heterogeneity, and relationship with dendritic cells. *Annu Rev Immunol*. (2009) 27:669–92. doi: 10.1146/annurev.immunol.021908.132557
27. Onai N, Kurabayashi K, Hosoi-Amaike M, Toyama-Sorimachi N, Matsushima K, Inaba K, et al. A clonogenic progenitor with prominent plasmacytoid dendritic cell developmental potential. *Immunity*. (2013) 38:943–57. doi: 10.1016/j.immuni.2013.04.006
28. Naik SH, Sathe P, Park H-Y, Metcalf D, Proietto AI, Dakic A, et al. Development of plasmacytoid and conventional dendritic cell subtypes from single precursor cells derived *in vitro* and *in vivo*. *Nat Immunol*. (2007) 8:1217–26. doi: 10.1038/ni1522
29. Schlitzer A, Sivakamasundari V, Chen J, Sumatoh HR Bin, Schreuder J, Lum J, et al. 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) 16:718–28. doi: 10.1038/ni.3200
30. Breton G, Zheng S, Valieris R, Tojal da Silva I, Satija R, Nussenzweig MC. Human dendritic cells (DCs) are derived from distinct circulating precursors that are precommitted to become CD1c+ or CD141+ DCs. *J Exp Med*. (2016) 213:2861–70. doi: 10.1084/jem.20161135
31. Rodrigues PFF, Alberti-Servera L, Eremin A, Grajales-Reyes GEE, Ivanek R, Tussiwand R. Distinct progenitor lineages contribute to the heterogeneity of plasmacytoid dendritic cells. *Nat Immunol*. (2018) 19:711–22. doi: 10.1038/s41590-018-0136-9
32. Murray EGD, Webb RA, Swann MBR. A disease of rabbits characterised by a large mononuclear leucocytosis, caused by a hitherto undescribed bacillusBacterium monocytogenes (n.sp.). *J Pathol Bacteriol*. (1926) 29:407–39. doi: 10.1002/path.1700290409
33. Metchnikoff E. Ueber den Kampf der Zellen gegen Erysipelkokken. Ein Beitrag zur Phagocytenlehre. *Arch für Pathol Anat und Physiol und für Klin Med*. (1887) 107:209–49. doi: 10.1007/BF01926053
34. Figdor C, Bont W, Touw I, de Roos J, Roosnek E, de Vries J. Isolation of functionally different human monocytes by counterflow centrifugation elutriation. *Blood*. (1982) 60:46–53.
35. Akiyama Y, Miller PJ, Thurman GB, Neubauer RH, Oliver C, Favilla T, et al. Characterization of a human blood monocyte subset with low peroxidase activity. *J Clin Invest*. (1983) 72:1093–105. doi: 10.1172/JCI111034
36. Akiyama Y, Stevenson GW, Schlick E, Matsushima K, Miller PJ, Stevenson HC. Differential ability of human blood monocyte subsets to release various cytokines. *J Leukoc Biol*. (1985) 37:519–30. doi: 10.1002/jlb.37.5.519
37. Elias J, Chien P, Gustilo K, Schreiber A. Differential interleukin-1 elaboration by density-defined human monocyte subpopulations. *Blood*. (1985) 66:298–301.
38. Ebert RH, Florey HW. THE extravascular development of the monocyte observed *in vivo*. *Br J Exp Pathol*. (1939) 20:342–56.
39. Metchnikoff E. *Leçons sur la Pathologie Comparée de L'inflammation : Faites à l'Institut Pasteur en Avril et mai 1891*. Paris: G Masson (1892).
40. van Furth R, Cohn ZA. The origin and kinetics of mononuclear phagocytes. *J Exp Med*. (1968) 128:415–35. doi: 10.1084/jem.128.3.415
41. Adams DO, Edelson PJ, Koren HS. *Methods for Studying Mononuclear Phagocytes*. New York, NY: Academic Press (1981). 1023 p.
42. Steinman RM. The Dendritic cell system and its role in immunogenicity. *Annu Rev Immunol*. (1991) 9:271–96. doi: 10.1146/annurev.iy.09.040191.001415
43. Takahashi K, Yamamura F, Naito M. Differentiation, maturation, and proliferation of macrophages in the mouse yolk sac: a light-microscopic, enzyme-cytochemical, immunohistochemical, and ultrastructural study. *J Leukoc Biol*. (1989) 45:87–96. doi: 10.1002/jlb.45.2.87
44. Ginhoux F, Greter M, Leboeuf M, Nandi S, See P, Gokhan S, et al. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science*. (2010) 330:841–5. doi: 10.1126/science.1194637
45. Schulz C, Perdiguero EG, Chorro L, Szabo-Rogers H, Cagnard N, Kierdorf K, et al. A lineage of myeloid cells independent of myb and hematopoietic stem cells. *Science*. (2012) 336:86–90. doi: 10.1126/science.1219179
46. Hashimoto D, Chow A, Noizat C, Teo P, Beasley MB, Leboeuf M, et al. Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity*. (2013) 38:792–804. doi: 10.1016/j.immuni.2013.04.004
47. Mass E, Ballesteros I, Farlik M, Halbritter F, Günther P, Crozet L, et al. Specification of tissue-resident macrophages during organogenesis. *Science*. (2016) 353:aaf4238. doi: 10.1126/science.aaf4238
48. Ginhoux F, Schultze JL, Murray PJ, Ochando J, Biswas SK. New insights into the multidimensional concept of macrophage ontogeny, activation and function. *Nat Immunol*. (2016) 17:34–40. doi: 10.1038/ni.3324
49. Yasaka T, Mantich NM, Boxer LA, Baehner RL. Functions of human monocyte and lymphocyte subsets obtained by countercurrent centrifugal elutriation: differing functional capacities of human monocyte subsets. *J Immunol*. (1981) 127:1515–8.
50. Passlick B, Flieger D, Ziegler-Heitbrock HW. Identification and characterization of a novel monocyte subpopulation in human peripheral blood. *Blood*. (1989) 74:2527–34.
51. Ziegler-Heitbrock HWL, Ulevitch RJ. CD14: cell surface receptor and differentiation marker. *Immunol Today*. (1993) 14:121–5. doi: 10.1016/0167-5699(93)90212-4
52. Grage-Griebenow E, Zawatzky R, Kahlert H, Brade L, Flad H-D, Ernst M. Identification of a novel dendritic cell-like subset of CD64+/CD16+ blood monocytes. *Eur J Immunol*. (2001) 31:48–56. doi: 10.1002/1521-4141(200101)31:1<48::AID-IMMU48>3.0.CO;2-5

53. Grage-Griebenow E, Flad H -D., Ernst M. Heterogeneity of human peripheral blood monocyte subsets. *J Leukoc Biol*. (2001) 69:11–20. doi: 10.1189/jlb.69.1.11
54. Maurer D, Fiebiger E, Reininger B, Wolff-Winiski B, Jouvin MH, Kilgus O, et al. Expression of functional high affinity immunoglobulin E receptors (Fc epsilon RI) on monocytes of atopic individuals. *J Exp Med*. (1994) 179:745–50. doi: 10.1084/jem.179.2.745
55. Crawford K, Gabuzda D, Pantazopoulos V, Xu J, Clement C, Reinherz E, et al. Cells monocytes are dendritic + circulating CD2. *J Immunol References*. (1999) 163:5920–8.
56. Cheng YX, Foster B, Holland SM, Klion AD, Nutman TB, Casale TB, et al. CD2 identifies a monocyte subpopulation with immunoglobulin E-dependent, high-level expression of Fc?RI. *Clin Exp Allergy*. (2006) 36:1436–45. doi: 10.1111/j.1365-2222.2006.02578.x
57. Reizis B, Bunin A, Ghosh HS, Lewis KL, Sisirak V. Plasmacytoid dendritic cells: recent progress and open questions. *Annu Rev Immunol*. (2011) 29:163–83. doi: 10.1146/annurev-immunol-031210-101345
58. Rönnblomo L, Ramstedt U, Alm G V. Properties of human natural interferon-producing cells stimulated by tumor cell lines. *Eur J Immunol*. (1983) 13:471–6. doi: 10.1002/eji.1830130608
59. Chehimi J, Starr SE, Kawashima H, Miller DS, Trinchieri G, Perussia B, et al. Dendritic cells and IFN-alpha-producing cells are two functionally distinct non-B, non-monocytic HLA-DR+ cell subsets in human peripheral blood. *Immunology*. (1989) 68:486–90.
60. Desmet; FFW-PKRV van den O. The American journal of surgical pathology. *Am J Surg Pathol*. (1990) 14:101–12. doi: 10.1097/00000478-199002000-00001
61. Grouard G, Rissoan M-C, Filgueira L, Durand I, Banchereau J, Liu Y-J. The enigmatic plasmacytoid T cells develop into dendritic cells with Interleukin (IL)-3 and CD40-ligand. *J Exp Med*. (1997) 185:1101. doi: 10.1084/jem.185.6.1101
62. Siegal FPP, Kadowaki N, Shodell M, Fitzgerald-Bocarsly PAA, Shah K, Ho S, et al. The nature of the principal type 1 interferon-producing cells in human blood. *Science*. (1999) 284:1835–7. doi: 10.1126/science.284.5421.1835
63. Cella M, Jarrossay D, Facchetti F, Alebardi O, Nakajima H, Lanzavecchia A, et al. Plasmacytoid monocytes migrate to inflamed lymph nodes and produce large amounts of type I interferon. *Nat Med*. (1999) 5:919–23. doi: 10.1038/11360
64. Dzionek A, Fuchs A, Schmidt P, Cremer S, Zysk M, Miltenyi S, et al. BDCA-2, BDCA-3, and BDCA-4: three markers for distinct subsets of dendritic cells in human peripheral blood. *J Immunol*. (2000) 165:6037–46. doi: 10.4049/jimmunol.165.11.6037
65. MacDonald KPAPA, Munster DJJ, Clark GJJ, Dzionek A, Schmitz J, Hart DNJNJ. Characterization of human blood dendritic cell subsets. *Blood*. (2002) 100:4512–20. doi: 10.1182/blood-2001-11-0097
66. O'Doherty U, Peng M, Gezelter S, Swiggard WJ, Betjes M, Bhardwaj N, et al. Human blood contains two subsets of dendritic cells, one immunologically mature and the other immature. *Immunology*. (1994) 82:487–93.
67. Thomas R, Davis LSS, Lipsky PEE. Isolation and characterization of human peripheral blood dendritic cells. *J Immunol*. (1993) 150: 821–34.
68. Schlitzer A, McGovern N, Ginhoux F. Dendritic cells and monocyte-derived cells: Two complementary and integrated functional systems. *Semin Cell Dev Biol*. (2015) 41:9–22. doi: 10.1016/j.semcdb.2015.03.011
69. Schlitzer A, Zhang W, Song M, Ma X. Recent advances in understanding dendritic cell development, classification, and phenotype. *F1000Research*. (2018) 7:1558. doi: 10.12688/f1000research.14793.1
70. Vremec D, Shortman K. What's in a name? Some early and current issues in dendritic cell nomenclature. *Front Immunol*. (2015) 6:267. doi: 10.3389/fimmu.2015.00267
71. Merad M, Sathe P, Helft J, Miller J, Mortha A. The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. *Annu Rev Immunol*. (2013) 31:563–604. doi: 10.1146/annurev-immunol-020711-074950
72. Bassler K, Schulte-Schrepping J, Warnat-Herresthal S, Aschenbrenner AC, Schultze JL. The myeloid cell compartment—cell by cell. *Annu Rev Immunol*. (2019) 37:269–93. doi: 10.1146/annurev-immunol-042718-041728
73. Guilliams M, Ginhoux F, Jakubzick C, Naik SH, Onai N, Schraml BU, et al. Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. *Nat Rev Immunol*. (2014) 14:571–8. doi: 10.1038/nri3712
74. Heidkamp GF, Sander J, Lehmann CHK, Heger L, Eissing N, Baranska A, et al. Human lymphoid organ dendritic cell identity is predominantly dictated by ontogeny, not tissue microenvironment. *Sci Immunol*. (2016) 1:eaai7677. doi: 10.1126/sciimmunol.aai7677
75. Bachem A, Güttler S, Hartung E, Ebstein F, Schaefer M, Tannert A, et al. Superior antigen cross-presentation and XCR1 expression define human CD11c+CD141+ cells as homologues of mouse CD8+ dendritic cells. *J Exp Med*. (2010) 207:1273–81. doi: 10.1084/jem.20100348
76. Crozat K, Guiton R, Contreras V, Feuillet V, Dutertre C-A, Ventre E, et al. The XC chemokine receptor 1 is a conserved selective marker of mammalian cells homologous to mouse CD8alpha+ dendritic cells. *J Exp Med*. (2010) 207:1283–92. doi: 10.1084/jem.20100223
77. Caminschi I, Proietto AII, Ahmet F, Kitsoulis S, Teh JSS, Lo JCYCY, et al. The dendritic cell subtype-restricted C-type lectin Clec9A is a target for vaccine enhancement. *Blood*. (2008) 112:3264–73. doi: 10.1182/blood-2008-05-155176
78. Huysamen C, Willment JA, Dennehy KM, Brown GD. CLEC9A Is a novel activation C-type lectin-like receptor expressed on BDCA3(+) dendritic cells and a subset of monocytes. *J Biol Chem*. (2008) 283:16693–701. doi: 10.1074/jbc.M709923200
79. Sancho D, Joffre OPP, Keller AMM, Rogers NCC, Martínez D, Hernanz-Falcón P, et al. Identification of a dendritic cell receptor that couples sensing of necrosis to immunity. *Nature*. (2009) 458:899–903. doi: 10.1038/nature07750
80. Haniffa M, Shin A, Bigley V, McGovern N, Teo P, See P, et al. Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells. *Immunity*. (2012) 37:60–73. doi: 10.1016/j.immuni.2012.04.012
81. Sittig SPP, Bakdash G, Weiden J, Sköld AEE, Tel J, Figdor CGG, et al. A comparative study of the T cell stimulatory and polarizing capacity of human primary blood dendritic cell subsets. *Mediators Inflamm*. (2016) 2016:1–11. doi: 10.1155/2016/3605643
82. Schulz O, Diebold SSS, Chen M, Näslund TII, Nolte MAA, Alexopoulou L, et al. Toll-like receptor 3 promotes cross-priming to virus-infected cells. *Nature*. (2005) 433:887–92. doi: 10.1038/nature03326
83. Schlitzer A, Ginhoux F. Organization of the mouse and human DC network. *Curr Opin Immunol*. (2014) 26:90–9. doi: 10.1016/j.coi.2013.11.002
84. Ito T, Inaba M, Inaba K, Toki J, Sogo S, Iguchi T, et al. A CD1a+/CD11c+ subset of human blood dendritic cells is a direct precursor of langerhans cells. *J Immunol*. (1999) 163:1409–19.
85. Ginhoux F, Guilliams M. Tissue-resident macrophage ontogeny and homeostasis. *Immunity*. (2016) 44:439–49. doi: 10.1016/j.immuni.2016.02.024
86. Kawamura S, Ohteki T. Monopoiesis in humans and mice. *Int Immunol*. (2018) 30:503–9. doi: 10.1093/intimm/dxy063
87. Ito T, Kanzler H, Duramad O, Cao W, Liu Y-J. Specialization, kinetics, and repertoire of type 1 interferon responses by human plasmacytoid predendritic cells. *Blood*. (2006) 107:2423–31. doi: 10.1182/blood-2005-07-2709
88. Dalod M, Salazar-Mather TP, Malmgaard L, Lewis C, Asselin-Paturel C, Brière F, et al. Interferon α/β and interleukin 12 responses to viral infections. *J Exp Med*. (2002) 195:517–28. doi: 10.1084/jem.20011672
89. Quinn ZA, Baranov VI, Tanner SD, Wrana JL. Simultaneous determination of proteins using an element-tagged immunoassay coupled with ICP-MS detection. *J Anal At Spectrom*. (2002) 17:892–6. doi: 10.1039/b202306g
90. Baranov VI, Quinn ZA, Bandura DR, Tanner SD. The potential for elemental analysis in biotechnology. *J Anal At Spectrom*. (2002) 17:1148–52. doi: 10.1039/B201494G
91. Ornatsky O, Baranov VI, Bandura DR, Tanner SD, Dick J. Multiple cellular antigen detection by ICP-MS. *J Immunol Methods*. (2006) 308:68–76. doi: 10.1016/j.jim.2005.09.020
92. Lou X, Zhang G, Herrera I, Kinach R, Ornatsky O, Baranov V, et al. Polymer-based elemental tags for sensitive bioassays. *Angew Chemie Int Ed*. (2007) 46:6111–4. doi: 10.1002/anie.200700796

93. Ornatsky O, Bandura D, Baranov V, Nitz M, Winnik MA, Tanner S. Highly multiparametric analysis by mass cytometry. *J Immunol Methods*. (2010) 361:1–20. doi: 10.1016/j.jim.2010.07.002
94. Bendall SC, Simonds EF, Qiu P, Amir E -a. D, Krutzik PO, Finck R, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*. (2011) 332:687–96. doi: 10.1126/science.1198704
95. Bendall SCC, Davis KLL, Amir EDD, Tadmor MDD, Simonds EFF, Chen TJJ, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell*. (2014) 157:714–25. doi: 10.1016/j.cell.2014.04.005
96. Newell EW, Sigal N, Bendall SC, Nolan GP, Davis MM. Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes. *Immunity*. (2012) 36:142–52. doi: 10.1016/j.immuni.2012.01.002
97. Newell EW, Sigal N, Nair N, Kidd BA, Greenberg HB, Davis MM. Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nat Biotechnol*. (2013) 31:623–9. doi: 10.1038/nbt.2593
98. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. (2008) 320:1344–9. doi: 10.1126/science.1158441
99. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. (2009) 10:57–63. doi: 10.1038/nrg2484
100. Xue J, Schmidt S V, Sander J, Draffehn A, Krebs W, Quester I, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity*. (2014) 40:274–88. doi: 10.1016/j.immuni.2014.01.006
101. Schmidt S V, Krebs W, Ulas T, Xue J, Baßler K, Günther P, et al. The transcriptional regulator network of human inflammatory macrophages is defined by open chromatin. *Cell Res*. (2016) 26:151–70. doi: 10.1038/cr.2016.1
102. Sander J, Schmidt S V, Cirovic B, McGovern N, Papantonopoulou O, Hardt A-L, et al. Cellular differentiation of human monocytes is regulated by time-dependent interleukin-4 signaling and the transcriptional regulator NCOR2. *Immunity*. (2017) 47:1051–66.e12. doi: 10.1016/j.immuni.2017.11.024
103. Lindstedt M, Lundberg K, Borrebaeck CAK. Gene family clustering identifies functionally associated subsets of human *in vivo* blood and tonsillar dendritic cells. *J Immunol*. (2005) 175:4839–46. doi: 10.4049/jimmunol.175.8.4839
104. Robbins SH, Walzer T, Dembélé D, Thibault C, Defays A, Bessou G, et al. Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. *Genome Biol*. (2008) 9:R17. doi: 10.1186/gb-2008-9-1-r17
105. Crozat K, Guiton R, Guilliams M, Henri S, Baranek T, Schwartz-Cornil I, et al. Comparative genomics as a tool to reveal functional equivalences between human and mouse dendritic cell subsets. *Immunol Rev*. (2010) 234:177–98. doi: 10.1111/j.0105-2896.2009.00868.x
106. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell*. (2014) 159:1312–26. doi: 10.1016/j.cell.2014.11.018
107. Gosselin D, Link VMM, Romanoski CEE, Fonseca GJJ, Eichenfield DZZ, Spann NJJ, et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell*. (2014) 159:1327–40. doi: 10.1016/j.cell.2014.11.023
108. Vu Manh T-P, Bertho N, Hosmalin A, Schwartz-Cornil I, Dalod M. Investigating evolutionary conservation of dendritic cell subset identity and functions. *Front Immunol*. (2015) 6:260. doi: 10.3389/fimmu.2015.00260
109. Ingersoll MAA, Spanbroek R, Lottaz C, Gautier ELL, Frankenberger M, Hoffmann R, et al. Comparison of gene expression profiles between human and mouse monocyte subsets. *Blood*. (2010) 115:e10-9. doi: 10.1182/blood-2009-07-235028
110. Cros J, Cagnard N, Woollard K, Patey N, Zhang S-Y, Senechal B, et al. Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. *Immunity*. (2010) 33:375–86. doi: 10.1016/j.immuni.2010.08.012
111. Wong KLL, Tai JJ-YJ-Y, Wong W-C, Han H, Sem X, Yeap W-H, et al. Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets. *Blood*. (2011) 118:e16-31. doi: 10.1182/blood-2010-12-326355
112. Frankenberger M, Hofer TPJPJ, Marei A, Dayyani F, Schewe S, Strasser C, et al. Transcript profiling of CD16-positive monocytes reveals a unique molecular fingerprint. *Eur J Immunol*. (2012) 42:957–74. doi: 10.1002/eji.201141907
113. Vä L, Scheele C, Pedersen K, Nielsen J. Proteome-and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes accession numbers. *Cell Rep*. (2015) 11:921–33. doi: 10.1016/j.celrep.2015.04.010
114. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett*. (2009) 583:3966–73. doi: 10.1016/j.febslet.2009.10.036
115. Narita T, Weinert BTT, Choudhary C. Functions and mechanisms of non-histone protein acetylation. *Nat Rev Mol Cell Biol*. (2019) 20:156–74. doi: 10.1038/s41580-018-0081-3
116. Wang Y-C, Peterson SEE, Loring JFF. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res*. (2014) 24:143–60. doi: 10.1038/cr.2013.151
117. Spitzer MH, Nolan GP. Mass cytometry: single cells, many features. *Cell*. (2016) 165:780–91. doi: 10.1016/j.cell.2016.04.019
118. Bodenmiller B, Zunder ER, Finck R, Chen TJ, Savig ES, Bruggner R V, et al. Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat Biotechnol*. (2012) 30:858–67. doi: 10.1038/nbt.2317
119. Schäkel K, Kannagi R, Kniep B, Goto Y, Mitsuoka C, Zwirner J, et al. 6-Sulfo LacNAc, a novel carbohydrate modification of PSGL-1, defines an inflammatory type of human dendritic cells. *Immunity*. (2002) 17:289–301. doi: 10.1016/S1074-7613(02)00393-X
120. Schäkel K, von Kietzell M, Hänsel A, Ebling A, Schulze L, Haase M, et al. Human 6-sulfo LacNAc-expressing dendritic cells are principal producers of early interleukin-12 and are controlled by erythrocytes. *Immunity*. (2006) 24:767–77. doi: 10.1016/j.immuni.2006.03.020
121. Micheletti A, Finotti G, Calzetti F, Lonardi S, Zoratti E, Bugatti M, et al. slan/M-DC8$_+$ cells constitute a distinct subset of dendritic cells in human tonsils. *Oncotarget*. (2016) 7:161–75. doi: 10.18632/oncotarget.12418
122. Hofer TPP, Zawada AMM, Frankenberger M, Skokann K, Satzl AAA, Gesierich W, et al. slan-defined subsets of CD16-positive monocytes: impact of granulomatous inflammation and M-CSF receptor mutation. *Blood*. (2015) 126:2601–10. doi: 10.1182/blood-2015-06-651331
123. Siedlarl M, Frankenberger M, Ziegler-Heitbrock LHWHW, Belge K-U. The M-DC8-positive leukocytes are a subpopulation of the CD14+CD16+monocytes. *Immunobiology*. (2000) 202:11–7. doi: 10.1016/S0171-2985(00)80047-9
124. Roussel M, Ferrell PB, Greenplate AR, Lhomme F, Le Gallou S, Diggins KE, et al. Mass cytometry deep phenotyping of human mononuclear phagocytes and myeloid-derived suppressor cells from human blood and bone marrow. *J Leukoc Biol*. (2017) 102:437–47. doi: 10.1189/jlb.5MA1116-457R
125. Günther P, Cirovic B, Baßler K, Händler K, Becker M, Dutertre CAA, et al. A rule-based data-informed cellular consensus map of the human mononuclear phagocyte cell space. *bioRxiv [Preprint]*. (2019) 658179. doi: 10.1101/658179
126. Thomas GD, Hamers AAJ, Nakao C, Marcovecchio P, Taylor AM, McSkimming C, et al. Human blood monocyte subsets. *Arterioscler Thromb Vasc Biol*. (2017) 37:1548–58. doi: 10.1161/ATVBAHA.117.309145
127. Amir ED, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol*. (2013) 31:545–52. doi: 10.1038/nbt.2594
128. Samusik N, Good Z, Spitzer MH, Davis KL, Nolan GP. Automated mapping of phenotype space with single-cell data. *Nat Methods*. (2016) 13:493–6. doi: 10.1038/nmeth.3863
129. Johansson MWW. Eosinophil activation status in separate compartments and association with asthma. *Front Med*. (2017) 4:75. doi: 10.3389/fmed.2017.00075
130. Han X, Jorgensen JLL, Brahmandam A, Schlette E, Huh YOO, Shi Y, et al. Immunophenotypic study of basophils by multiparameter flow cytometry. *Arch Pathol Lab Med*. (2008) 132:813–9. doi: 10.1043/1543-2165(2008)132[813:ISOBBM]2.0.CO;2

131. Mukai K, Gaudenzio N, Gupta S, Vivanco N, Bendall SCC, Maecker HTT, et al. Assessing basophil activation by using flow cytometry and mass cytometry in blood stored 24 hours before analysis. *J Allergy Clin Immunol*. (2017) 139:889–99.e11. doi: 10.1016/j.jaci.2016.04.060
132. Alcántara-Hernández M, Leylek R, Wagar LE, Engleman EG, Keler T, Marinkovich MP, et al. High-dimensional phenotypic mapping of human dendritic cells reveals interindividual variation and tissue specialization. *Immunity*. (2017) 47:1037–50.e6. doi: 10.1016/j.immuni.2017.11.001
133. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*. (2009) 6:377–82. doi: 10.1038/nmeth.1315
134. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc*. (2010) 5:516–35. doi: 10.1038/nprot.2009.236
135. Aevermann BD, Novotny M, Bakken T, Miller JA, Diehl AD, Osumi-Sutherland D, et al. Cell type discovery using single-cell transcriptomics: implications for ontological representation. *Hum Mol Genet*. (2018) 27:R40–7. doi: 10.1093/hmg/ddy100
136. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature*. (2017) 541:331–8. doi: 10.1038/nature21350
137. Junker JP, Van Oudenaarden A. Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell*. (2014) 157:8–11. doi: 10.1016/j.cell.2014.02.010
138. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet*. (2013) 14:618–30. doi: 10.1038/nrg3542
139. The Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. (2018) 562:367–72. doi: 10.1038/s41586-018-0590-4
140. Eberwine J, Sul J-Y, Bartfai T, Kim J. The promise of single-cell sequencing. *Nat Methods*. (2013) 11:25–7. doi: 10.1038/nmeth.2769
141. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol*. (2017) 18:35–45. doi: 10.1038/nri.2017.76
142. Cao J, Packer JSS, Ramani V, Cusanovich DAA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. (2017) 357:661–7. doi: 10.1126/science.aam8940
143. Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. (2017) 356:eaah4573. doi: 10.1126/science.aah4573
144. Collin M, Bigley V. Human dendritic cell subsets: an update. *Immunology*. (2018) 154:3–20. doi: 10.1111/imm.12888
145. Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. *Nature*. (2018) 553:418–26. doi: 10.1038/nature25022
146. Helft J, Anjos-Afonso F, van der Veen AG, Chakravarty P, Bonnet D, Reis e Sousa C. Dendritic cell lineage potential in human early hematopoietic progenitors. *Cell Rep*. (2017) 20:529–37. doi: 10.1016/j.celrep.2017.06.075
147. Kolodziejczyk AA, Kim JK, Tsang JCH, Ilicic T, Henriksson J, Natarajan KN, et al. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell*. (2015) 17:471–85. doi: 10.1016/j.stem.2015.09.011
148. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol*. (2015) 33:155–60. doi: 10.1038/nbt.3102
149. Wang J, Huang M, Torre E, Dueck H, Shaffer S, Murray J, et al. Gene expression distribution deconvolution in single-cell RNA sequencing. *Proc Natl Acad Sci USA*. (2018) 115:E6437–46. doi: 10.1073/pnas.1721085115
150. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet*. (2015) 16:133–45. doi: 10.1038/nrg3833
151. Hoppe PS, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Moritz N, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*. (2016) 535:299–302. doi: 10.1038/nature18320
152. Giladi A, Paul F, Herzog Y, Lubling Y, Weiner A, Yofe I, et al. Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nat Cell Biol*. (2018) 20:836–46. doi: 10.1038/s41556-018-0121-4
153. Franziska Paul A, Arkin ara, Giladi A, Torben Porse B, Tanay A, Amit I. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell*. (2015);163:1663–77. doi: 10.1016/j.cell.2015.11.013
154. Shahi P, Kim SCC, Haliburton JRR, Gartner ZJJ, Abate ARR. Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci Rep*. (2017) 7:44447. doi: 10.1038/srep44447
155. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PKK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. (2017) Sep;14:865–8. doi: 10.1038/nmeth.4380
156. Engel P, Boumsell L, Balderas R, Bensussan A, Gattei V, Horejsi V, et al. CD nomenclature 2015: human leukocyte differentiation antigen workshops as a driving force in immunology. *J Immunol*. (2015) 195:4555–63. doi: 10.4049/jimmunol.1502033
157. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The human cell atlas meeting participants. *bioRxiv [Preprint]*. (2017). doi: 10.1101/121202
158. Wagner A, Regev A, Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol*. (2016) 34:1145–60. doi: 10.1038/nbt.3711
159. Aran D, Looney APP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*. (2019) 20:163–72. doi: 10.1038/s41590-018-0276-y
160. Hou R, Denisenko E, Forrest ARR. scMatch: a single-cell gene expression profile annotation tool using reference datasets. Kelso J, editor. *Bioinformatics*. (2019) btz292. doi: 10.1093/bioinformatics/btz292
161. Warnat-Herresthal S, Perrakis K, Taschler B, Becker M, Seep L, Baßler K, et al. Diagnostic value of blood gene expression-based classifiers as exemplified for acute myeloid leukemia. *bioRxiv [Preprint]*. (2018) 382143. doi: 10.1101/382143
162. Angra S, Ahuja S. Machine learning and its applications: a review. In: *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*. Chirala: IEEE (2017). p. 57–60. doi: 10.1109/ICBDACI.2017.8070809
163. Guilliams M, Henri S, Tamoutounour S, Ardouin L, Schwartz-Cornil I, Dalod M, et al. From skin dendritic cells to a simplified classification of human and mouse dendritic cell subsets. *Eur J Immunol*. (2010) 40:2089–94. doi: 10.1002/eji.201040498
164. Ginhoux F, Jung S. Monocytes and macrophages: developmental pathways and tissue homeostasis. *Nat Rev Immunol*. (2014) 14:392–404. doi: 10.1038/nri3671
165. Arendt D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet*. (2008) 9:868–82. doi: 10.1038/nrg2416
166. Trapnell C. Defining cell types and states with single-cell genomics. *Genome Res*. (2015) 25:1491–8. doi: 10.1101/gr.190595.115
167. Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, et al. The origin and evolution of cell types. *Nat Rev Genet*. (2016) 17:744–57. doi: 10.1038/nrg.2016.127
168. Marioni JC, Arendt D. How single-cell genomics is changing evolutionary and developmental biology. *Annu Rev Cell Dev Biol*. (2017) 33:537–53. doi: 10.1146/annurev-cellbio-100616-060818


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publication II

Christ, A.*, **Günther, P.***, Lauterbach, M.A.R., Duewell, P., Biswas, D., Pelka, K., Scholz, C.J., Oosting, M., Haendler, K., Baßler, K., Klee, K., Schulte-Schrepping, J., Ulas, T., Moorlag, S.J.C.F.M., Kumar, V., Park, M.H., Joosten, L.A.B., Groh, L.A., Riksen, N.P., Espevik, T., Schlitzer, A., Li, Y., Fitzgerald, M.L., Netea, M.G., Schultze, J.L. and Latz, E. (2018) 'Western Diet Triggers NLRP3-Dependent Innate Immune Reprogramming.' *Cell*, 172(1–2), pp. 162-175.e14. DOI: 10.1016/j.cell.2017.12.013.

* Shared first authorship

The manuscript is reprinted with permission from Cell Press.

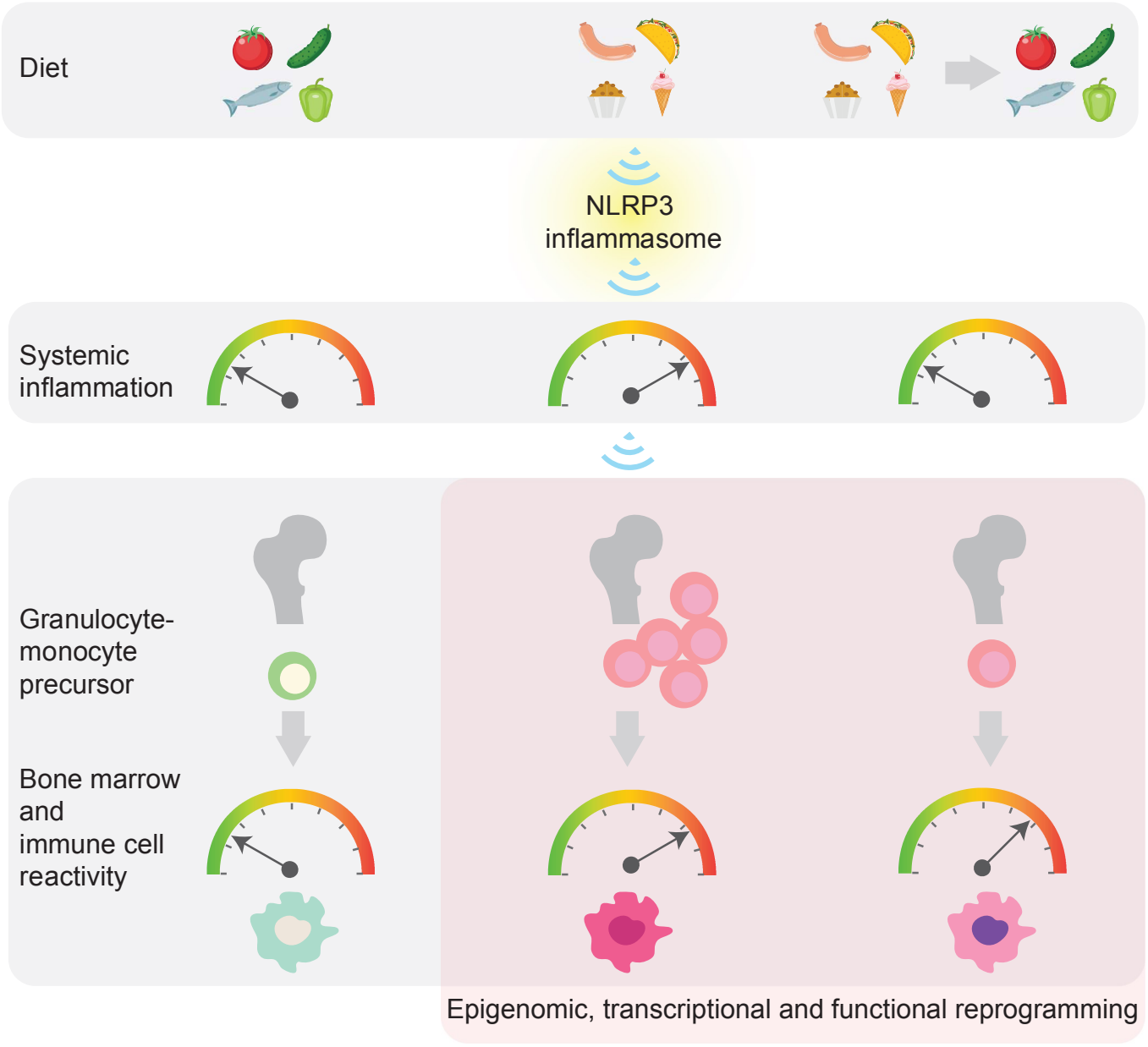
My contribution to this publication includes:

Methodology and Software: improvement of a bioinformatics pipeline for the preprocessing of ATAC-Seq data; development of strategies for data analysis, integration and visualization

Investigation and Interpretation: preprocessing, quality control and normalization of RNA-seq and ATAC-seq data; basic and advanced data analysis of RNA-seq data; basic and advanced data analysis of ATAC-seq data; visualization of all figures related to RNA-seq or ATAC-seq data

Writing: First draft, Review and Editing

Graphical Abstract



# Western diet triggers NLRP3-dependent innate immune reprograming

Anette Christ[1,2*], Patrick Günther[3*], Mario A.R. Lauterbach[1], Peter Duewell[4], Debjani Biswas[2], Karin Pelka[5], Claus J. Scholz[3], Marije Oosting[6], Kristian Haendler[3], Kevin Baßler[3], Kathrin Klee[3], Jonas Schulte-Schrepping[3], Thomas Ulas[3], Simone J.C.F.M. Moorlag[6], Vinod Kumar[7], Min Hi Park[8,9], Leo A.B. Joosten[6], Laszlo A. Groh[6], Niels P. Riksen[6], Terje Espevik[10], Andreas Schlitzer[3], Yang Li[7], Michael L. Fitzgerald[8], Mihai G. Netea[3,6], Joachim L. Schultze[3,11] and Eicke Latz[1,2,10,11**]

[1]Institute of Innate Immunity, University Hospital Bonn, University of Bonn, 53127 Bonn, Germany; [2]Department of Infectious Diseases & Immunology, UMass Medical School, 01605 Worcester, MA, USA; [3]Department for Genomics & Immunoregulation, and Myeloid Cell Biology, Life and Medical Sciences Institute (LIMES), University of Bonn, 53115 Bonn, Germany; [4]Center of Integrated Protein Science Munich (CIPSM) and Division of Clinical Pharmacology, Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, Munich, Germany; [5]Broad Institute of MIT and Harvard, 02142 Cambridge, MA, USA; [6]Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands; [7]Department of Genetics, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, The Netherlands; [8]Lipid Metabolism Unit, Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, 02114 Boston, MA, USA; [9]Texas A&M Lerma Rangel College of Pharmacy, 78363 Kingsville, TX, USA; [10]Centre of Molecular Inflammation Research, Norwegian University of Science and Technology, 7491 Trondheim, Norway; [11]German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany

[*] these authors contributed equally

** Corresponding author and lead contact

**Running title:** NLRP3-driven myeloid cell reprogramming by Western diet

**[**]Corresponding author and lead contact:**

Eicke Latz, MD, PhD, Institute of Innate Immunity, Biomedical Center (1OG 007)

University Hospitals, University of Bonn, Sigmund-Freud Str. 25, 53127 Bonn, Germany;

**Long-term epigenetic reprogramming of innate immune cells in response to microbes, also termed *trained immunity*, causes prolonged altered cellular functionality to protect from secondary infections. Here we investigated whether sterile triggers of inflammation induce trained immunity and thereby influence innate immune responses. Western diet (WD) feeding of *Ldlr*$^{-/-}$ mice induced systemic inflammation, which was undetectable in serum soon after mice were shifted back to a chow diet (CD). In contrast, myeloid cell responses towards innate stimuli remained broadly augmented. WD-induced transcriptomic and epigenomic reprogramming of myeloid progenitor cells, led to increased proliferation and enhanced innate immune responses. QTL analysis in human monocytes trained with oxidized low-density lipoprotein (oxLDL) and stimulated with LPS suggested inflammasome-mediated trained immunity. Consistently, *Nlrp3*$^{-/-}$*/Ldlr*$^{-/-}$ mice lacked WD-induced systemic inflammation, myeloid progenitor proliferation and reprogramming. Hence, NLRP3 mediates trained immunity following WD and could thereby mediate the potentially deleterious effects of trained immunity in inflammatory diseases.**

A major shift in the burden of infectious diseases has occurred over the last two centuries. Whereas approximately half of all deaths were accounted for by microbial infections in the late 19$^{th}$ century, today this burden has dropped to 15% (GBD 2015 Mortality and Causes of Death Collaborators, 2016). Three prominent changes driving this change include: *i.* widespread sanitation improvement in the 19$^{th}$ century, *ii.* continued introduction and larger deployment of vaccines in this period, and *iii.* deployment of antibiotics. However, as infectious disease burden has dropped predominantly in the last century, the development of chronic non-communicable

diseases has dramatically increased. Today, in Western societies over 80% of deaths are due to non-communicable diseases including those associated with aging and diseases caused or influenced by the consumption of Western-type calorically rich diets, such as type II diabetes, obesity and cardiovascular diseases.

For atherosclerosis development, increased circulating levels of LDL cholesterol have been linked mechanistically and genetically to clinical event risks (Mega et al., 2015). However, it is also well appreciated that disease progression is strongly associated with inflammatory processes involving cells of the innate immune system, mainly monocyte-derived macrophages (Hansson and Hermansson, 2011; Ridker et al., 2017). Hence, it is of fundamental and translational importance to understand explicitly the mechanisms that link the consumption of calorically rich diets to increased inflammation.

To respond to invading pathogens vertebrates have evolved innate and adaptive immune systems. Contrary to the adaptive immunity arm, which can induce antigen-specific memory formation upon pathogen encounter, the innate immunity arm quickly mounts non-antigen specific protective responses towards pathogens. Protective anti-pathogenic innate immune responses are evoked by the activation of a series of innate immune signaling receptors including amongst others, Toll-like receptors (TLRs) and the nucleotide-binding oligomerization (NOD)-like receptors (NLRs) (Cao, 2016). However, these receptors can further recognize ‘sterile’ danger signals, which are thought to trigger inflammation in non-communicable diseases (Zimmer et al., 2015). These observations link activation of innate immune receptors with the development of cardiovascular diseases.

Emerging evidence has indicated that the vertebrate innate immune system has evolved elaborate adaptive mechanisms allowing them to respond more vigorously to future infections. This type of functionally adapted response after an initial trigger known as

‘innate immune memory’ or ‘trained immunity’ and is mediated by epigenetic and metabolic reprogramming which can last for prolonged periods of time (Netea et al., 2016; 2011). Innate immune memory has likely evolved to provide non-specific protection from secondary infections, and most studies that characterized innate immune memory effects have focused on the ability of microbial triggers to induce cellular reprogramming. However, it is also conceivable that ‘sterile’ inflammatory triggers, such as WD, can induce trained immunity. In such a scenario, a secondary stimulus would then trigger altered and potentially pathological immune responses. Indeed, recent cell culture data suggest that oxLDL particles, which are known to trigger innate immune responses, can induce memory responses via epigenetic modification of human monocytes (Bekkering et al., 2014).

Here we have investigated whether WD feeding triggers trained immunity in the *Ldlr*$^{-/-}$ atherosclerosis mouse model. We show that WD induced systemic inflammation which subsided after shifting mice to CD. Furthermore, WD triggered a proliferative hematopoietic cell expansion associated with functionally reprogrammed granulocyte monocyte precursor cells (GMPs). These responses were maintained over prolonged times after reversing the diet from WD to CD indicating that WD can induce trained immunity. Mechanistically, we identified the NLRP3 inflammasome as the central receptor, which mediates WD-induced systemic inflammation and myeloid precursor reprogramming, opening therapeutic opportunities to interfere with WD-associated pathologies.

## Results

### WD induces systemic inflammation and functional reprogramming of myeloid cells

Given the inflammatory nature of a WD, we sought to test whether WD feeding in an experimental model of atherosclerosis induces innate immune memory. We chose to study the effects of WD after 4 weeks feeding since at this early time point only minimal atherosclerotic lesions can be observed (Duewell et al., 2010) and thus a contribution of plaque-derived immune cells and potential anti-inflammatory effects during plaque regression is minimized. *Ldlr*$^{-/-}$ mice were fed a regular CD, a WD for 4 weeks, or first WD for 4 weeks followed by 4 weeks CD (Figure 1A). Circulating cholesterol peaked by 4 weeks and returned to baseline in mice that were shifted back to CD after WD feeding (Figure 1B, Figure S1A). Likewise, circulating levels of growth factors, cytokines or chemokines, as well as acute phase proteins, which were all elevated after 4 weeks WD feeding, returned to baseline in mice fed the WD for 4 weeks followed by 4 weeks CD (Figure 1C; Figure S1C; Table S1). Mice had minimal, non-significant weight gain over the course of 4 weeks WD and no significant weight loss was observed after changing mice back to CD for 4 weeks following the 4 weeks WD (Figure S1B). Together, these studies showed that WD feeding induces a transient hypercholesterolemia concomitant with a systemic inflammatory response. To test whether the WD provoked functional immune cell reprogramming we isolated bone marrow cells and splenic CD11b$^{+}$ monocytic cells from mice fed either regular CD, 4 weeks WD, or 4 weeks WD followed by 4 weeks CD. These cells were stimulated *ex vivo* with a panel of TLR ligands (LPS, PGN, R848 and CpG) and secretion of cytokines and chemokines was analyzed. Immune cells isolated from WD fed mice had significantly enhanced TLR responses indicating a primed cell state. Intriguingly, even though the systemic cytokines had

normalized upon shifting mice back to the CD after WD, the TLR responses of cells isolated from the WD fed and then CD rested mice remained augmented when compared to cells isolated from CD fed mice. Of note, in addition to the quantitative changes of cytokine responses, we found qualitative changes. For example, bone marrow cells isolated from WD fed mice that were then CD rested had even stronger mKC and TNF responses, yet displayed decreased IL-6 responses when compared to cells isolated from mice on WD (Figure 1D, E; Figure S1D, E). These results suggest that WD feeding induced a complex myeloid cell reprograming leading to long-lasting and qualitatively altered hyper-responsiveness even after resting mice from WD feeding.

**WD triggers myelopoiesis and transcriptional reprogramming of myeloid precursor cells**

We next determined the effect of WD on circulating blood cell populations (Figure S2A). Absolute numbers of circulating red blood cells (RBCs) as well as myeloid cell subsets, including monocytes and granulocytes, were markedly increased after WD feeding (Figure 2A, B). Additionally, WD feeding induced an increased activation status in circulating myeloid subsets, as indicated by CD86 surface expression (Figure 2C; Figure S3A). Splenic inflammatory monocyte and granulocyte numbers were significantly increased as well (Figure S2B), though CD86 surface expression remained unaltered (Figure S2C; Figure S3A).

To test whether the observed changes in specific leukocyte subsets in the blood were also apparent at the bone marrow level, we determined the quantities of hematopoietic precursor subsets by comparing WD to CD fed mice. We found that the abundance of hematopoietic stem cell progenitors (HSPCs), multipotent progenitor cells (MPPs), as well as granulocyte-monocyte progenitor cells (GMPs) were all significantly increased

after WD (Figure 2D). To better understand the mechanisms whereby WD induces myelopoiesis and functional reprogramming of myeloid cells, we next isolated myeloid progenitor subsets from the bone marrow compartment by FACS sorting (Figure S2D). As GMPs are the most differentiated myeloid progenitor subsets that give rise to monocytes and granulocytes, we decided to study these cells by an unbiased approach and performed transcriptional RNA profiling by RNAseq and subsequent computational analysis (Figure S2E).

Principal component analysis (Figure 2E) and unsupervised hierarchical clustering of the 1000 most variable genes (Figure 2F) demonstrated that GMPs globally reprogram transcriptional responses after WD. Analysis of the most changed genes (Figure 2G) suggested that WD induced up-regulation of genes involved in cell proliferation, including *KLF2*, *DUSP1*, *CDKN1A*, *CCND1*, *SOX4*, *SMO* and *BCL2*, and a skewing of GMP development towards the monocytic cell lineage (*JUND, FOSB, FLT3 and HDAC9)*. To test for a WD-induced lineage bias in GMPs, we first identified specific gene signatures for monocytic or granulocytic lineage-commitment by differential expression (DE) gene analysis between the monocytic and granulocytic branches of the GMP developmental trajectory (Figure 2H) generated from publicly available single-cell RNAseq data sets (GSE70235). Monocytic signature genes were highly enriched (*CD34, CSF1R, CFP, LY86*), while most of the granulocytic signature genes were down-regulated (*S100A8/9, ETS1, PGLYRP1, CD63, CEACAM1*) in GMPs isolated from WD fed as compared to CD fed mice (Figure 2I). These data were in accordance with a linear support vector regression analysis indicating an increase in monocytic lineage potential after WD (Figure 2J). We next analyzed which Gene Ontology (GO) terms and KEGG pathways were enriched in the DE genes after WD (Figure S2F). GO enrichment analysis using the genes that were up-regulated after WD, was further visualized as a

biological GO enrichment network (Figure 2K). Together, these analyses identified that genes up-regulated in GMPs from WD fed mice were associated with regulatory processes involved in hematopoiesis, cell proliferation, programmed cell death, metabolic processes and cell migration, as well as immune cell differentiation, leukocyte activation and immune processes involved in cellular stress responses and WNT receptor signaling (*i.e.,* hedgehog signaling pathway, IL-17 signaling pathway/Th17 differentiation, JAK-STAT signaling pathway, viral carcinogensis). Furthermore, transcription factor (TF) binding prediction analysis suggested that the WD-induced transcriptional skewing of GMPs into the monocyte direction was influenced by the transcription factors *GATA1-3, KLF4* and *TCF7*, certain *FOS, JUN*, and *ATF* family members, as well as *CREB1* and *EGR1* (functional involvement in induction of cell reprogramming, survival, and monocytic differentiation; Figure S2G). Together, these data indicate that the bone marrow compartment strongly responds to WD feeding with an increased myelopoiesis and a deep transcriptional reprogramming of myeloid precursor cells.

**WD causes functional reprogramming of myeloid precursor cells**

Given that WD feeding induced a global transcriptional reprogramming of GMPs, we next asked whether the immune responses towards a model ligand (*i.e.,* LPS) were altered *in vivo*. We fed *Ldlr*$^{-/-}$ mice the CD or WD for 4 weeks and intravenously injected LPS (10µg) or vehicle control (PBS) six hours before sacrifice and cell harvest (Figure 3A). LPS treatment of WD fed mice resulted in higher serum levels of most inflammatory cytokines and chemokines when compared to LPS treated mice fed CD (Figure 3B). Furthermore, the activation status of circulating and splenic monocyte subsets as measured by CD86 surface expression was amplified in LPS challenged WD fed mice

when compared to LPS challenged CD fed mice (Figure 3C, D). In contrast, CD86 surface expression in circulating and splenic granulocytes remained unchanged upon LPS treatment in WD fed mice (Figure 3C, D). These data suggest that WD feeding primes the immune system for systemic and local innate immune responses towards an inflammatory trigger.

To better define how WD-induced reprogramming of myeloid precursor cells influences the responses towards LPS, we next performed RNASeq analysis of sorted GMPs from mice that received the WD or CD and were exposed to an LPS or PBS challenge 6 hours prior to sacrifice. Pearson correlation matrix analysis of the transcriptional changes in the different treatment conditions showed that WD treatment mimicked the effects of LPS in CD fed mice and that LPS and WD treatment had synergistic effects on gene transcription (Figure 3E), which was also evident when the gene expression differences are represented as a Volcano plot (Figure 3F). Detailed inspection of the most highly differentially expressed genes revealed that many type I interferon (IFN) response genes, including *IRG1*, *CCL5*, *CXCL10*, *GBP4 and 6*, *IRF1*, *IFI203*, were expressed more strongly in WD fed mice receiving LPS compared to CD fed mice receiving LPS. Furthermore, functional annotation enrichment analysis for GO terms and KEGG pathways demonstrated that the WD primed GMPs for LPS responses and led to a shift towards increased inflammatory signaling (Figure 3G). As expected, WD feeding was associated with a down-regulation of cholesterol biosynthesis pathways, innate immune defense responses and in particular anti-viral responses were greatly enriched (Figure S3B-E). These data demonstrate that WD feeding results in GMP priming that triggers enhanced innate immune responses towards LPS.

**Long-lasting reprogramming of GMPs following WD**

We observed that a four weeks CD resting period normalized the WD-induced systemic cytokine response in the circulation, whereas the innate immune responses of myeloid cells remained enhanced and qualitatively altered (Figure 1C-E). We next sought to decipher to what extend the transcriptional changes in GMPs induced by WD persisted and how this affected the GMP responsiveness towards LPS. Mice fed CD, 4 weeks WD or 4 weeks WD followed by 4 weeks CD were intravenously challenged with LPS or PBS six hours prior to cell isolation, and then RNAseq analysis of purified GMPs was performed (Figure 4A; Figure S4A). Principal component analysis (Figure 4B) and Pearson correlation matrix analysis (Figure 4C) demonstrated that the transcriptomic reprogramming of GMPs after WD did not reverse in GMPs isolated from mice that had been switched to CD feeding for 4 weeks after WD feeding. A detailed assessment of genes that were regulated by the different dietary interventions demonstrated that certain gene sets were commonly expressed in GMPs isolated from WD as well as from WD fed mice that were rested on CD (Figure S4B). Interestingly, other sets of genes were only expressed in GMPs isolated from WD fed mice or only in GMPs isolated from WD fed mice that were rested on CD. We therefore analyzed which GO terms were commonly or selectively enriched in GMPs isolated from mice subjected to the different dietary interventions (Figure S4B). We found genes that were commonly up-regulated in GMPs from both feeding groups were associated with immune processes and lymphocyte activation (Figure S4B). This shows that even after resting mice from WD, GMPs remain in an activated state and keep a persistent transcriptionally remodeled state that is qualitatively distinct from that of cells isolated from CD fed mice.

Next, we aimed to better define the effects of the different dietary interventions on the functional responses of GMPs to *in vivo* LPS challenge. To visualize changes in the

transcriptional programs that occur after LPS challenge in mice fed CD, 4 weeks WD or 4 weeks WD followed by 4 weeks CD, we performed a co-expression network analysis of all genes with changed expression and visualized the changes as standardized gene expression (Figure 4D-F) or fold changes (Figure 4G-I). This analysis revealed that LPS challenge induced the expression of many more gene clusters in mice fed a WD than CD (Figure 4D, E, G, H) and that the LPS response remained altered in GMPs in mice shifted back to CD after WD feeding (Figure 4F, I). We found that numerous of the LPS regulated genes under CD feeding (Figure 4G) were more strongly up-regulated or in some cases even counter-regulated after WD feeding (Figure 4H). Of note, the LPS-induced gene expression in GMPs isolated from mice that had been switched to CD after WD feeding remained altered and largely similar to the WD only conditions (Figure 4I). We additionally performed TF motif enrichment analysis for DE genes in GMPs isolated from mice subjected to the different dietary and LPS challenges. Genes that were higher expressed upon LPS challenge in WD fed compared to CD fed mice exhibited increased abundance for TF binding sites such as RELA/B, NF-kB1/2, STAT1-4/5a,b/ 6, IKZF1/2 as well as IRF1/3/7/8/9, JUN/FOS and KDM4A,B,C,D in the regulatory regions (Figure S4C-E). In addition, assessment of TF binding prediction analysis for genes expressed in GMPs that were not regulated upon LPS treatment in CD fed mice, but up-regulated upon WD and LPS treatment revealed a defined set of transcriptional regulators including NRF1, YY1/YY2, STAT1/2, IRF1/3/5/7/8/9, GATA1-3, JUN/FOS and MEF2A/C/D (Figure S4D). Notably, TFs that were the most enriched in GMPs from mice treated with WD and LPS remained partly enriched in GMPs from mice fed 4 weeks WD followed by 4 weeks CD and treated with LPS, including amongst others, YY1/2, NRF1/2, MEF2A/C/D and ERG (Figure S4E).

Moreover, fold-change analysis of the most strongly LPS-induced genes revealed a preservation of the distinctive type I IFN signature in GMPs (for example GBP6, CXCL10, Ifi44, Ifit3b) from the WD fed and CD rested mice compared to those mice fed only CD (Figure 5A). Whereas the marked increase in circulating monocyte quantities seen after WD feeding was not maintained in CD rested mice (Figure 5B), the quantity of circulating granulocytes was found to be further increased in CD rested mice compared to WD fed mice. Additionally, splenic granulocytes displayed a more activated phenotype upon WD feeding, and tended to remain more activated in WD and CD rested mice compared to CD fed mice (Figure 5D).

To test whether WD changes the overall accessibility of chromatin we performed global epigenetic profiling of open chromatin by Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-Seq) in GMPs isolated from 4 weeks WD or CD fed mice (Figure 5E-F). While WD clearly induced changes in the chromatin landscape, these changes were discrete in comparison to LPS stimulation (Figure 5E). Next, differentially regulated peaks were clustered and presented as a heatmap (Figure 5F). We identified 8 clusters of peaks with some genomic loci showing similar chromatin status in GMPs derived from WD-fed or LPS-treated mice but differing from GMPs from CD-fed control mice (Figure 5 F, clusters 3). Genes associated with this cluster were linked to GO-terms such as ‘leukocyte differentiation’ or ‘T cell activation’. Furthermore, peaks particularly enriched in GMPs from WD-fed mice (cluster 4) were associated with genes linked to ‘IL-6 production’ and the ‘JAK/STAT pathway’. Of interest, enhancer regions including Tet2 and Tlr4 were found to be more open in both GMPs isolated from WD fed mice and WD followed by CD fed mice ($p < 0.05$, $|FC|>2$; Figure 5G). In contrast, gene loci for Oxysterol-binding protein-related protein 3 (Osbpl3) and Abca1 ($p < 0.05$, $|FC|>2$) were more closed in GMPs isolated from mice fed WD, but reopened in

mice fed WD followed by CD and were thus comparable to the CD fed mice (Figure S4F). These data indicate that WD can effectively reprogram GMPs both on the transcriptional and epigenetic level. Importantly, the broad transcriptional changes in GMPs from LPS injected mice fed WD and then CD for 4 weeks more closely resembled the altered responses seen in GMPs isolated from WD fed and LPS challenged mice suggesting persistence of a long-lasting reprogrammed state after removal of the WD insult.

**The IL-1 pathway is part of WD-induced GMP reprogramming**

The training effect we observed in our murine model might be further modulated by genetic variation in major pathways involved. To determine such genetic effects and identify those pathways, we performed a functional trained immunity (FTI-) QTL study in human monocytes. We used the pro-inflammatory oxidized form of LDL (oxLDL) that has been associated with WD and is known to induce trained immunity in human monocytes (Bekkering et al., 2014). We subjected adherent monocytes to oxLDL or control, rested cells after a wash step for 5 days and re-stimulated with LPS using monocytes from 122 healthy individuals (Figure 6A) from the 200 FG cohort (Li et al., 2016). Training of monocytes with oxLDL resulted in an increased responsiveness to LPS re-stimulation, compared to cells pre-incubated with culture medium. The FTI-QTL analysis identified several single nucleotide polymorphisms (SNPs) in the putative regulatory gene regions of (1) *PYCARD*, the gene encoding the inflammasome adapter ASC, and (2) *IL1RAP*, the gene encoding the IL-1 receptor antagonist (IL-1ra). Importantly, several of these genetic variants showed a significant effect on the capacity of oxLDL to induce trained immunity (Figure 6B, D: Manhattan plots representing clustered reference SNPs (rs) in the proximity of the gene coding regions for *PYCARD* or *IL1RA*P), assessed by the

diminished release of TNF ($p < 10^{-4}$, linear regression model; Figure 6C, E) or IL-6 ($p < 10^{-4}$, linear regression model; Figure 6E). A role of IL-1ra in oxLDL mediated innate immune training could be confirmed in an independent cohort of humans. Addition of recombinant IL-1ra during the oxLDL-induced training period resulted in lower responses upon LPS challenge after the training period, as analyzed by TNF and IL-6 levels in the culture supernatants (Figure 6F). Given the important role of inflammasomes in regulating IL-1 and the suggestion by the companion papers that IL-1 is involved in cellular reprogramming, we further tested the effects of IL-1 blockage on the systemic WD-induced inflammatory response *in vivo*. We subjected *Ldlr*$^{-/-}$ mice to recombinant IL-1ra or PBS vehicle control during a 4 weeks course of CD or WD feeding. The results demonstrated that IL-1 blockade during WD feeding did not significantly alter hypercholesterolemia, yet it tended to (Figure 6G). Additionally, IL-1 blockade reduced the WD-induced systemic inflammation, demonstrated by the decrease of inflammatory cytokines (Figure 6H), as well as the acute phase protein SAA3 (Figure 6I). Together, these data provide evidence for the importance of the inflammasome and the downstream IL-1R signaling pathway for the induction of trained immunity by oxLDL in vitro or by WD in vivo.

**The NLRP3 inflammasome is required for sensing WD**

Previous work has identified that oxLDL is capable of activating the NLRP3 inflammasome and that NLRP3 activation has been shown to be part of the pathogenesis of atherosclerosis (Duewell et al., 2010; Sheedy et al., 2013). We hypothesized that triggering of the NLRP3 inflammasome by pro-inflammatory factors arising after WD could contribute to trained immunity, as observed by the hematopoietic response in mice. We therefore generated *Nlrp3*$^{-/-}$*/Ldlr*$^{-/-}$ mice and analyzed their

systemic and cellular responses to WD. Strikingly, WD did not induce peripheral monocytosis in *Nlrp3*$^{-/-}$/*Ldlr*$^{-/-}$ mice compared to WD fed *Ldlr*$^{-/-}$ mice (Figure 7A). WD furthermore failed to induce GMP activation as assessed by CD86 surface expression (Figure 7B) and WD did not induce increased GMP proliferation in *Nlrp3*$^{-/-}$/*Ldlr*$^{-/-}$ mice (Figure 7C). Additionally, the WD-induced augmentation of the LPS induced systemic inflammatory cytokine responses observed in *Ldlr*$^{-/-}$ were largely blunted in *Nlrp3*$^{-/-}$/*Ldlr*$^{-/-}$ mice, especially in inflammasome dependent cytokines (Figure 7E). Transcriptional profiling of GMPs isolated from WD or CD fed *Ldlr*$^{-/-}$ or *Nlrp3*$^{-/-}$/*Ldlr*$^{-/-}$ mice combined with computational network analysis (Figure 7F) and rank plot visualization (Figure 7G) demonstrated that the WD-induced transcriptional reprogramming of GMPs was mostly dependent on NLRP3. In agreement with the lack of inflammatory effects of WD, *Nlrp3*$^{-/-}$/*Ldlr*$^{-/-}$ mice showed markedly reduced atherosclerotic plaque size after 8 weeks of WD feeding (Figure 7D).

Together, our data strongly support a key function of the NLRP3 inflammasome in myeloid cell reprogramming in the context of WD feeding.

## Discussion

Genetic and lifestyle factors are key drivers of chronic non-communicable inflammatory diseases, which currently represent the vast majority of death burden in Western societies (Mega et al., 2015). Importantly, an unhealthy lifestyle, which includes factors such as noxious diets, little exercise and sleep, and unfavorable environmental triggers, can independently of the genetic risk strongly increase the rate of incident cardiovascular events and increase the susceptibility for other chronic inflammatory diseases (Khera et al., 2016). Therefore, it is of great importance to better define how these factors mechanistically influence inflammatory processes.

Alterations in diets can have deleterious effects on immune responses and drive the development of a range of inflammatory diseases (Thorburn et al., 2014). Yet, whether and how 'inflammatory' diets can provoke trained immunity leading to potentially long-lasting alterations of immune responses was unknown. Early studies documented that WD feeding of rabbits induced enhanced cytokine production towards LPS stimulation in cells within the aortic tissue suggesting that WD could prime for augmented innate immune responses (Clinton et al., 1991; Fleet et al., 1992). Here, we have investigated whether WD consumption can trigger trained immunity and whether this results in modified secondary immune responses in a mouse model for atherosclerosis, because wild-type mice do not spontaneously develop atherosclerosis, unless serum cholesterol levels are raised experimentally by genetic deletion of LDLR or APOE. The LDLR deficient mouse model is relevant to human atherosclerosis as patients with LDLR mutations are highly susceptible for the development of early atherosclerosis. In our studies we have focused on deciphering inflammatory effects of WD in the early, initiation phase of atherosclerosis before significant amounts of plaque have deposited in the vessel walls in $Ldlr^{-/-}$ mice. This minimized the potential contribution of plaque-derived immune cells and also putative changes that might occur due to plaque regression that can occur after changing the diet from WD to CD. Our studies revealed that while WD provoked only transient systemic inflammatory responses, it led to long-lasting alterations of myeloid cell responses towards different innate immune stimuli. These studies thus establish that not only microbes but also complex sterile inflammatory triggers, such as a WD, can induce trained immunity.

To better understand how WD induces trained immunity, and because other studies have linked hypercholesterolemia to reprogramming of hematopoietic stem cells (Murphy et al., 2011; Seijkens et al., 2014; van Kampen et al., 2014; Yvan-Charvet et al.,

2010), we performed functional and transcriptional analyses of GMPs, the proximate precursors of circulating myeloid subsets. Notably, blood leukocyte counts and in particular the numbers of circulating monocytes were shown to be tightly associated with hypercholesterolemia and incidence of cardiovascular events (Friedman et al., 1974; Swirski and Nahrendorf, 2013). Furthermore, inflammatory blood monocytes that arise from activated hematopoietic precursors during WD feeding develop into disease promoting macrophages in atherosclerotic plaques and likely other organs (Swirski et al., 2007). Our studies revealed that WD induces GMP proliferation and skewing towards the development of activated and potentially harmful monocytes, which was accompanied by long-lived transcriptional and epigenetic reprogramming. While the ATAC sequencing results suggest that epigenetic changes are induced by WD feeding and might contribute to the transcriptomic reprogramming, other mechanisms of epigenetic regulation may also be induced. Indeed, opening of chromatin is only one aspect of the epigenetic processes that control transcription, and it is likely that there are further mechanisms that drive the complex long-term transcriptomic changes we have demonstrated. Nevertheless, our studies suggest that the immune system misinterprets a WD as a threat to the host as it sets in motion powerful anti-infectious mechanisms such as hematopoiesis and maintenance of a hyper-responsive state with the generation of myeloid cells that are programmed to respond to secondary inflammatory triggers more potently.

Recent studies support the idea that cell-fate decisions and a potential reprogramming already happens at the level of non-committed HSPCs, particularly under conditions of systemic stress induced by bacterial or viral infections (Boettcher and Manz, 2017) but also during hyperlipidemia and hyperglycemia (Brasacchio et al., 2009; Nagareddy et al., 2014; Singer et al., 2014). The work by the International Trained Immunity Consortium

(INTRIM), reported in this issue of Cell now demonstrates that trained immunity can be induced on the level of HSPCs. *Mitroulis et al.* show that the trained immunity model ligand β-glucan induces metabolic and transcriptional rewiring in HSPCs through IL-1R signaling, which protects the hematopoietic system from chemotherapy-induced myeloablation.

By showing that WD induces a 'memory' in myeloid progenitors our work raised the question of which mechanisms are involved in sensing the complex inflammatory trigger. LDL cholesterol overload, oxidation of LDL (oxLDL) as well as fatty acids can all induce a continuous state of intracellular stress and activate innate immune pathways (Dorrestein et al., 2014; Sheedy et al., 2013). Additionally, WD-induced microbial alterations could also drive for innate immune cell re-programming, which could be tested using germ-free mice in future studies. However, the metabolic, neuroendocrine, and molecular control mechanisms that mediate the biological effects of WD are wide-ranging and their individual contributions are difficult to decipher *in vivo.* We thus performed a reductionist *in vitro* trained immunity QTL study in human monocytes to exploit human genetic variability as a discovery approach. Intriguingly, this analysis revealed that oxLDL-induced trained immunity was influenced by several independent genetic polymorphisms in the putative regulatory region of *PYCARD*, the gene encoding the principal inflammasome adapter protein ASC, and also in the gene for the endogenous IL-1 inhibitory molecule *IL-1RAP*. Together with a large array of studies linking NLRP3 to the inflammatory response in metabolic diseases (Duewell et al., 2010; Rhoads et al., 2017; Sheedy et al., 2013), this provided evidence for a potential key role of the NLRP3 inflammasome pathway in the induction of trained immunity under conditions of WD feeding. Strongly supporting this hypothesis, we observed an almost complete protection from WD-induced changes in systemic inflammation, induction of

hematopoiesis and functional and transcriptomic reprogramming of GMPs in *Ldlr*$^{-/-}$*/NLRP3*$^{-/-}$ mice. Future studies using cell specific deletions will have to be performed to clarify the important question as to which cell type activation of the inflammasome is required to induce the observed epigenetic reprogramming effects we have demonstrated.

Nevertheless, NLRP3 qualifies as a principal WD insult sensor that mediates most aspects of the WD-induced inflammation and trained immunity we have demonstrated in bone marrow precursor cells. In support of this hypothesis, patients with primary hypercholesterolemia reportedly have increased levels of inflammasome-dependent IL-18 and evidence for systemic inflammation (Messal et al., 2006; Narverud et al., 2011). Additionally, cholesterol overload in hematopoietic progenitors could also lead to a sustained state of intracellular stress, linked to increased inflammatory signaling and a potential long-term reprogramming. Supportively, the studies by *Bekkering et al.* reported in this issue of Cell show that dysbalanced cholesterol biosynthesis with an accumulation of mevalonate can drive trained immunity induction. On the other hand it appears plausible that NLRP3 could also instigate inflammatory pathologies towards a range of sterile danger signals not limited to WD, as numerous triggers of relevance in human chronic inflammatory pathologies have been linked to NLRP3 activation (Heneka et al., 2013; Hornung et al., 2008; Martinon et al., 2006).

Our study together with work by *Mitroulis et al.* in this issue of Cell implies that an inflammasome-mediated product, such as IL-1β, is likely the central endogenous mediator of the mechanisms resulting in the induction of trained immunity. Indeed, early work had already established that injections of IL-1β before experimental infection can prevent lethality from infections (Van der Meer et al., 1988). Additionally, IL-1β, produced by activated adipose tissue macrophages in obesity, was shown to interact with the IL-

1R expressed on CMPs, which stimulated the proliferation of GMPs (Nagareddy et al., 2014). Hence, it is likely that IL-1β is the key messenger generated by NLRP3 following WD-mediated activation. It is further tempting to speculate that these mechanisms are part of the clinical benefit observed in the recent CANTOS trial, which tested IL-1β blockade in humans at cardiovascular risk (Ridker et al., 2017).

Curiously, functional annotation analysis of the transcriptomic changes demonstrated that GMPs responded in a fashion that shares features of a viral infection characterized by the appearance of a strong type I IFN signature. It would be interesting to better understand the complex interplay between type I IFNs signaling and the IL-1β signaling after WD exposure, given that type I IFNs can regulate inflammasome activation and IL-1β synthesis (Labzin et al., 2016). Furthermore, a causal link between WD-induced type I IFN production and the risk of virus infections and auto-immunity is conceivable and should prospectively be investigated in rodent models of infections or in human populations. In addition, we detected signaling pathways in GMPs associated with cell proliferation, anti-apoptosis, loss of stem cell quiescence and favored lineage differentiation into monocytes (Olsson et al., 2016), which potentially raises GMP life span and facilitates memory induction. These cell phenotypes were largely maintained even four weeks after shifting mice to CD feeding.

In summary, WD feeding can result in a remodeled cellular compartment in the bone marrow leading to increased proliferation, skewing of hematopoiesis and generating exaggerated responds to exogenous and endogenous triggers of inflammation. By integrating the inflammatory effects of WD, the NLRP3 inflammasome appears to be critical for this response. Hence, this may pave the road for new therapeutic CVD interventions, such as the use of small molecule inhibitors that block the NLRP3 signaling pathway.

## Author Contribution

Conceptualization, A.C. and E.L.; Methodology, A.C., P.G. and E.L.; Software, P.G., K.H., K.B., C.S., K.K., T.U., J.S.S. and JLS; Investigation, A.C., M.A.R.L., K.P., P.D., D.B., M.P., M.O., S.M., V.K., T.E., A.S., L.A.B.J., N.P.R., Y.L., L.A.G. and M.G.N.; Resources, M.G.N, J.L.S. and E.L.; Writing- Original Draft, A.C., P.G., M.G.N. and E.L; Writing-Review & Editing, A.C., P.G., J.L.S., M.L.F. and E.L.; Visualization, A.C. and E.L.; Supervision, E.L.; Project Administration, E.L.; Funding Acquisition, A.C., M.L.F., P.D., J.L.S., M.G.N., E.L.

## Acknowledgements

EL, MLF and AC are supported by grants from NIH (1R01HL112661 and R01HL101274). EL is supported by DFG grants (SFB 670 and TRR 83, 57) and an ERC Consolidator grant (InflammAct). EL and PD are supported by the SFB1123 grant by the DFG. JLS is supported by DFG grants (SFB645, 704), the European Union's Horizon 2020 research and innovation program under grant agreement No.733100 (SYSCID), and a grant from the Federal Ministry for Economic Affairs and Energy (BMWi Project FASTGENOMICS). EL is a co-founder of IFM-Therapeutics and a member of its scientific advisory board. MGN is supported by an ERC Consolidator Grant (#310372) and a Spinoza grant of the Netherlands Organization for Scientific Research (NWO). NPR is supported by a Dr. Dekker grant from the Netherlands Heart Foundation (2012T051). NPR, MGN, and LABJ received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 667837. MGN, JLS and EL are members of the Excellence Cluster ImmunoSensation.

## References


Bekkering, S., Quintin, J., Joosten, L.A.B., der Meer, van, J.W.M., Netea, M.G., and Riksen, N.P. (2014). Oxidized low-density lipoprotein induces long-term proinflammatory cytokine production and foam cell formation via epigenetic reprogramming of monocytes. Arteriosclerosis, Thrombosis, and Vascular Biology *34*, 1731–1738.

Boettcher, S., and Manz, M.G. (2017). Regulation of Inflammation- and Infection-Driven Hematopoiesis. Trends Immunol.

Brasacchio, D., Okabe, J., Tikellis, C., Balcerczyk, A., George, P., Baker, E.K., Calkin, A.C., Brownlee, M., Cooper, M.E., and El-Osta, A. (2009). Hyperglycemia induces a dynamic cooperativity of histone methylase and demethylase enzymes associated with gene-activating epigenetic marks that coexist on the lysine tail. Diabetes *58*, 1229–1236.

Cao, X. (2016). Self-regulation and cross-regulation of pattern-recognition receptor signalling in health and disease. Nat Rev Immunol *16*, 35–50.

Clinton, S.K., Fleet, J.C., Loppnow, H., Salomon, R.N., Clark, B.D., Cannon, J.G., Shaw, A.R., Dinarello, C.A., and Libby, P. (1991). Interleukin-1 gene expression in rabbit vascular tissue in vivo. The American Journal of Pathology *138*, 1005–1014.

Dorrestein, P.C., Mazmanian, S.K., and Knight, R. (2014). Finding the Missing Links among Metabolites, Microbes, and the Host. Immunity *40*, 824–832.

Duewell, P., Kono, H., Rayner, K.J., Sirois, C.M., Vladimer, G., Bauernfeind, F.G., Abela, G.S., Franchi, L., Núñez, G., Schnurr, M., et al. (2010). NLRP3 inflammasomes are required for atherogenesis and activated by cholesterol crystals. Nature *464*, 1357–1361.

Fleet, J.C., Clinton, S.K., Salomon, R.N., Loppnow, H., and Libby, P. (1992). Atherogenic diets enhance endotoxin-stimulated interleukin-1 and tumor necrosis factor gene expression in rabbit aortae. J. Nutr. *122*, 294–305.

Friedman, G.D., Klatsky, A.L., and Siegelaub, A.B. (1974). The leukocyte count as a predictor of myocardial infarction. N. Engl. J. Med. *290*, 1275–1278.

GBD 2015 Mortality and Causes of Death Collaborators (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet *388*, 1459–1544.

Hansson, G.K., and Hermansson, A. (2011). The immune system in atherosclerosis. Nat Immunol *12*, 204–212.

Heneka, M.T., Kummer, M.P., Stutz, A., Delekate, A., Schwartz, S., Vieira-Saecker, A., Griep, A., Axt, D., Remus, A., Tzeng, T.-C., et al. (2013). NLRP3 is activated in Alzheimer's disease and contributes to pathology in APP/PS1 mice. Nature *493*, 674–

678.

Hornung, V., Bauernfeind, F., Halle, A., Samstad, E.O., Kono, H., Rock, K.L., Fitzgerald, K.A., and Latz, E. (2008). Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. Nat Immunol *9*, 847–856.

Khera, A.V., Emdin, C.A., Drake, I., Natarajan, P., Bick, A.G., Cook, N.R., Chasman, D.I., Baber, U., Mehran, R., Rader, D.J., et al. (2016). Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N. Engl. J. Med. *375*, 2349–2358.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Meth *12*, 357–360.

Labzin, L.I., Lauterbach, M.A.R., and Latz, E. (2016). Interferons and inflammasomes: Cooperation and counterregulation in disease. J. Allergy Clin. Immunol. *138*, 37–46.

Li, Y., Oosting, M., Deelen, P., Ricaño-Ponce, I., Smeekens, S., Jaeger, M., Matzaraki, V., Swertz, M.A., Xavier, R.J., Franke, L., et al. (2016). Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. Nat Med *22*, 952–960.

Martinon, F., Pétrilli, V., Mayor, A., Tardivel, A., and Tschopp, J. (2006). Gout-associated uric acid crystals activate the NALP3 inflammasome. Nature *440*, 237–241.

Mega, J.L., Stitziel, N.O., Smith, J.G., Chasman, D.I., Caulfield, M.J., Devlin, J.J., Nordio, F., Hyde, C.L., Cannon, C.P., Sacks, F.M., et al. (2015). Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. The Lancet *385*, 2264–2271.

Messal, El, M., Beaudeux, J.-L., Drissi, A., Giral, P., Chater, R., Bruckert, E., Adlouni, A., and Chapman, M.J. (2006). Elevated serum levels of proinflammatory cytokines and biomarkers of matrix remodeling in never-treated patients with familial hypercholesterolemia. Clinica Chimica Acta *366*, 185–189.

Murphy, A.J., Akhtari, M., Tolani, S., Pagler, T., Bijl, N., Kuo, C.-L., Wang, M., Sanson, M., Abramowicz, S., Welch, C., et al. (2011). ApoE regulates hematopoietic stem cell proliferation, monocytosis, and monocyte accumulation in atherosclerotic lesions in mice. J Clin Invest *121*, 4138–4149.

Nagareddy, P.R., Kraakman, M., Masters, S.L., Stirzaker, R.A., Gorman, D.J., Grant, R.W., Dragoljevic, D., Hong, E.S., Abdel-Latif, A., Smyth, S.S., et al. (2014). Adipose Tissue Macrophages Promote Myelopoiesis and Monocytosis in Obesity. Cell Metabolism *19*, 821–835.

Narverud, I., Ueland, T., Nenseter, M.S., Retterstøl, K., Telle-Hansen, V.H., Halvorsen, B., Ose, L., Aukrust, P., and Holven, K.B. (2011). Children with familial hypercholesterolemia are characterized by an inflammatory imbalance between the tumor necrosis factor Î± system and interleukin-10. Atherosclerosis *214*, 163–168.

Netea, M.G., Joosten, L.A.B., Latz, E., Mills, K.H.G., Natoli, G., Stunnenberg, H.G., O'Neill, L.A.J., and Xavier, R.J. (2016). Trained immunity: A program of innate immune memory in health and disease. Science *352*, aaf1098–aaf1098.

Netea, M.G., Quintin, J., and der Meer, van, J.W.M. (2011). Perspective. Cell Host Microbe *9*, 355–361.

Olsson, A., Venkatasubramanian, M., Chaudhri, V.K., Aronow, B.J., Salomonis, N., Singh, H., and Grimes, H.L. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. Nature *537*, 698–702.

Rhoads, J.P., Lukens, J.R., Wilhelm, A.J., Moore, J.L., Mendez-Fernandez, Y., Kanneganti, T.-D., and Major, A.S. (2017). Oxidized Low-Density Lipoprotein Immune Complex Priming of the Nlrp3 Inflammasome Involves TLR and FcγR Cooperation and Is Dependent on CARD9. J Immunol *198*, 2105–2114.

Ridker, P.M., Everett, B.M., Thuren, T., MacFadyen, J.G., Chang, W.H., Ballantyne, C., Fonseca, F., Nicolau, J., Koenig, W., Anker, S.D., et al. (2017). Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. N. Engl. J. Med. *377*, 1119–1131.

Seijkens, T., Hoeksema, M.A., Beckers, L., Smeets, E., Meiler, S., Levels, J., Tjwa, M., de Winther, M.P.J., and Lutgens, E. (2014). Hypercholesterolemia-induced priming of hematopoietic stem and progenitor cells aggravates atherosclerosis. Faseb J *28*, 2202–2213.

Sheedy, F.J., Grebe, A., Rayner, K.J., Kalantari, P., Ramkhelawon, B., Carpenter, S.B., Becker, C.E., Ediriweera, H.N., Mullick, A.E., Golenbock, D.T., et al. (2013). CD36 coordinates NLRP3 inflammasome activation by facilitating intracellular nucleation of soluble ligands into particulate ligands in sterile inflammation. Nat Immunol *14*, 812–820.

Singer, K., DelProposto, J., Morris, D.L., Zamarron, B., Mergian, T., Maley, N., Cho, K.W., Geletka, L., Subbaiah, P., Muir, L., et al. (2014). Diet-induced obesity promotes myelopoiesis in hematopoietic stem cells. Molecular Metabolism *3*, 664–675.

Swirski, F.K., and Nahrendorf, M. (2013). Leukocyte Behavior in Atherosclerosis, Myocardial Infarction, and Heart Failure. Science *339*, 161–166.

Swirski, F.K., Libby, P., Aikawa, E., Alcaide, P., Luscinskas, F.W., Weissleder, R., and Pittet, M.J. (2007). Ly-6Chi monocytes dominate hypercholesterolemia-associated monocytosis and give rise to macrophages in atheromata. J Clin Invest *117*, 195–205.

Thorburn, A.N., Macia, L., and Mackay, C.R. (2014). Diet, Metabolites, and "Western-Lifestyle" Inflammatory Diseases. Immunity *40*, 833–842.

Van der Meer, J.W., Barza, M., Wolff, S.M., and Dinarello, C.A. (1988). A low dose of recombinant interleukin 1 protects granulocytopenic mice from lethal gram-negative infection. Proc Natl Acad Sci USA *85*, 1620–1623.

van Kampen, E., Jaminon, A., van Berkel, T.J.C., and Van Eck, M. (2014). Diet-induced (epigenetic) changes in bone marrow augment atherosclerosis. J Leukoc Biol *96*, 833–841.

Yvan-Charvet, L., Pagler, T., Gautier, E.L., Avagyan, S., Siry, R.L., Han, S., Welch, C.L., Wang, N., Randolph, G.J., Snoeck, H.W., et al. (2010). ATP-binding cassette

transporters and HDL suppress hematopoietic stem cell proliferation. Science *328*, 1689–1693.

Zimmer, S., Grebe, A., and Latz, E. (2015). Danger signaling in atherosclerosis. Circulation Research *116*, 323–340.

## Figure legends

**Figure 1- WD feeding induces systemic inflammation and functional reprogramming. A.** Schematic of dietary interventions. Female *Ldlr*$^{-/-}$ mice were fed either CD, WD for 4 weeks or WD for 4 weeks followed by CD for 4 weeks (WD>CD). **B.** Systemic serum cholesterol in response to dietary intervention in *Ldlr*$^{-/-}$ mice. **C.** Heat-map representing normalized serum cytokine levels from mice fed as indicated **D.** Bone marrow cells or **E.** splenic CD11b$^{+}$ monocytes, isolated from *Ldlr*$^{-/-}$ mice following dietary intervention treated *ex vivo* with vehicle or different TLR stimuli for 6h. Log 2 transformed data represented as spider plots for the following stimulations: Pam3Csk4, LPS, R848 and CpG. For (B) n = 6-10 animals; for (C) n = 3-5 animals per group; $\pm$ SEM, $p < 0.05$ vs. CD; vs. un-stimulated cells (D, E). Experiments were performed twice independently and data are representative of a single experiment. See also Figure S1 and Table S1.

**Figure 2- WD induces hematopoiesis and transcriptional reprogramming of GMPs. A.** Total counts of the indicated blood cell populations in CD or WD fed (4 weeks) female *Ldlr*$^{-/-}$ mice. **B.** Relative numbers (%) and **C.** activation status of circulating myeloid subsets isolated from female *Ldlr*$^{-/-}$ mice. **D.** Percentage of hematopoietic precursor cells as indicated in female *Ldlr*$^{-/-}$ mice fed either CD or WD (4 weeks). **E.** PCA of RNAseq data of GMPs isolated from CD or WD fed mice. PCA is based on variable genes (non-

adj. p-value < 0.05, n = 4672) **F.** Gene and sample wise hierarchical clustering based on the 1000 genes with the highest variance within the dataset of GMPs purified from WD or CD fed mice. Gene expression values are z-score standardized. **G.** MA-plot showing DE genes in GMPs of WD or CD fed mice. DE genes (|FC|>1.5, non-adj. p-value < 0.05) are colored in red (up-regulated in WD) and blue (down-regulated in WD) and notable genes are highlighted. **H.** Trajectory analysis of single-cell RNAseq data (GSE70235) with computational clustering (top) representing the expression of Csf1R or S100a8, overlaid onto the developmental trajectory to identify monocytic or granulocytic lineage determination. Cells of the monocytic (turquoise) and granulocytic (dark green) branches were used to determine signature genes to test for lineage potential in I and J. **I.** Expression differences of monocytic (turquoise) and granulocytic (dark green) signature genes in GMPs from mice fed as indicated. **J.** Enrichment of monocytic and granulocytic signatures in GMPs isolated from CD and WD fed mice, and computationally inferred by linear support vector regression analysis. Enrichments are significant (non-adj. p-value < 0.001). **K.** GO term enrichment network analysis of differentially expressed genes from GMPs isolated from WD and CD fed mice. Each dot represents a significantly enriched GO term and connections indicate shared genes between GO terms. Significance (FDR) is indicated by color (lower FDR: more intense color) and size (lower FDR: bigger nodes/ thicker borders) of nodes (up-regulated genes) or borders (down-regulated genes); ± SEM, $p < 0.05$ vs. CD; for (A-C) n = 3-5 animals per group. Experiments were performed twice independently and data are representative of a single experiment. See also Figure S2.

**Figure 3- WD primes for inflammatory responses to LPS. A.** Schematic representation of the WD and LPS manipulations. Female *Ldlr*$^{-/-}$ mice were fed CD or

WD for 4 weeks. 6 hours prior to sacrifice mice were intravenously challenged by LPS or PBS. **B.** Serum cytokine levels in LPS-treated CD or WD fed mice were normalized to levels in PBS control groups and represented as fold-change. Activation status of **C.** circulating and **D.** splenic myeloid subsets from female *Ldlr*$^{-/-}$ mice treated as indicated. **E.** Pearson correlation analysis of top 1000 DE genes of transcriptomes from GMPs isolated from mice treated as indicated. **F.** Volcano plot indicating transcriptomic changes between CD/LPS and WD/LPS. Significantly up-regulated (FC >1.5, FDR-adj. p-value < 0.05, red) and down-regulated (FC < -1.5, FDR-adj. p-value < 0.05, blue) genes are shown and the most significantly regulated genes are highlighted. **G.** Number of significantly up- (red) and down- (grey) regulated GMP genes within immune system associated GO terms in WD/LPS compared to CD/LPS treated mice. n = 3-8 animals per group in B; n = 3 animals in C-D; ± SEM, p < 0.05 vs. CD (B-D); vs. CD-GMPs (H). Experiments were performed twice independently and data are representative of a single experiment. See also Figure S3.

**Figure 4- WD induces long-lasting reprogramming in GMPs. A.** Schematic of diet and LPS manipulations. **B.** PCA of genes with highest variance (non-adj. p-value < 0.05, n = 4672). **C.** Pearson correlation analysis of top 1000 DE genes of transcriptomes from GMPs isolated from mice treated as indicated. **D. – I.** Co-expression network analysis (genes = 4,360, correlation > 0.85) based on highly correlated genes among the 11306 expressed genes. To allow identification of specific gene signatures z-score transformed average gene expression (D-F) or fold changes in gene expression (G-I) of respective conditions were overlaid onto the co-expression network. See also Figure S4.

**Figure 5- WD induces epigenetic reprogramming of GMPs. A.** FC-FC plot comparing the LPS response of GMPs with amplitude of difference in LPS responses visualized as color code (red: up-regulated genes; blue: down-regulated genes). **B.** Total counts of the indicated blood cell populations in female mice fed as shown. Activation status of circulating **C.** and splenic **D.** myeloid subsets from mice treated as indicated. **E.** Volcano plot displaying open chromatin loci as determined by ATACseq in GMPs isolated from mice treated as indicated. Average signal is represented as log2 fold change. Significantly up- (non-adj. p-value < 0.05, log2FC > 1) and down-regulated (non-adj. p-value < 0.05, log2FC < -1) peaks are shown. **F.** Hierarchical clustering, standardized and visualized as a heatmap show the significant differentially accessible genomic loci (p-value < 0.05) in GMPs isolated from mice treated as indicated. **G.** Coverage of ATACseq signal for TET2 (top panel) and TLR4 loci (bottom panel). n = 3-5 for groups in (B-D); ± SEM, p < 0.05 vs CD. Experiments were performed twice independently and data are representative of a single experiment. See also Figure S4.

**Figure 6- *PYCARD* and *IL-1RAP* SNPs influence oxLDL-induced training effects.** **A.** Schematic overview of the *in vitro* FTI-QTL protocol. Manhattan plots representing reference single nucleotide polymorphisms (rs) in the **B.** *PYCARD* gene locus or **D.** in the *IL1RAP* gene locus. **C.** TNF production capacity in different genotyping groups from B and **E.** TNF and IL-6 production capacity between different genotyping groups from D. **F.** Monocyte oxLDL training was performed in presence or absence of recombinant IL-1ra followed by LPS stimulation. Shift in cytokine levels represented as fold-change (oxLDL/+/- IL-1ra). **G.** Systemic serum cholesterol levels, **H.** cytokines and **I.** acute phase response in *Ldlr*$^{-/-}$ mice treated as indicated. n = 3-4 mice per group (G-I); means ± SEM, p < 0.05 vs. non-IL-1ra treatment (RPMI only) (F); vs. *i.* CD and *ii.* PBS

treatment (G-I).

**Figure 7- NLRP3 dependent myeloid progenitor priming. A.** Total counts of circulating cell subsets from female *Ldlr*$^{-/-}$ and *Nlrp3*$^{-/-}$*/Ldlr*$^{-/-}$ mice treated as indicated. **B.** Activation status and **C.** proliferative capacity of GMPs isolated from *Ldlr*$^{-/-}$ or *Nlrp3*$^{-/-}$*/Ldlr*$^{-/-}$ mice fed as indicated. **D.** Atherosclerotic plaque lesion size in WD fed (8 weeks) female *Ldlr*$^{-/-}$ or *Nlrp3*$^{-/-}$*/Ldlr*$^{-/-}$ mice. **E.** Serum cytokine response 6 hours post LPS injection in female *Ldlr*$^{-/-}$ and *Nlrp3*$^{-/-}$*/Ldlr*$^{-/-}$ mice fed as indicated. **F.** Co-expression network analysis of transcriptional changes induced by WD in female *Ldlr*$^{-/-}$ and *Nlrp3*$^{-/-}$*/Ldlr*$^{-/-}$ mice. Fold changes are overlaid onto co-expression networks. **G.** Rank plot visualization of fold changes of genes represented in F. n = 2-5 for groups in A-C, E (experiments were performed twice independently and data are representative of a single experiment), n = 5-9 animals in D; ± SEM, $p < 0.05$ vs. CD (A-C, E), vs. *Ldlr*$^{-/-}$ (D).

**Legends for Supplemental Figures**

**Supplemental Figure 1- A.** Systemic lipoprotein levels in female *Ldlr*$^{-/-}$ mice fed as shown. **B.** Body weights of female *Ldlr*$^{-/-}$ mice treated as indicated **C.** Acute phase reactant levels in female *Ldlr*$^{-/-}$ mice treated as indicated. Cytokine and chemokine response of **D.** bone marrow cells or **E.** splenic CD11b$^{+}$ monocytes isolated from female *Ldlr*$^{-/-}$ mice following diet manipulation and treated *ex vivo* with vehicle or different TLR stimuli for 6h as indicated. n = 5 mice per group in B and n = 6-9 mice per group in C; ± SEM, $p < 0.05$ vs. CD (C), vs. un-stimulated cells (D, E). Figure S1. related to Figure 1.

**Supplemental Figure 2-**

**A.** FACS gating strategy. **B.** Relative numbers and **C.** activation status of splenic myeloid subsets isolated from female *Ldlr*$^{-/-}$ mice treated as indicated **D.** FACS sorting gating strategy for the different myeloid progenitor subsets. **E.** Schematic representation of the computational analysis as indicated **F.** Functional enrichment analysis showing overrepresentation of genes up-regulated in GMPs isolated from female *Ldlr*$^{-/-}$ mice treated as indicated in Gene Ontology terms or KEGG pathways. **G.** Transcription factor binding prediction analysis of significantly up-regulated genes in GMPs isolated from female *Ldlr*$^{-/-}$ mice fed WD or CD. Motif enrichment analysis was performed on 20kb area centered around the transcriptionally start site of up-regulated genes. $Log_2FC$ in genes expressed in GMPs and visualized for predicted transcriptional regulators. Boxes represent TFs with high motif similarity (FDR-adj. p-value < 0.001). Colors indicate the normalized enrichment score of motif enrichment. n = 3-5 animals per group; ± SEM, $p < 0.05$ (B, C). Figure S2. related to Figure 2.

**Supplemental Figure 3-**

**A.** Representative flow cytometry histogram analysis from myeloid cell subsets from CD and WD fed female *Ldlr*$^{-/-}$ mice. **B. - E.** Functional enrichment analysis showing overrepresentation of genes up-regulated (B, C) or down-regulated (D, E) in GMPs isolated from mice treated as indicated. Figure S3. related to Figure 3.

**Supplemental Figure 4-**

**A.** Schematic describing the procedure of computational analysis of RNAseq data for Figure 4 and S4. **B.** Functional analysis of genes commonly up-regulated in GMPs from mice treated as indicated. Top five enriched gene ontology terms of these gene lists are shown. DE genes: non-adjust. p-value < 0.05, FC> 1.5. **C. - E.** Transcription factor binding prediction of regulated genes in GMPs isolated from mice treated as indicated

and selected genes as depicted in the schematics. Selection of input genes for analysis: C: Genes up-regulated in WD/LPS vs CD/LPS (FC > 1.5, non-adj. p-value < 0.05), D: Genes not regulated in LPS response under CD (-1.3 < FC < 1.3), but significantly up-regulated in LPS response upon WD feeding in GMPs (FC > 1.3, non-adj. p-value < 0.05) and E: Genes not regulated in LPS response under CD (-1.3 < FC < 1.3), but significantly up-regulated in LPS response upon WD>CD feeding in GMPs (FC > 1.3, non-adj. p-value < 0.05). Motif enrichment analysis was performed on a 20kb area centered around the transcription start site of up-regulated genes. $Log_2$ FC is visualized for predicted transcriptional regulators from conditions as indicated. Boxes represent TFs with high motif similarity (FDR-adj. p-value < 0.001). Color indicates the normalized enrichment score of motif enrichment. Selected clusters are shown. **F.** Coverage of ATACseq signal for OSBPL3 (left panel) and ABCA1 loci (right panel). Figure S4. related to Figure 4 and Figure 5.

**Supplemental Table 1:** Serum cytokine levels in response to CD, 4 weeks WD, or 4 weeks WD followed by a resting period of 4 weeks. Table S1. related to Figure 1.

## STAR methods

### Contact for Reagent and Resource sharing

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding author Eicke Latz (eicke.latz@uni-bonn.de).

## Experimental Model and Subject details

### *Mice*

Experiments were approved by the Institutional Animal Care and Use Committees of the University of Massachusetts Medical School and performed according to local ethics regulations (IACUC 1945, UMass Medical School), and in accordance with the NIH guidelines. *C57Bl/6J* were purchased from The Jackson Laboratory. *Ldlr*$^{-/-}$ and *Nlrp3*$^{-/-}$ were originally purchased from The Jackson Laboratory and kept in house. *Nlrp3*$^{-/-}$*/Ldlr*$^{-/-}$ were bred in house. All mice were previously backcrossed over ten generations to the *C57Bl6/J* background. For all *in vivo* animal studies (Western Diet feeding studies, LPS challenge, *in vivo* cell proliferation) age (8 weeks of age) and sex-matched female wild type, *Ldlr*$^{-/-}$*, Nlrp3*$^{-/-}$*/Ldlr*$^{-/-}$ *mice* were used with at least 3 mice per genotype. If mouse numbers per genotype were < 3, experiments were at least performed twice independently. During experimental settings mice had ad libitum access to food and water, and were housed under a 12 hour light-dark cycle.

### *Human subjects*

Experiments were conducted according to the principles expressed in the Declaration of Helsinki as Ethics Statement. Healthy individuals of Western European descent gave written informed consent to donate venous blood to use for research (Li et al., 2016). 122 volunteers were between 23-73 years old, and consisted of 77% males and 23% females. During all *in vitro* culture assays human isolated monocytes were kept in RPMI 1640 culture medium supplemented with 50 µg/mL gentamycin, 2 mM glutamax and 1 mM pyruvate at 37°C in a 5%$CO_2$ atmosphere.

**Methods Details**

***Mouse in vivo studies***

To induce hyperlipidemia/ hypercholesterolemia female mice were fed a Western diet (Teklad 88137) consiting of 17.3% protein, 21.2% fat (saturated fat 12.8%, monounsaturated fat 5.6%, polyunsaturated fat 1%) and 48.5% carbohydrates. Chow diet (Prolab Isopro RMH 30; LabDiet) consisted of 25% protein, 14% fat (ether extract) and 60% carbohydrates. To study diet effects on long-term reprogramming of myeloid (progenitor) subsets, female mice were fed a WD for 4 weeks, and subsequently subjected to regular chow diet for additional 4 weeks. To study diet-induced innate immune cell priming *in vivo* female mice received an intravenous injection of PBS (vehicle control) or *E. coli* derived ultrapure LPS (0111:B4; 10µg/ mouse) six hours prior to sacrifice. To study the involvement of the IL-1/IL-1R signaling pathway in innate immune cell reprogramming *in vivo* female mice were daily i.p. injected with PBS (vehicle control) or Anakinra (human recominant IL-1ra; 10 mg/kg BW) during a 4 weeks course of CD or WD feeding. Blood, spleen and bone marrow (BM) were collected for single cell isolation and additional flow cytometry analysis, for immunohistochemical tissue preparation, for RNA isolation and further gene expression analysis, or for additional *in vitro* stimulation experiments.

***Isolation of mouse plasma lipoproteins by FPLC***

For lipoprotein separation by FPLC, total sera cholesterol levels were first determined with standard enzymatic assays (see below for kit description). Then a 50 µl sera aliquot was pre-warmed to 37°C for 5 min. followed by filtration through a PVDF 0.45 µm membrane filter. The filtered samples (20 µl) were subsequently fractionated by fast

performance liquid chromatography (FPLC) gel filtration on a Superose 6 PC 3.2 /30 column at 4°C (GE Healthcare, Uppsala, Sweden). The elution fractions were monitored using absorbance at 280 nm, with a constant flow of 40 µl/min and fractions (40 µl) were collected beginning 18 min. after sample injection. Cholesterol in each fraction was measured by the *Infinity cholesterol reagents* (Thermo Fisher Scientific, VA, USA), and the area under the curve for vLDL, LDL and HDL was determined in comparison to a standard of known amounts of human vLDL, LDL and HDL run in parallel.

***Mouse serum MultiPlex Cytokine Measurements***

Multi-cytokines in mouse sera were measured with the BioPlex Mouse Cytokine 23-plex assay together with BioPlex mouse cytokine single-plexes of IL-18 and VEGF (Bio-Rad Laboratories, Hercules, CA) on a Bio-Plex 200 system powered by Luminex xMAP Technology. In some experiments a Procarta Plex 10-plex bead assay from ThermoFisher and measured on a MAGPIX instrument using Luminex xMAP Technology. The analysis was performed according to the manufacturer's protocol with the following modifications: use of only half of the volumes, adapted incubation times (serum: 60-30-30), serum centrifugation for 10 min. at 13200 x rpm before analysis.

***Measurement of mouse serum acute phase proteins***

Acute phase proteins in mouse sera were measured using commercial ELISA kits for SAA3 (Invitrogen/ ThermoFisher Scientific) and CRP (R&D) in accordance with the manufacturer's instructions.

***Murine White Blood Cell (WBC) counts***

Total WBC counts in freshly isolated murine blood were performed by collecting blood in EDTA-coated tubes and analyzing them using a hematology cell counter (Abaxis HM5).

***Flow cytometry of circulating and splenic leukocyte subsets in mouse***

Total number of myeloid cells, monocyte subsets and neutrophils were identified as depicted in Figure S2A. Blood was collected via cardiac puncture into ethylene-diamine-tetraacetate (EDTA) lined tubes and immediately placed on ice. Mice were perfused wth 10ml of sterile phosphate buffered saline (PBS) via the left ventricle before harvesting and further processing of organs. Spleens were collected into RPMI1640/ 0.1% fetal calf serum (FCS) containing vials and stored on ice until further processing. All following steps were performed on ice. Red blood cells (RBCs) were lysed (3 times for 5 minutes) and leukocytes were centrifuged, washed and resuspended in flow cytometry buffer for staining (PBS containing 0.5% bovine serum albumine (BSA) and 2mM EDTA). Spleens were minced into small pieces, gently crushed and filtered twice with PBS through a 70µm cell strainer. RBC lysis was performed once for 3 minutes before resuspending splenic cells in flow cytometry buffer and antibody staining with a cocktail of antibodies against CD45 PerCp-Cy5.5, CD11b BV510, CD11c BV421, Ly6C-APC, Ly6G-FITC, F4/80 PE, CD64 PE-Cy7, MHCII APC-Cy7, CD86 Pacific Blue (all from BioLegend). Monocytes were identified as $CD45^{hi}$ $CD11b^{+}$ $Ly6G^{-}$ and further separated into $Ly6C^{hi}$, $Ly6C^{int}$ and $Ly6C^{-}$; neutrophils were identified as $CD45^{hi}$ $CD11b^{+}$ $Ly6G^{+}$ $Ly6C^{lo}$. Cells were analyzed by LSRII using the FACS Diva software (BD). All flow cytometry data were analyzed using FlowJo Software (Tree Star Inc.).

***In vivo proliferation assay***

For *in vivo* proliferation studies, mice were injected intraperitoneally with 1 mg of 5-(ethynyl-2'-deoxyuridine) (EdU) 24 hours prior to sacrifice. In preparation for subsequent flow cytometry analysis, cell populations were immunostained as described above, fixed in 4% paraformaldehyde (PFA) and further treated for EdU staining following the manufacturer's instructions of the Click-iT System (ThermoScientific). AlexaFluor647-conjugated azide was used to detect incorporated EdU. Stained cells were analyzed by LSRII using the BD Diva software and analyzed by FlowJo. Proliferation was quantified and expressed as percentage of EdU-positive cells.

***Ex vivo stimulation of bone marrow cells and splenic monocytes***

Bone marrow and splenic single cell suspensions were prepared as described above. Splenic CD11b+ monocytes were further purified by positive selection using a cocktail of CD11b magnetic beads (Miltenyi), reaching a purity of 95%. Cells in 96-well flat-bottom wells (100.000 cells/ well) were cultured in RPMI1640 supplemented with L-Glut, 10%FCS, 1% Ciprofloxacin at 37°C in a 5%$CO_2$ atmosphere. After a 2 hour resting period in the incubator (37°C, 5%$CO_2$), cells were stimulated with TLR-ligands for 6 hours: Pam3CysK4 (10ng/ml; TLR2 ligand), LPS (10ng/ml; TLR4 ligand), R848 (10µg/ml; TLR7/8 ligand), and CpG (10µM; TLR9 ligand). Subsequently, supernatants were collected for cytokines measurement by MultiPlex bead array (ThermoFisher).

***Flow cytometric cell sorting of murine bone marrow hematopoietic stem cells***

Hematopoietic stem cell precursor cells (HSPCs) and myeloid progenitor cells including common myeloid progenitor cells (CMPs) and granulocyte-monocyte progenitor cells (GMPs) from bone marrow were analyzed by flow cytometry as described in Figure S2D. Bone marrow was harvested from femurs and tibias by flushing bones with cold Hank's

balanced salt solution (HBSS) containing 0.1% BSA and 2mM EDTA. RBC lysis was performed once for 5 minutes, followed by antibody staining with a cocktail of antibodies against lineage-committed cells (CD3e, CD19, CD45R (B220), CD11b, TER-119, CD2, CD8b, CD4, Ly6G; all FITC-labeled and from eBioscience), Sca-1 Pacific Blue, c-kit (CD117) APC-Cy7, CD34 APC, CD16/CD32 (FcyRII/III) PerCp-Cy5.5 (all from BioLegend). HSPCs were identified as lin- Sca-$1^+$ c-$kit^+$; CMPs as $lin^-$ $Sca1^-$ c-$kit^+$ $CD34^{int}$ $FcγRII/III^{int}$; GMPs as $lin^-$ $Sca1^-$ c-$kit^+$ $CD34^{int}$ $FcγRII/III^{hi}$. Isolated and stained bone marrow cells were either directly analyzed by LSRII (BD), or flow sorted by FACS Aria II for further RNA isolation and RNA/ ATAC-sequencing, or *in vitro* assays.

***RNA Isolation***

$1.7 \times 10^5$–$16 \times 10^5$ GMPs were sorted and subsequently lysed in TRIZOL (Invitrogen) and total RNA was extracted using the miRNAeasy kit (Quiagen) according to the manufacturer's protocol. The precipitated RNA was solved in RNase-free water. The quality of the RNA was assessed by measuring the ratio of absorbance at 260 nm and 280 nm using a Nanodrop 2000 Spectrometer (Thermo Scientific) and by visualization of 28S and 18S band integrity on a Tapestation 2200 (Agilent).

***Generation of cDNA Libraries and Sequencing***

Total RNA was converted into libraries of double-stranded cDNA molecules as a template for high-throughput sequencing using the Illumina TruSeq RNA Sample Preparation Kit v2. Briefly, mRNA was purified from 100–500 ng of total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in Illumina proprietary fragmentation buffer. First strand cDNA was synthesized using random oligonucleotides and SuperScript II. Second

strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities and enzymes were removed. After adenylation of 3′ ends of DNA fragments, Illumina adaptor oligonucleotides were ligated to prepare for hybridization. DNA fragments with ligated adaptor molecules were selectively enriched using Illumina PCR primers in a 15 cycles PCR reaction. Size-selection and purification of cDNA fragments with preferentially 200 bp in insert length was performed using SPRIBeads (Beckman-Coulter). Size distribution of cDNA libraries was measured using the Agilent high sensitivity DNA assay on a Tapestation 2200 (Agilent). cDNA libraries were quantified using KAPA Library Quantification Kits (Kapa Biosystems). After cluster generation on a cBot, 75 bp single read sequencing was performed on a HiSeq1500 and de-multiplexed using CASAVA v1.8.2.

***RNAseq Data Analysis***

Pre-processing of RNAseq data was performed by a standardized and reproducible pipeline based on the Docker system (Docker image is available via docker hub, limesbonn/hisat2). Briefly, alignment to the mouse reference genome mm10 from UCSC was conducted by Hisat2 (Hisat2, 2.0.6) (Kim et al., 2015)using standard settings Aligned BAM files were imported into Partek Genomics Suite (PGS) software (version 6.6, Partek) for further analysis, and mRNA quantification was performed using mm10 RefSeq Transcripts (2015-11-03) as annotation file. Afterwards, raw gene counts were normalized using the DESeq2 algorithm in R (package version 1.14.1). After DESeq normalization, the normalized read counts were imported back into PGS and floored by setting all read counts less than 1 to a value of exactly 1. Subsequently, we removed all

genes that had a maximum overall group mean below 10. After filtering the data comprised 11306 present genes. Correlation was calculated as Pearson correlation of normalized read counts. Differentially expressed genes were identified by a two-way ANOVA model taking the diet and treatment as factors, with genes with highest variance (n = 4672) being defined as genes with a significant p-value (p-value < 0.05) for the interaction of diet and treatment. Principal component analysis was performed using PGS. Heatmaps and several plots were generated using the ggplot2 plotting library. Functional annotation of gene sets was performed using the clusterProfiler R package (version 3.2.11). RNAseq data from *Nlrp3*$^{-/-}$ / *Ldlr*$^{-/-}$ mice were normalized and analyzed separately from data obtained from *Ldlr*$^{-/-}$ mice following the same workflow as described above. Network visualization of Gene Ontology Enrichment Analysis (GOEA) was performed with Cytoscape (version 3.3) and the BiNGO plugin (Version 2.44) with a FDR-corrected hypergeometric *p*-value of 0.001. The Cytoscape plugin enrichment map (Version 1.2) displayed the GOEA results as a network of GO-terms, and for assistance of annotation we used the word cloud cytoscape app (version 3.1.0).

***Co-expression network analysis***

In order to perform a topological mapping of global changes in gene expression we used a co-expression network approach based on correlation. The gene expression matrices containing the normalized read counts for the 11306 present genes were imported into BioLayout (Version 3.3) and co-regulation networks were generated with a Pearson correlation cutoff of 0.85. Co-expressed genes were visualized as a network (n = 4,360 genes) by Cytoscape applying a Force-directed layout. To identify the topology of the co-expression network the positive standardized gene expression (scaled and centered

average expression) was calculated and mapped onto the co-expression network. In addition, we mapped the expression differences in a LPS response under CD as fold change to the network. To visualize the diet dependent effects on a LPS response we overlaid the color-coded fold changes between WD/LPS and CD/LPS or WD>CD/LPS and CD/LPS, respectively.

***Transcription factor prediction analysis***

In order to predict potential regulators of the observed gene expression changes we performed transcription factor binding prediction (TFBP) in the region 20kb around the transcriptional start site (TSS). To identify transcriptional regulators of genes up-regulated in a WD vs. CD LPS response, genes with a significant difference in gene expression (FC >1.5, non-adj. p-value < 0.05) between WD/LPS and CD/LPS were used for TFBP. In addition, we examined transcriptional regulators of genes only regulated in a WD/LPS or an WD>CD/LPS but not in a LPS response under standard chow diet. Therefore, genes were used for TFBP with no considerable gene expression changes between CD/LPS and CD (-1.3 < FC < 1.3) and a significant difference between WD/LPS and CD/LPS or WD>CD/LPS and CD/LPS, respectively. For the gene-based motif enrichment analysis, we used iRegulon (version 1.3) in combination with the provided 10K position weight matrices (PWMs) motif collection. To identify only highly enriched transcription factor predictions we filtered the results by an enrichment score of 3 and only reported transcription factors with a maximum false discovery rate (FDR-adj. p-value) of less than 0.001. Resulting transcriptional regulators were grouped into clusters with high motif similarity using the STAMP algorithm with iRegulon standard

parameters. The transcriptional regulators in their respective clusters were ordered by decreasing fold change and their gene expression was visualized as log2FC.

### *Assessment of GMP-commitment using single cell RNAseq data combined with support vector regression*

To investigate the effect of nutrition on GMP commitment towards either the monocytic or the granulocytic developmental branch, we utilized a publicly available dataset, which comprised single cell transcriptome data of various murine hematopoietic progenitor cells (GSE70240). Here, we used the available normalized count data of only GMPs that were found either in the monocytic or granulocytic cluster (109 cells). Next, we constructed a developmental trajectory using Monocle 2. For this purpose, we transformed the expression levels in transcripts per million mapped reads (TPM) to estimate mRNA counts per cell (RPC) using the *Census* implementation of Monocle 2. We set the lower detection limit to an RPC of 1 and defined variable genes by taking advantage of a dispersion plot for which the mean expression cutoff was set to 1 and the dispersion cutoff was set to 1.75. Based on these variable genes, we computed a trajectory, which consisted of three branches with two of them were either monocytic (enriched for Csfr1) or granulocytic (enriched for S100a8) committed. Next, we calculated differentially expressed (DE) genes between the GMPs either committed towards the monocytic or the granulocytic lineage using Seurat. As tests for differential expression, we utilized a likelihood-ratio test for single cell gene expression, a standard AUC classifier and Tobit-test. When limiting the evaluation to genes which showed on average at least 1-fold difference (natural log-scale) between the two populations and which were detected in at least 25% of the cells in either of the two populations, all three

tests resulted in the same number of DE genes (195 genes). For support vector regression, we defined the calculated DE genes as support vectors. The signature file was generated by population-wise calculation of the total TPM count per gene and filtering for the defined support vectors. The mixture file was based on variable genes (ANOVA, non-adj. p-value < 0.05) across all conditions in the present dataset. Support vector regression was performed with 100 permutations and by disabling quantile normalization.

***Cell preparation for ATAC-sequencing***

For ATAC-sequencing, 5.000 - 50.000 flow sorted GMPs were collected in ice-cold HBSS containing 0.1%BSA and 2mM EDTA, and immediately processed following previously published protocols. In brief, sorted cells were spun down at 500 x g for 5 minutes at 4°C, washed once in cold PBS, and spun down in 50 µl cold lysis buffer at 500 x g for 10 minutes at 4°C. Immediately thereafter the transposition reaction was started by adding Nextera's Tn5 Transposase (TDE1) in reaction buffer. The transposition reaction mix was incubated for 30 minutes at 37°C, DNA was purified using a Quiagen MinElute PCR purification kit. ATAC-seq libraries were purified using a PCR purification MinElute kit (Quiagen) and quantified using KAPA library quantification kits (Kapa biosystems) and a D1000 assay on a Tapestation 2200 (Agilent). Libraries were sequenced in a 50 bp single read run on a HiSeq 1500 (Illumina).

***Analysis of ATACseq data***

For Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) data analysis, short reads were aligned to the mouse genome version mm10 with Bowtie

v1.1.1. Duplicate reads were flagged with Picard v1.134 prior to peak calling with MACS2 v2.1.0.20140616. Consensus peak regions across all samples were generated with the 'reduce' function of the Bioconductor GenomicRanges v1.26.3 package, followed by read counting per sample, which was done with bedtools multiBamCov v2.25.0. Blacklisted analysis regions with coordinate adjustments by the LiftOver tool to mm10, as well as regions that were not consistently present across samples of the same analysis group were excluded from the analysis. Differentially accessible regions were determined with the DESeq2 v1.14.1 package and mapped to nearest genes with the ChIPseeker v1.10.3 package. To read coverage tracks of ATAC signal we combined replicate samples (2 biological replicates, for WD>CD/PBS) and created normalized (to 10,000,000 reads) bigWig files using the HOMER (version 4.6) command *makeUCSCfile.* For visualization, the read coverage tracks were smoothed by averaging over a sliding window of 200 bases and all tracks for a given region were scaled to the highest overall peak and visualized by the Gviz package (version 1.18.1). Global changes in chromatin accessibility are visualized as heatmap and similarly accessible regions are clustered by hierarchical clustering using pheatmap package v 1.0.8 in R.

***Histology of mouse atherosclerotic lesions***

Mice were anesthesized by an overdose isoflurane and perfused through the left cardiac ventricle with PBS containing 1% PFA. Hearts were dissected, fixed overnight in 4%PFA frozen in OCT Tissue Tek Compound (Sakura), and cut into 4 µM thick serial sections. To measure plaque volume in the aortic root, plaque area was measured for each valve for consecutive sections at 20 µm intervals that covered the entire lesion. H&E staining was performed as described previously (Duewell et al., Nature 2010). Slides were

analyzed in a blinded manner using a Leica DM3000 light microscope (Leica Microsystems) coupled to a computerized morphometric system (Leica Qwin 3.5.1).

***Isolation of human PBMCs and in vitro trained immunity model***

Venous blood was drawn from the cubital vein of healthy volunteers into 10 mL EDTA tubes (Monoject). PBMCs were isolated by density centrifugation of blood diluted 1:1 in pyrogen-free saline over Ficoll-Paque (Pharmacia Biotech). Cells were washed twice in saline and resuspended in RPMI 1640 culture medium (Invitrogen) supplemented with 50 µg/mL gentamycin, 2 mM glutamax (GIBCO), and 1 mM pyruvate (GIBCO). Cells were counted in a Coulter counter (Coulter Electronics) and adjusted to 5 x $10^6$ cells per milliliter. A 100 µL volume was added into flat-bottom 96-well plates (Corning) and cells were incubated at 37˚C. After 1 hour, cells were washed three times with 200 µL warm PBS to remove non-adherent cells. Subsequently, monocytes were incubated with culture medium only (negative control) or 10 µg/mL oxLDL for 24 hours in 10% pooled human serum. In an additional experimental setup Anakinra (recombinant IL-1ra; 10ug/ml) was added during oxLDL incubation for 24 hours. Cells were washed once with 200 µL warm PBS and incubated for 5 days in culture medium with 10% human pooled serum. The medium was changed once on day 3 of incubation. On day 6, the supernatant was discarded and the cells were re-stimulated with 200 µL culture medium or 10 ng/mL *E. coli* LPS (serotype 055:B5, Sigma-Aldrich). After 24 hours, supernatants were collected and stored at -20˚C until cytokine measurement. Cytokine production was measured in supernatants using commercial ELISA kits for TNF and IL-6 (R&D systems) in accordance with the manufacturer's instructions.

***Mapping of genetic variants in human PBMCs***

Raw cytokine levels were first log-transformed. Then, the ratio of cytokine measurements between the second and the first stimulation was computed to capture the enhanced trained response or tolerant response. Individuals of 200FG cohort were genotyped using the Illumina HumanOmniExpressExome-8 v1.0 and the data was imputed, as previously described. In total there were 122 samples with both DNA and cytokine measurements. We focused on the genetic effect of polymorphisms within a window of 250kb around *PYCARD*, *NLRP3*, *CASPASE1* and *IL-1RAP* genes on the oxLDL-induced trained immunity. The ratio of log-transformed cytokine data was mapped to genotype data using a linear model with age and gender as covariates.

**Quantification and Statistical analysis**

Statistical parameters including the exact value of n, the definition of center, dispersion and precision measures (mean±SEM) and statistical significance are reported in the Figures and Figure Legends. All data are expressed as the mean± SEM. Data is judged to be statistically significant when $p < 0.05$ by two-tailed Student's t test (if normally distributed); nonparametric data are analyzed using a Mann-Whitney *U*-test. To compare several groups, a one-way ANOVA (with Dunnet's post-test) or repeated measures ANOVA with multiple testing correction are used. In figures asterisks denote statistical significance (*, $p < 0.05$; **, $p < 0.01$; ***). Statistical analysis was performed in GraphPad PRISM 7 (Graph Pad *Software* Inc.).

**Data and software availability**

Raw data files for the RNA sequencing and ATAC sequencing analysis have been deposited in the NCBI Gene Expression omnibus under accession number GEO: GSE97926.

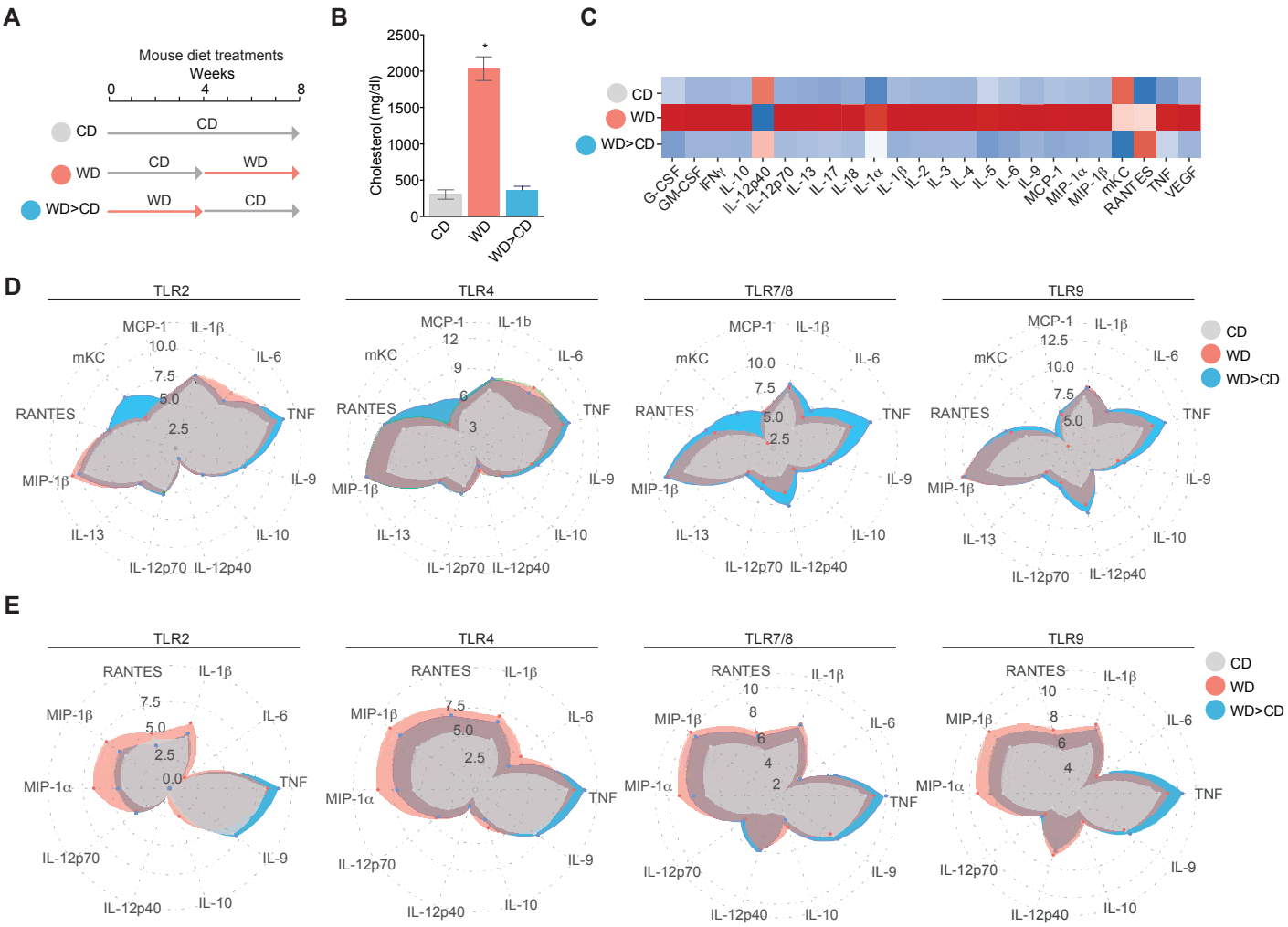

A
Mouse diet treatments
Weeks
0
4
8
CD
CD
WD
CD
WD
WD>CD
WD
CD
B
Cholesterol (mg/dl)
2500
2000
1500
1000
500
0
*
CD
WD
WD>CD
C
CD
WD
WD>CD
G-CSF
GM-CSF
IFNγ
IL-10
IL-12p40
IL-12p70
IL-13
IL-17
IL-18
IL-1α
IL-1β
IL-2
IL-3
IL-4
IL-5
IL-6
IL-9
MCP-1
MIP-1α
MIP-1β
mKC
RANTES
TNF
VEGF
D
TLR2
TLR4
TLR7/8
TLR9
MCP-1
IL-1β
IL-1b
mKC
IL-6
RANTES
TNF
MIP-1β
IL-9
IL-13
IL-10
IL-12p70
IL-12p40
CD
WD
WD>CD
E
TLR2
TLR4
TLR7/8
TLR9
RANTES
IL-1β
MIP-1β
IL-6
MIP-1α
TNF
IL-12p70
IL-9
IL-12p40
IL-10
CD
WD
WD>CD

**Figure 1**

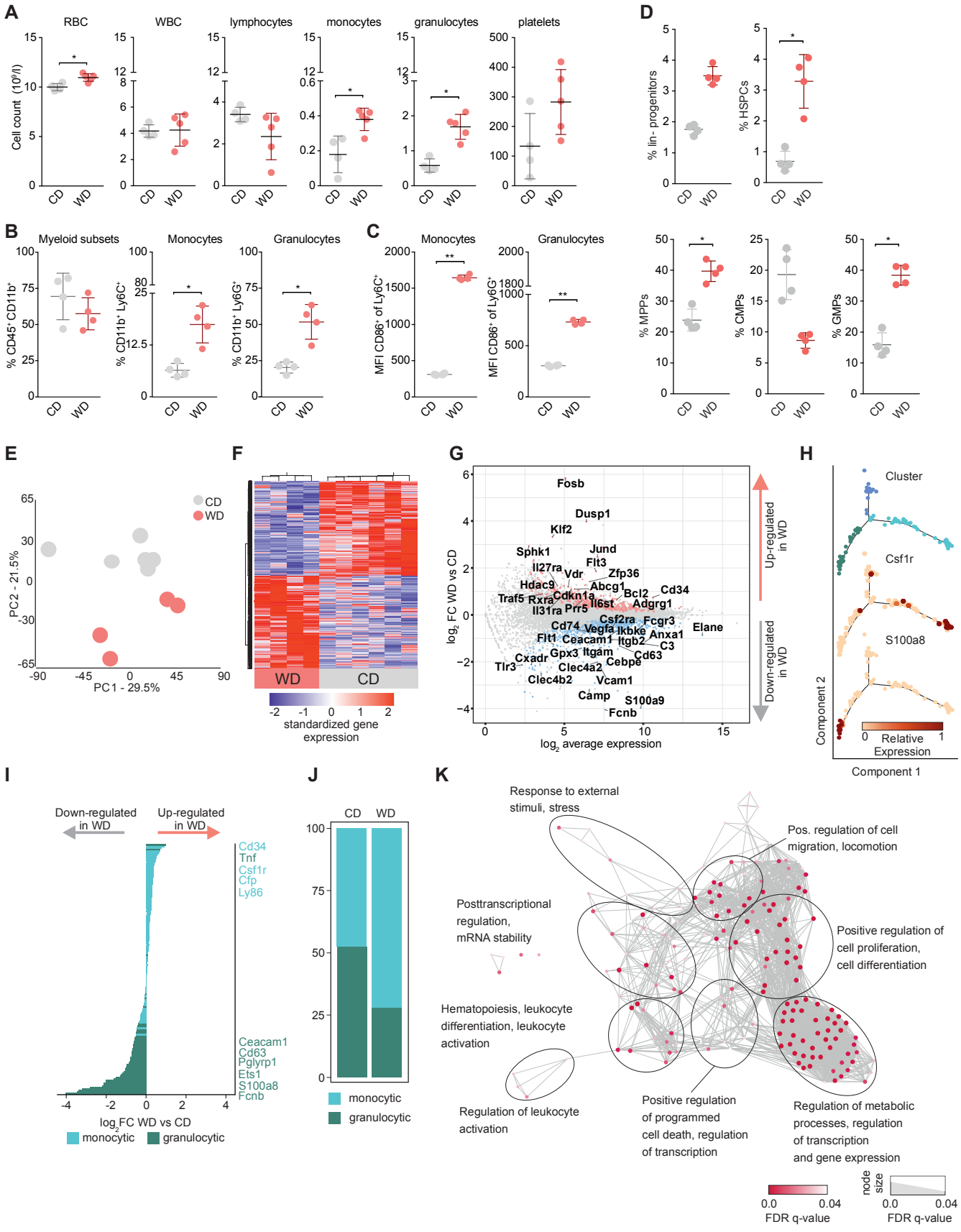

A
RBC
WBC
lymphocytes
monocytes
granulocytes
platelets
Cell count (10⁹/l)
CD
WD
D
% lin- progenitors
% HSPCs
B
Myeloid subsets
Monocytes
Granulocytes
% CD45⁺ CD11b⁺
% CD11b⁺ Ly6C⁺
% CD11b⁺ Ly6G⁺
C
Monocytes
Granulocytes
MFI CD86⁺ of Ly6C⁺
MFI CD86⁺ of Ly6G⁺
% MPPs
% CMPs
% GMPs
E
CD
WD
PC2 - 21.5%
PC1 - 29.5%
F
WD
CD
standardized gene expression
G
Fosb
Dusp1
Klf2
Sphk1
Jund
Il27ra
Flt3
Vdr
Zfp36
Hdac9
Abcg1
Cdkn1a
Bcl2
Cd34
Traf5
Rxra
Il6st
Adgrg1
Il31ra
Prr5
Csf2ra
Fcgr3
Cd74
Vegfa
Ikbke
Elane
Anxa1
Flt1
Ceacam1
Itgb2
C3
Gpx3
Itgam
Cd63
Cxadr
Cebpe
Tlr3
Clec4a2
Clec4b2
Vcam1
Camp
S100a9
Fcnb
log₂ FC WD vs CD
log₂ average expression
Up-regulated in WD
Down-regulated in WD
H
Cluster
Csf1r
S100a8
Component 2
Component 1
Relative Expression
I
Down-regulated in WD
Up-regulated in WD
Cd34
Tnf
Csf1r
Cfp
Ly86
Ceacam1
Cd63
Pglyrp1
Ets1
S100a8
Fcnb
log₂FC WD vs CD
monocytic
granulocytic
J
CD
WD
monocytic
granulocytic
K
Response to external stimuli, stress
Pos. regulation of cell migration, locomotion
Posttranscriptional regulation, mRNA stability
Positive regulation of cell proliferation, cell differentiation
Hematopoiesis, leukocyte differentiation, leukocyte activation
Regulation of leukocyte activation
Positive regulation of programmed cell death, regulation of transcription
Regulation of metabolic processes, regulation of transcription and gene expression
FDR q-value
node size
FDR q-value

**Figure 2**

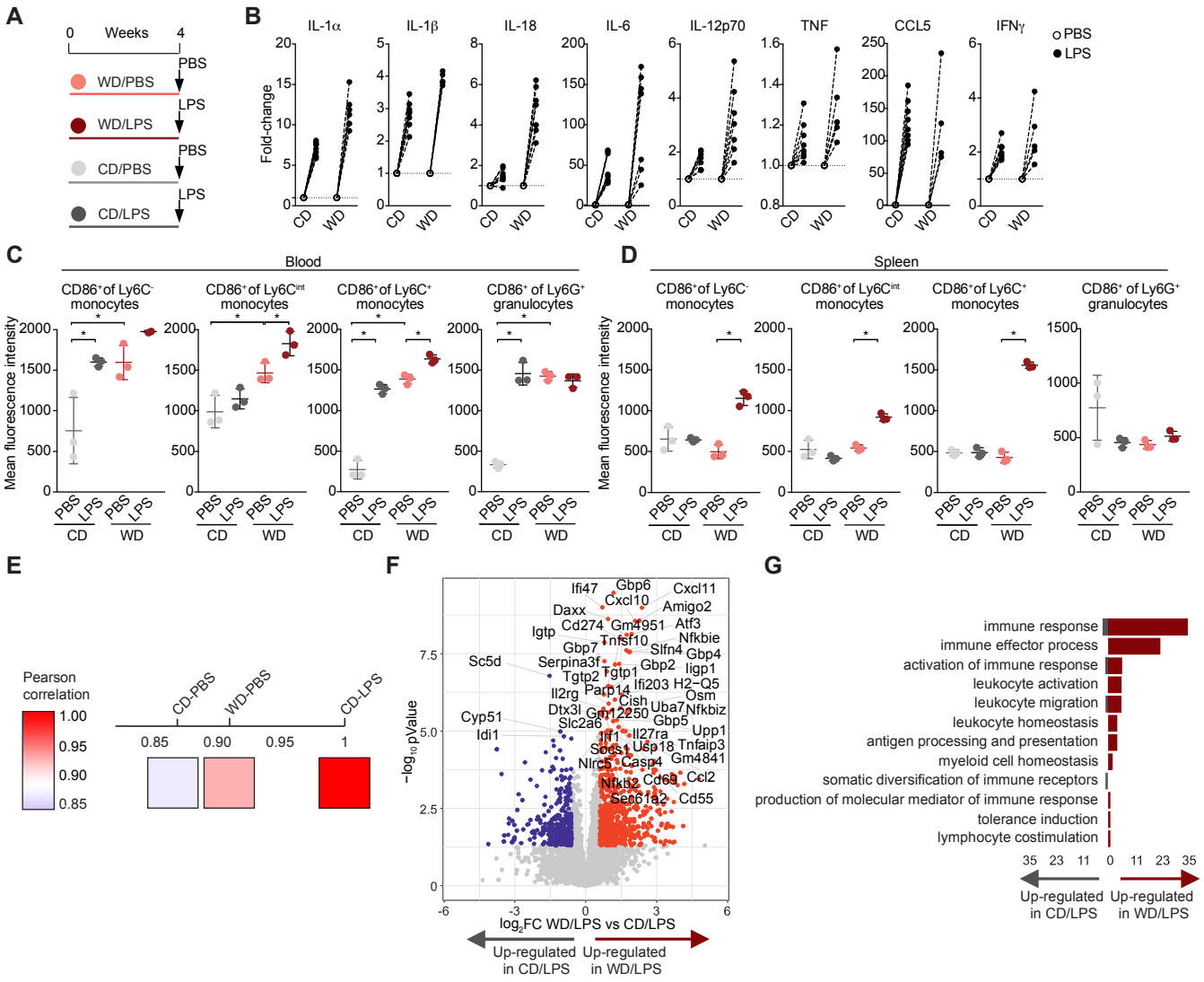

A
0 Weeks 4
WD/PBS
WD/LPS
CD/PBS
CD/LPS
PBS
LPS
B
IL-1α
IL-1β
IL-18
IL-6
IL-12p70
TNF
CCL5
IFNγ
Fold-change
PBS
LPS
CD
WD
C
Blood
CD86+ of Ly6C- monocytes
CD86+ of Ly6Cint monocytes
CD86+ of Ly6C+ monocytes
CD86+ of Ly6G+ granulocytes
Mean fluorescence intensity
D
Spleen
E
Pearson correlation
CD-PBS
WD-PBS
CD-LPS
F
−log10pValue
log2FC WD/LPS vs CD/LPS
Up-regulated in CD/LPS
Up-regulated in WD/LPS
G
immune response
immune effector process
activation of immune response
leukocyte activation
leukocyte migration
leukocyte homeostasis
antigen processing and presentation
myeloid cell homeostasis
somatic diversification of immune receptors
production of molecular mediator of immune response
tolerance induction
lymphocyte costimulation
Up-regulated in CD/LPS
Up-regulated in WD/LPS

**Figure 3**

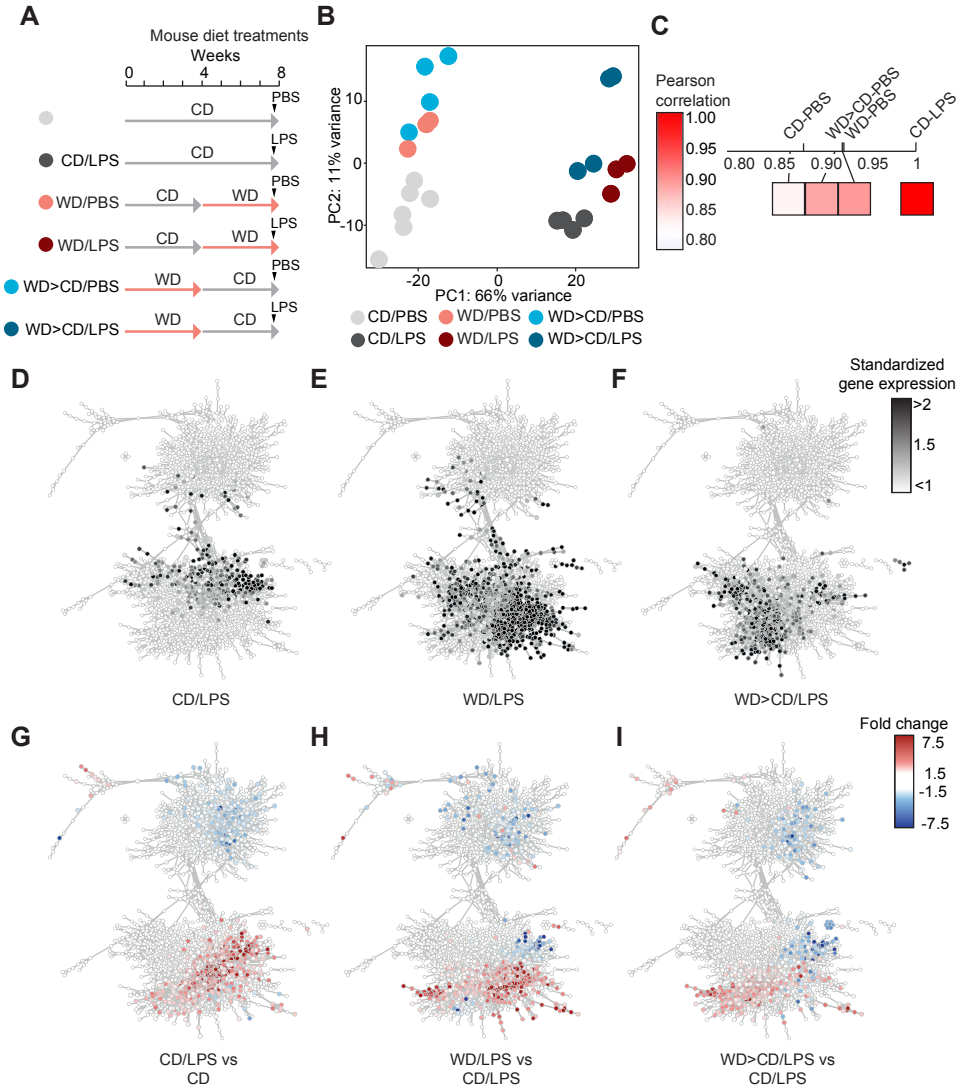

A
Mouse diet treatments
Weeks
0
4
8
CD
PBS
CD/LPS
CD
LPS
WD/PBS
CD
WD
PBS
WD/LPS
CD
WD
LPS
WD>CD/PBS
WD
CD
PBS
WD>CD/LPS
WD
CD
LPS
B
PC2: 11% variance
PC1: 66% variance
CD/PBS
WD/PBS
WD>CD/PBS
CD/LPS
WD/LPS
WD>CD/LPS
C
Pearson correlation
1.00
0.95
0.90
0.85
0.80
CD-PBS
WD>CD-PBS
WD-PBS
CD-LPS
D
E
F
Standardized gene expression
>2
1.5
<1
CD/LPS
WD/LPS
WD>CD/LPS
G
H
I
Fold change
7.5
1.5
-1.5
-7.5
CD/LPS vs CD
WD/LPS vs CD/LPS
WD>CD/LPS vs CD/LPS

**Figure 4**

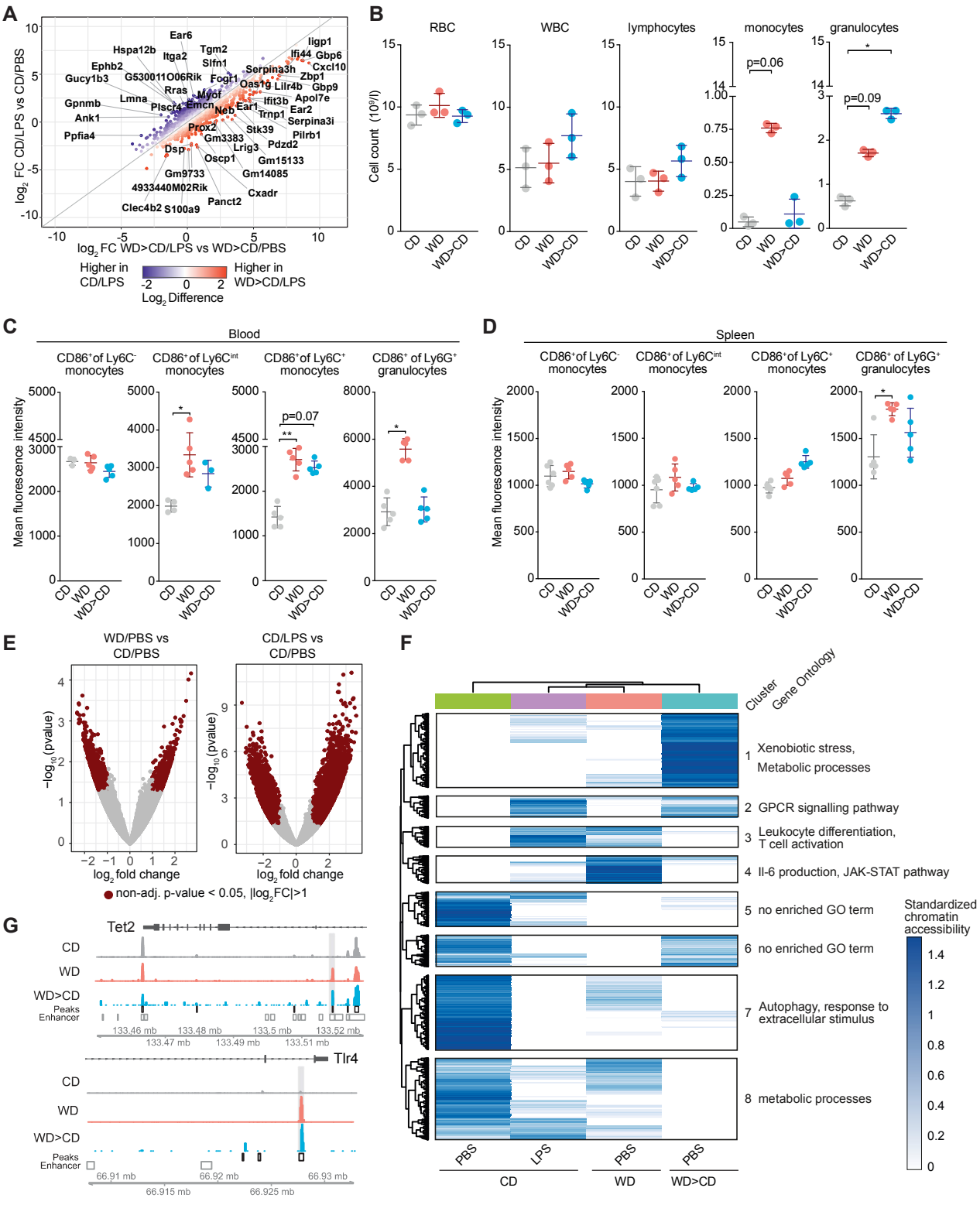

A
B
RBC
WBC
lymphocytes
monocytes
granulocytes
Cell count (10⁹/l)
p=0.06
p=0.09
C
Blood
CD86⁺of Ly6C⁻ monocytes
CD86⁺of Ly6C^int monocytes
CD86⁺of Ly6C⁺ monocytes
CD86⁺ of Ly6G⁺ granulocytes
Mean fluorescence intensity
p=0.07
D
Spleen
E
WD/PBS vs CD/PBS
CD/LPS vs CD/PBS
−log10 (pvalue)
log2 fold change
non-adj. p-value < 0.05, |log2FC|>1
F
Cluster
Gene Ontology
1 Xenobiotic stress, Metabolic processes
2 GPCR signalling pathway
3 Leukocyte differentiation, T cell activation
4 Il-6 production, JAK-STAT pathway
5 no enriched GO term
6 no enriched GO term
7 Autophagy, response to extracellular stimulus
8 metabolic processes
Standardized chromatin accessibility
PBS
LPS
CD
WD
WD>CD
G
Tet2
Tlr4
Peaks
Enhancer

**Figure 5**

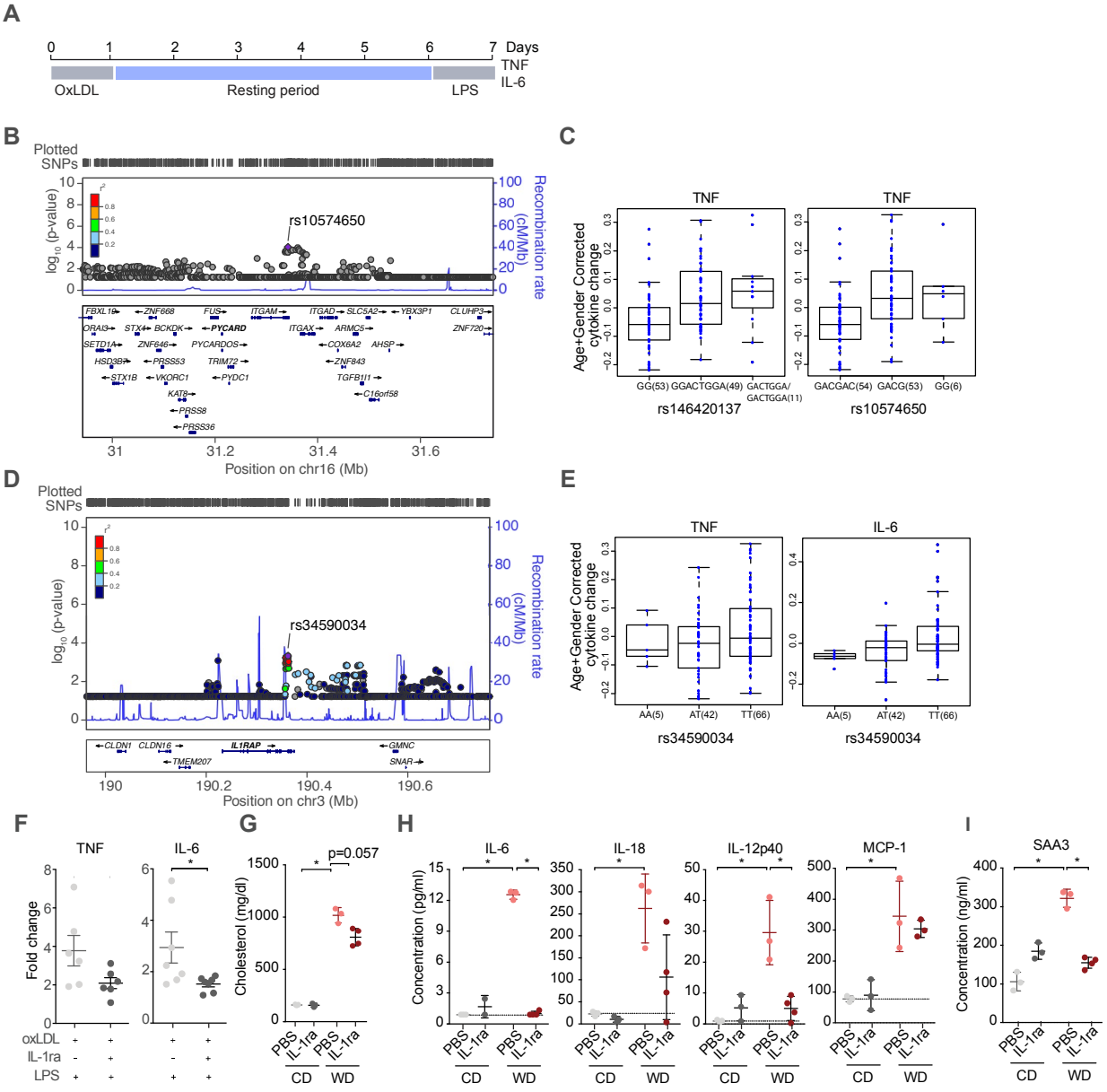

A
0 1 2 3 4 5 6 7 Days
TNF
IL-6
OxLDL
Resting period
LPS
B
Plotted SNPs
rs10574650
log10 (p-value)
Recombination rate (cM/Mb)
Position on chr16 (Mb)
C
TNF
TNF
Age+Gender Corrected cytokine change
GG(53) GGACTGGA(49) GACTGGA/GACTGGA(11)
GACGAC(54) GACG(53) GG(6)
rs146420137
rs10574650
D
Plotted SNPs
rs34590034
log10 (p-value)
Recombination rate (cM/Mb)
Position on chr3 (Mb)
E
TNF
IL-6
Age+Gender Corrected cytokine change
AA(5) AT(42) TT(66)
rs34590034
rs34590034
F
TNF
IL-6
Fold change
oxLDL
IL-1ra
LPS
G
p=0.057
Cholesterol (mg/dl)
PBS IL-1ra PBS IL-1ra
CD WD
H
IL-6
IL-18
IL-12p40
MCP-1
Concentration (pg/ml)
PBS IL-1ra PBS IL-1ra
CD WD
I
SAA3
Concentration (ng/ml)
PBS IL-1ra PBS IL-1ra
CD WD

**Figure 6**

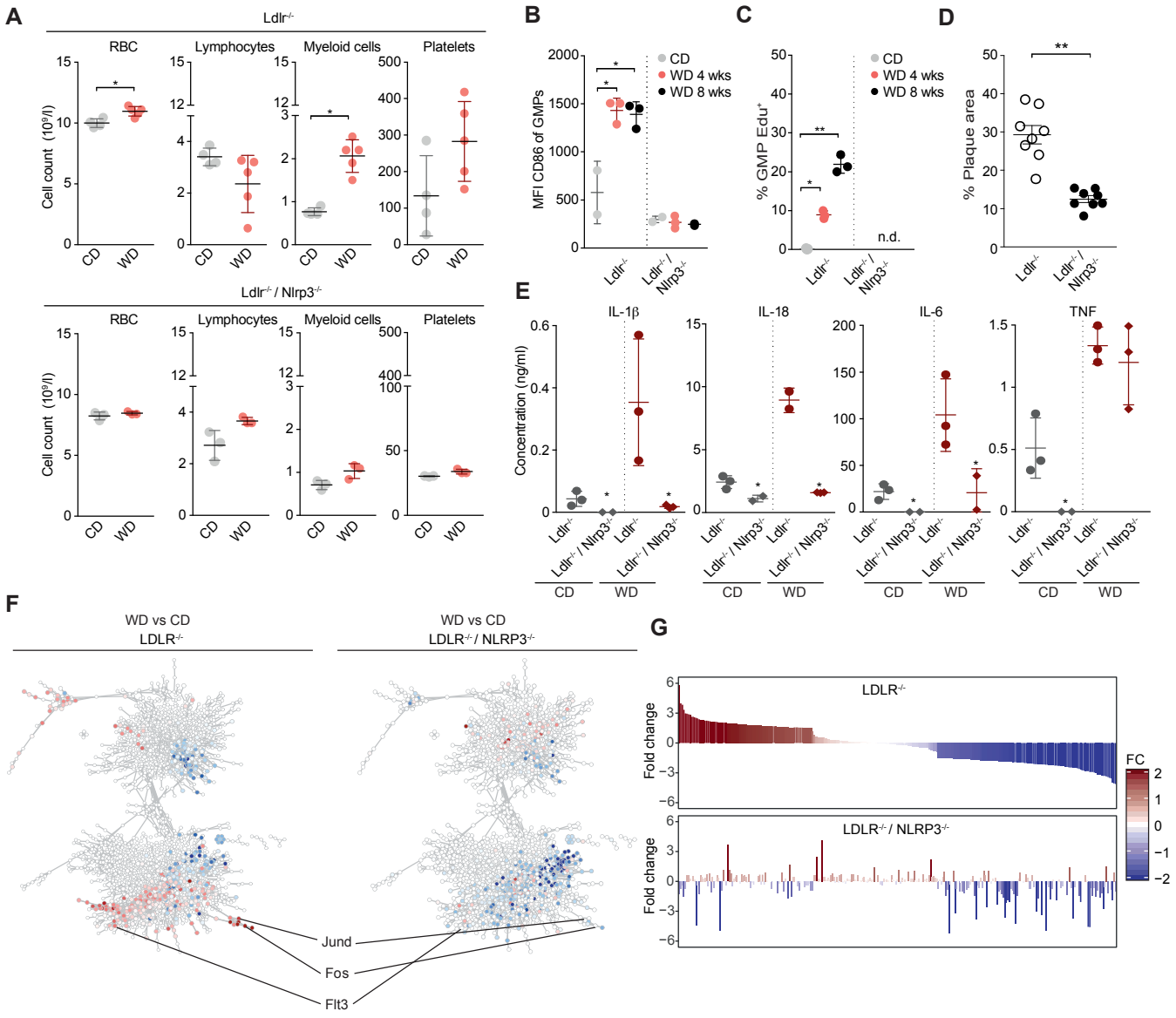

A
Ldlr-/-
RBC
Lymphocytes
Myeloid cells
Platelets
Cell count (10⁹/l)
CD
WD
Ldlr-/- / Nlrp3-/-
B
MFI CD86 of GMPs
CD
WD 4 wks
WD 8 wks
Ldlr-/-
Ldlr-/- / Nlrp3-/-
C
% GMP Edu+
n.d.
D
% Plaque area
E
IL-1β
IL-18
IL-6
TNF
Concentration (ng/ml)
F
WD vs CD
LDLR-/-
WD vs CD
LDLR-/- / NLRP3-/-
Jund
Fos
Flt3
G
LDLR-/-
LDLR-/- / NLRP3-/-
Fold change
FC

Figure 7

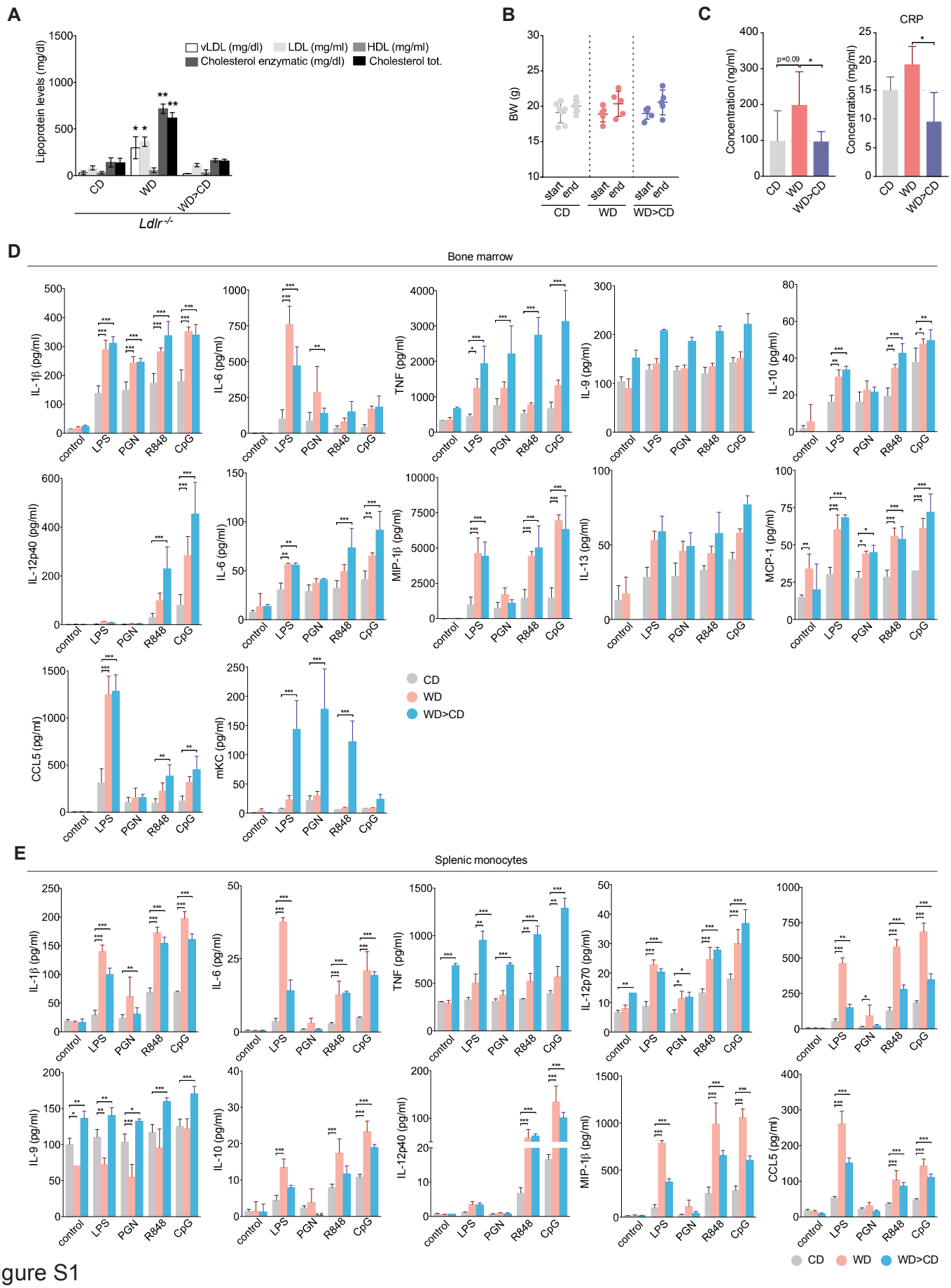

A
Lipoprotein levels (mg/dl)
vLDL (mg/dl)
LDL (mg/ml)
HDL (mg/ml)
Cholesterol enzymatic (mg/dl)
Cholesterol tot.
CD
WD
WD>CD
Ldlr-/-
B
BW (g)
start
end
C
Concentration (ng/ml)
Concentration (mg/ml)
CRP
p=0.09
D
Bone marrow
IL-1β (pg/ml)
IL-6 (pg/ml)
TNF (pg/ml)
IL-9 (pg/ml)
IL-10 (pg/ml)
IL-12p40 (pg/ml)
MIP-1β (pg/ml)
IL-13 (pg/ml)
MCP-1 (pg/ml)
CCL5 (pg/ml)
mKC (pg/ml)
control
LPS
PGN
R848
CpG
E
Splenic monocytes
IL-12p70 (pg/ml)

Figure S1

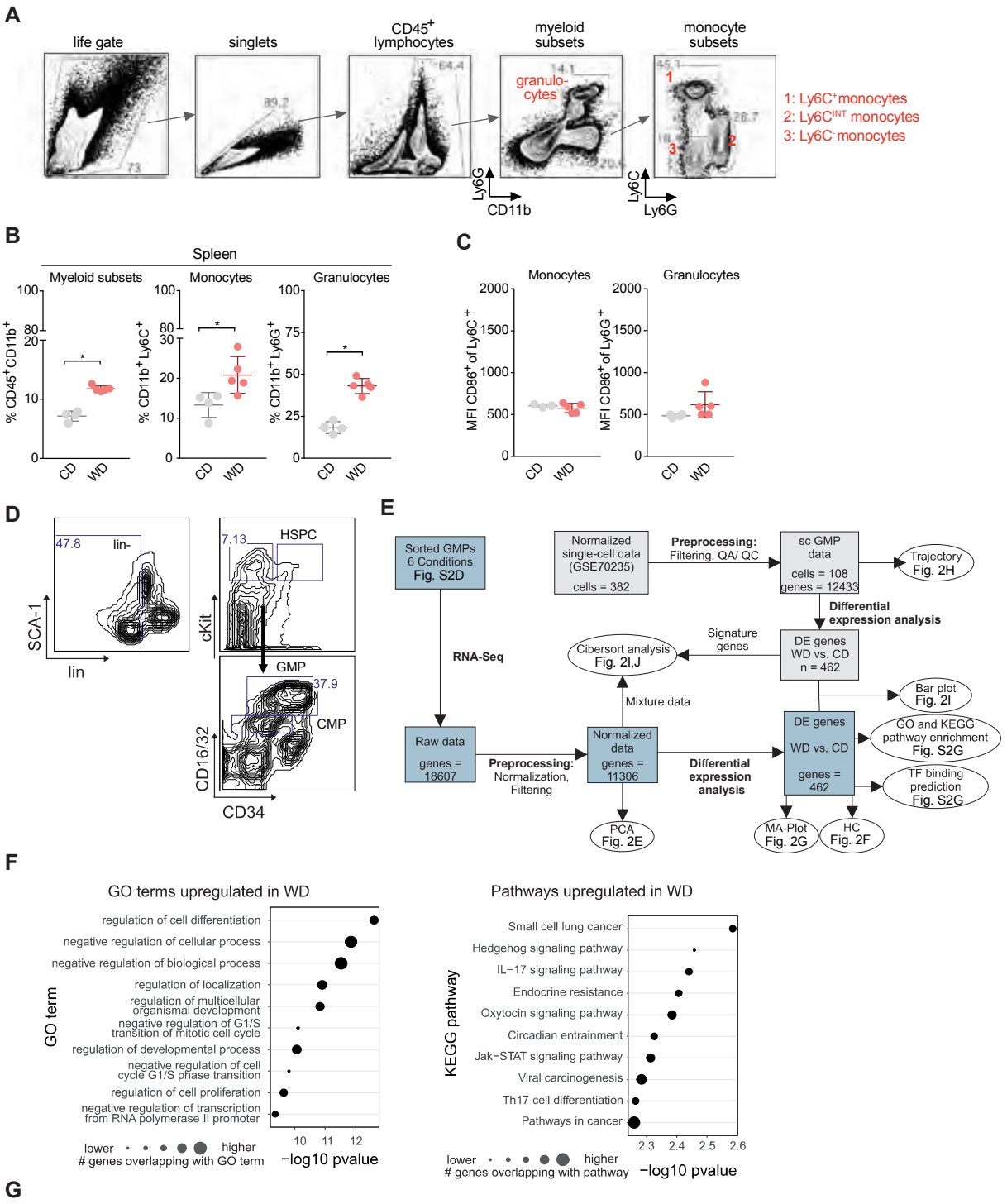

A
life gate
singlets
CD45+ lymphocytes
myeloid subsets
monocyte subsets
granulo-cytes
1: Ly6C+ monocytes
2: Ly6CINT monocytes
3: Ly6C- monocytes
Ly6G
CD11b
Ly6C
Ly6G
B
Spleen
Myeloid subsets
Monocytes
Granulocytes
% CD45+CD11b+
% CD11b+Ly6C+
% CD11b+Ly6G+
CD
WD
C
Monocytes
Granulocytes
MFI CD86+ of Ly6C+
MFI CD86+ of Ly6G+
D
47.8
lin-
SCA-1
lin
7.13
HSPC
cKit
GMP
37.9
CMP
CD16/32
CD34
E
Sorted GMPs 6 Conditions Fig. S2D
RNA-Seq
Raw data genes = 18607
Preprocessing: Normalization, Filtering
Normalized data genes = 11306
PCA Fig. 2E
Differential expression analysis
Normalized single-cell data (GSE70235) cells = 382
Preprocessing: Filtering, QA/ QC
sc GMP data cells = 108 genes = 12433
Trajectory Fig. 2H
Differential expression analysis
DE genes WD vs. CD n = 462
Signature genes
Cibersort analysis Fig. 2I,J
Mixture data
Bar plot Fig. 2I
DE genes WD vs. CD genes = 462
GO and KEGG pathway enrichment Fig. S2G
TF binding prediction Fig. S2G
MA-Plot Fig. 2G
HC Fig. 2F
F
GO terms upregulated in WD
regulation of cell differentiation
negative regulation of cellular process
negative regulation of biological process
regulation of localization
regulation of multicellular organismal development
negative regulation of G1/S transition of mitotic cell cycle
regulation of developmental process
negative regulation of cell cycle G1/S phase transition
regulation of cell proliferation
negative regulation of transcription from RNA polymerase II promoter
GO term
lower higher
# genes overlapping with GO term
−log10 pvalue
Pathways upregulated in WD
Small cell lung cancer
Hedgehog signaling pathway
IL−17 signaling pathway
Endocrine resistance
Oxytocin signaling pathway
Circadian entrainment
Jak−STAT signaling pathway
Viral carcinogenesis
Th17 cell differentiation
Pathways in cancer
KEGG pathway
lower higher
# genes overlapping with pathway
−log10 pvalue
G
WD
CD
TF binding prediction
Ahr Arnt
Mef2c Srf
Tef Hlf Atf2
Nr3c1 Hoxa10
Zfat Znx1
E2f4
Gata3 Gata1 Gata2 Klf4 Tcf7
Fosb Fos Jund Atf3 Junb Creb1 Atf4 Atf2 Crem
Egr1 Egr2
Irf9 Irf3 Irf8 Irf7 Churc1 Irf1 Prdm1 Stat1
Runx1 Runx3
log2 FC WD vs CD
Normalized enrichment score

Figure S2

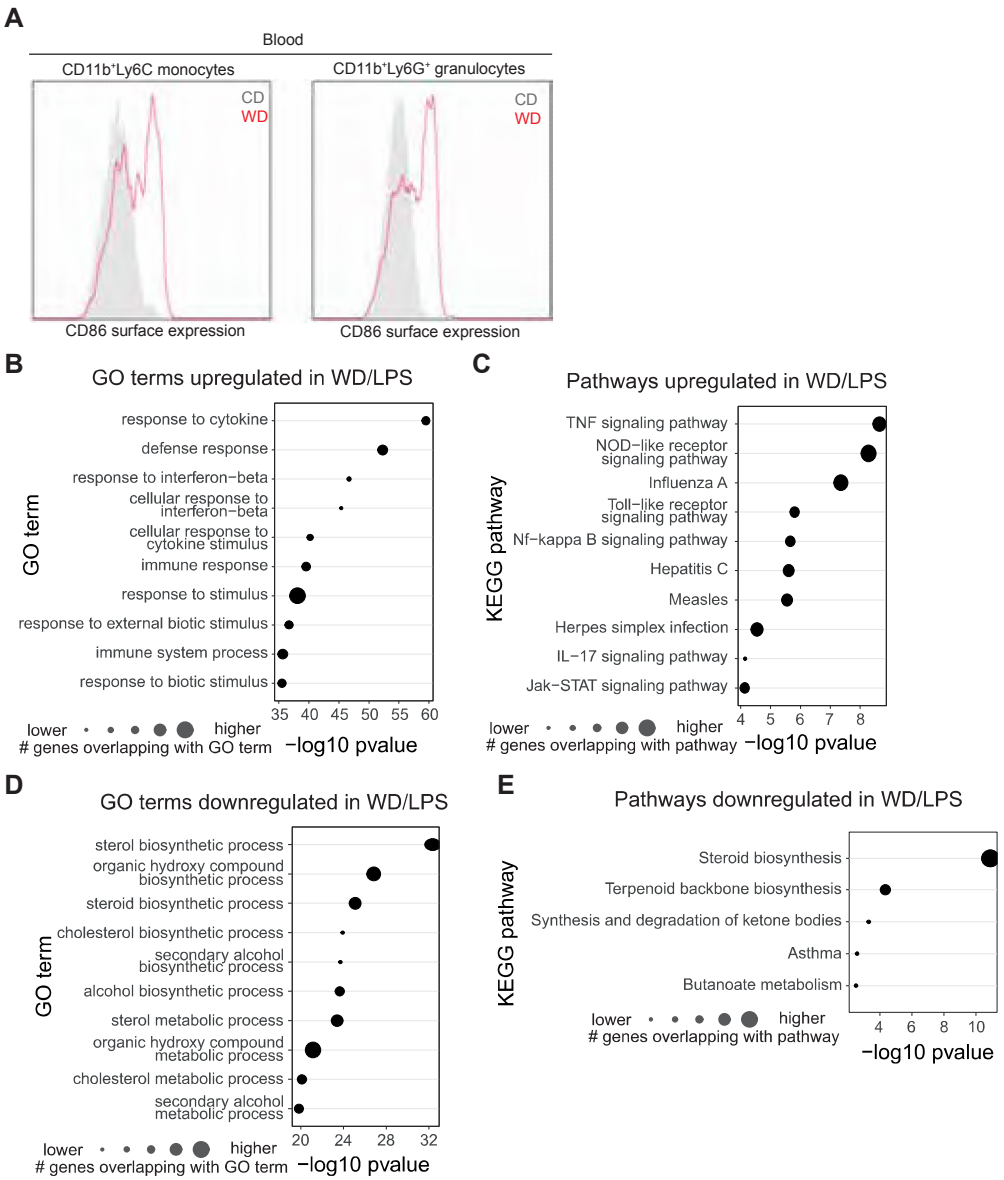

A
Blood
CD11b+Ly6C monocytes
CD11b+Ly6G+ granulocytes
CD
WD
CD86 surface expression
B
GO terms upregulated in WD/LPS
response to cytokine
defense response
response to interferon−beta
cellular response to interferon−beta
cellular response to cytokine stimulus
immune response
response to stimulus
response to external biotic stimulus
immune system process
response to biotic stimulus
GO term
lower higher
# genes overlapping with GO term
−log10 pvalue
C
Pathways upregulated in WD/LPS
TNF signaling pathway
NOD−like receptor signaling pathway
Influenza A
Toll−like receptor signaling pathway
Nf−kappa B signaling pathway
Hepatitis C
Measles
Herpes simplex infection
IL−17 signaling pathway
Jak−STAT signaling pathway
KEGG pathway
lower higher
# genes overlapping with pathway
−log10 pvalue
D
GO terms downregulated in WD/LPS
sterol biosynthetic process
organic hydroxy compound biosynthetic process
steroid biosynthetic process
cholesterol biosynthetic process
secondary alcohol biosynthetic process
alcohol biosynthetic process
sterol metabolic process
organic hydroxy compound metabolic process
cholesterol metabolic process
secondary alcohol metabolic process
GO term
lower higher
# genes overlapping with GO term
−log10 pvalue
E
Pathways downregulated in WD/LPS
Steroid biosynthesis
Terpenoid backbone biosynthesis
Synthesis and degradation of ketone bodies
Asthma
Butanoate metabolism
KEGG pathway
lower higher
# genes overlapping with pathway
−log10 pvalue

Figure S3

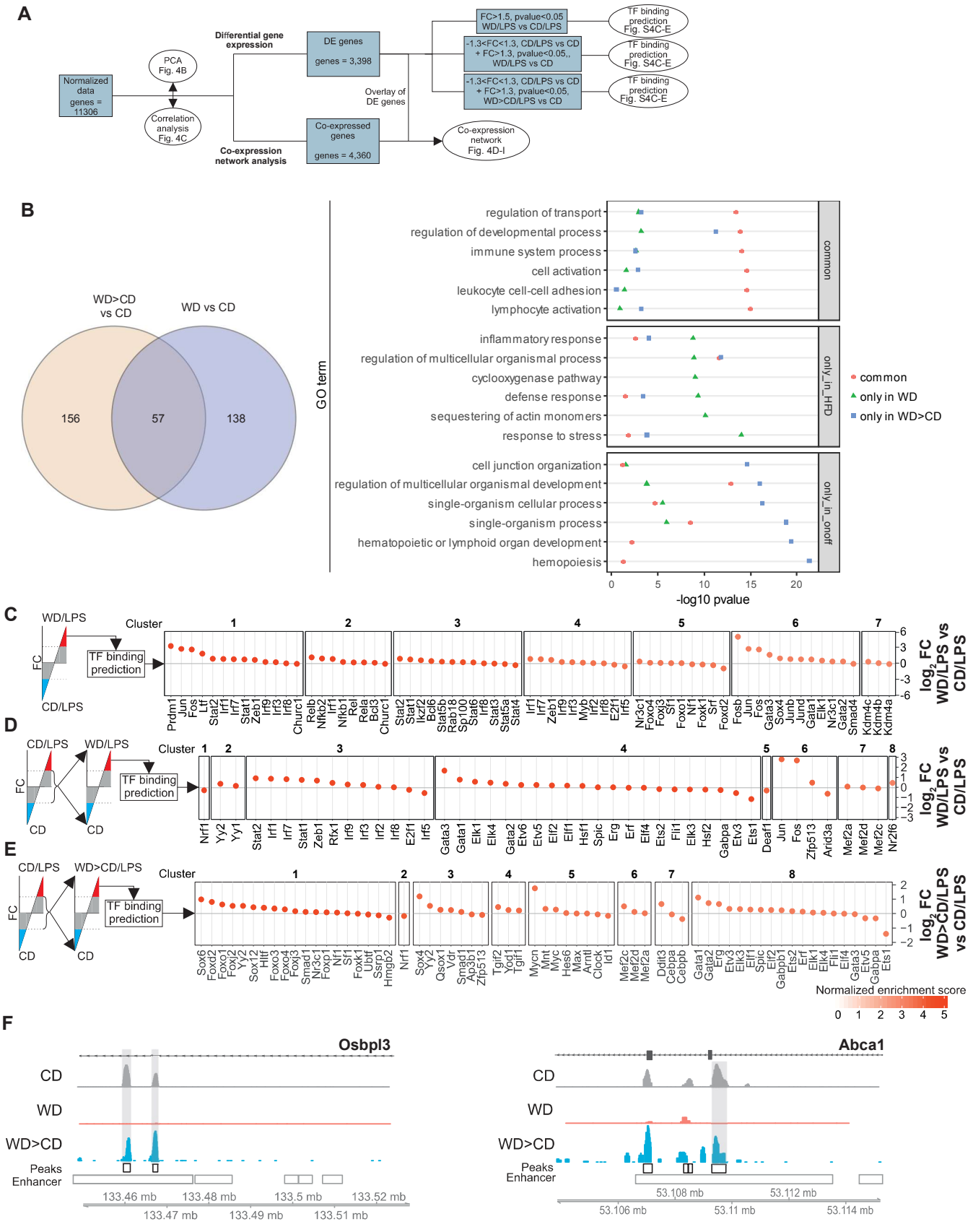

A
Normalized data genes = 11306
PCA Fig. 4B
Correlation analysis Fig. 4C
Differential gene expression
DE genes genes = 3,398
FC>1.5, pvalue<0.05 WD/LPS vs CD/LPS
-1.3<FC<1.3, CD/LPS vs CD + FC>1.3, pvalue<0.05,, WD/LPS vs CD
-1.3<FC<1.3, CD/LPS vs CD + FC>1.3, pvalue<0.05, WD>CD/LPS vs CD
TF binding prediction Fig. S4C-E
TF binding prediction Fig. S4C-E
TF binding prediction Fig. S4C-E
Overlay of DE genes
Co-expressed genes genes = 4,360
Co-expression network analysis
Co-expression network Fig. 4D-I
B
WD>CD vs CD
WD vs CD
156
57
138
GO term
regulation of transport
regulation of developmental process
immune system process
cell activation
leukocyte cell-cell adhesion
lymphocyte activation
common
inflammatory response
regulation of multicellular organismal process
cyclooxygenase pathway
defense response
sequestering of actin monomers
response to stress
only_in_HFD
cell junction organization
regulation of multicellular organismal development
single-organism cellular process
single-organism process
hematopoietic or lymphoid organ development
hemopoiesis
only_in_onoff
common
only in WD
only in WD>CD
0
5
10
15
20
-log10 pvalue
C
WD/LPS
CD/LPS
FC
TF binding prediction
Cluster
1
2
3
4
5
6
7
log2 FC WD/LPS vs CD/LPS
D
CD/LPS
WD/LPS
CD
CD
TF binding prediction
Cluster
1
2
3
4
5
6
7
8
log2 FC WD/LPS vs CD/LPS
E
CD/LPS
WD>CD/LPS
CD
CD
TF binding prediction
Cluster
1
2
3
4
5
6
7
8
log2 FC WD>CD/LPS vs CD/LPS
Normalized enrichment score
0
1
2
3
4
5
F
Osbpl3
CD
WD
WD>CD
Peaks
Enhancer
133.46 mb
133.48 mb
133.5 mb
133.52 mb
133.47 mb
133.49 mb
133.51 mb
Abca1
CD
WD
WD>CD
Peaks
Enhancer
53.108 mb
53.112 mb
53.106 mb
53.11 mb
53.114 mb

Figure S4

## Supplemental Tables

**Table 1. Serum cytokine levels.** Female *Ldlr*$^{-/-}$ mice were fed either chow diet (CD), western diet (WD) for 4 weeks or WD for 4 weeks followed by CD for 4 weeks (WD>CD) and serum cytokines were measured. pg/ml = picogram per milliliter

| Cytokine (pg/ml) | CD | WD | WD>CD |
|---|---|---|---|
| IL-1α | 12.15 | 107.45 | 54.45 |
| IL-1β | 554.00 | 2288.98 | 472.31 |
| IL-2 | 41.22 | 247.23 | 39.90 |
| IL-3 | 13.36 | 102.23 | 13.62 |
| IL-4 | 7.47 | 42.08 | 7.95 |
| IL-5 | 60.22 | 156.47 | 30.42 |
| IL-6 | 22.81 | 62.56 | 16.20 |
| IL-9 | 318.81 | 802.64 | 302.35 |
| IL-10 | 87.03 | 385.39 | 78.13 |
| IL-12p40 | 264.02 | 211.49 | 253.39 |
| IL-12p70 | 157.05 | 627.88 | 185.46 |
| IL-13 | 368.36 | 1530.54 | 477.75 |
| IL-17 | 52.85 | 199.10 | 69.98 |
| G-CSF | 159.48 | 272.95 | 130.05 |
| GM-CSF | 212.68 | 561.09 | 211.18 |
| IFNγ | 19.02 | 93.98 | 18.69 |
| CXCL1 | 163.69 | 140.47 | 87.91 |
| MCP-1 | 436.50 | 1222.79 | 338.51 |
| MIP-1α | 18.58 | 37.06 | 17.17 |
| MIP-1β | 58.84 | 209.35 | 51.37 |
| CCL5 | 76.34 | 85.01 | 89.40 |
| TNF-α | 1965.22 | 3002.10 | 2207.60 |
| IL-18 | 1498.37 | 3184.53 | 1618.29 |
| VEGF | 22.63 | 115.80 | 22.64 |

Publication III

See, P.*, Dutertre, C.A.*, Chen, J.*, **Günther, P.**, McGovern, N., Irac, S.E., Gunawan, M., Beyer, M., Händler, K., Duan, K., Bin Sumatoh, H.R., Ruffin, N., Jouve, M., Gea-Mallorquí, E., Hennekam, R.C.M., Lim, T., Yip, C.C., Wen, M., Malleret, B., Low, I., Shadan, N.B., Fen, C.F.S., Tay, A., Lum, J., Zolezzi, F., Larbi, A., Poidinger, M., Chan, J.K.Y., Chen, Q., Rénia, L., Haniffa, M., Benaroch, P., Schlitzer, A., Schultze, J.L., Newell, E.W. and Ginhoux, F. (2017) 'Mapping the Human DC Lineage through the Integration of High-Dimensional Techniques'. *Science*, 356(6342), p. eaag3009. DOI: 10.1126/science.aag3009.

* Shared first authorship



My contribution to this publication includes:

assistance in library preparation of single-cell MARS-Seq libraries; development and improvement of a bioinformatics pipeline for the preprocessing of MARS-Seq data; quality control and data cleaning of raw data; improving strategy for barcode correction; creating a downstream analysis pipeline including normalization and filtering of data; performing initial analysis of single-cell data; performing advanced analysis of scRNA-seq data; benchmarking different dimensionality reduction and clustering methods; developing a strategy for signature score calculation; applying trajectory analysis of single-cell data; identification of DC-specific signatures from microarray data; interpretation of scRNA-seq results; visualizations of the scRNA-seq data; writing manuscript

**Title:**

# Mapping the human DC lineage through the integration of high dimensional techniques

**One Sentence Summary:**

High dimensional mapping of the human dendritic cell lineage reveals novel pre-dendritic cell populations in human blood.

**Authors:**

Peter See[1#], Charles-Antoine Dutertre[1,2#], Jinmiao Chen[1#], Patrick Günther[3], Naomi McGovern[1], Sergio Erdal Irac[2], Merry Gunawan[4], Marc Beyer[3,5], Kristian Händler[3], Kaibo Duan[1], Hermi Rizal Bin Sumatoh[1], Nicolas Ruffin[6], Mabel Jouve[6], Ester Gea-Mallorquí[6], Raoul C.M. Hennekam[7], Tony Lim[8], Chan Chung Yip[9], Ming Wen[2], Benoit Malleret[1,10], Ivy Low[1], Nurhidaya Binte Shadan[1], Charlene Foong Shu Fen[11], Alicia Tay[1], Josephine Lum[1], Francesca Zolezzi[1], Anis Larbi[1], Michael Poidinger[1], Jerry K.Y. Chan[1,12,13,14], Qingfeng Chen[15], Laurent Renia[1], Muzlifah Haniffa[4], Philippe Benaroch[6], Andreas Schlitzer[1,16], Joachim L. Schultze[3,5], Evan W. Newell[1], and Florent Ginhoux[1]*

**Affiliations:**

[1]Singapore Immunology Network (SIgN), A*STAR, 8A Biomedical Grove, Immunos Building, Level 4, Singapore 138648, Singapore

[2]Program in Emerging Infectious Disease, Duke-NUS Medical School, 8 College Road, 169857 Singapore

[3]Genomics & Immunoregulation, LIMES-Institute, University of Bonn, 32115 Bonn, Germany

[4]Institute of Cellular Medicine, Newcastle University, Newcastle, UK

[5]Platform for Single Cell Genomics and Epigenomics at the German Center for Neurodegenerative Diseases and the University of Bonn, 53175 Bonn, Germany

[6]Institut Curie, PSL Research University, INSERM U 932, F-75005, Paris, France

[7]Department of Pediatrics, Academic Medical Centre, University of Amsterdam, Amsterdam, Netherlands

[8]Department of Anatomical Pathology, Singapore General Hospital, Singapore

[9]Department of HPB and Transplant Surgery, Singapore General Hospital, Singapore

[10]Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[11]Singapore Health Services Flow Cytometry Core Platform, 20 College Road, The Academia, Discovery Tower Level 10, Singapore 169856

[12]Department of Reproductive Medicine, Division of Obstetrics and Gynaecology, KK Women's and Children's Hospital, Singapore

[13]Cancer and Stem Cell Biology Program, Duke-NUS Graduate Medical School, Singapore

[14]Experimental Fetal Medicine Group, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[15]Humanized mouse unit, Institute of Molecular and Cell Biology (IMCB), A*STAR, Singapore

[16]Myeloid Cell Biology, LIMES-Institute, University of Bonn, 53115 Bonn, Germany

# These authors contributed equally to this work.

* Correspondence to Florent Ginhoux

**Abstract:**

Dendritic cells (DC) are professional antigen-presenting cells that orchestrate immune responses. The human DC population comprises two main functionally-specialized lineages, whose origins and differentiation pathways remain incompletely defined. Here we combine two high-dimensional technologies — single-cell mRNA sequencing and Cytometry by Time-of-Flight (CyTOF), to identify human blood $CD123^{+}CD33^{+}CD45RA^{+}$ DC precursors (pre-DC). Pre-DC share surface markers with plasmacytoid DC (pDC) but have distinct functional properties that were previously attributed to pDC. Tracing the differentiation of DC from the bone marrow to the peripheral blood revealed that the pre-DC compartment contains distinct lineage-committed sub-populations including one early uncommitted $CD123^{high}$ pre-DC subset and two $CD45RA^{+}CD123^{low}$ lineage-committed subsets exhibiting functional differences. The discovery of multiple committed pre-DC populations opens promising new avenues for the therapeutic exploitation of DC subset-specific targeting.

**Main Text:**

**Introduction (4500 max including ref, notes and caption)**

Dendritic cells (DC) are professional pathogen-sensing and antigen-presenting cells that are central to the initiation and regulation of immune responses (*1*). The DC population is classified into two lineages: plasmacytoid DC (pDC), and conventional DC (cDC), the latter comprising cDC1 and cDC2 sub-populations (*2*, *3*). Dissecting the origins and differentiation pathways giving rise to DC sub-populations is necessary to understand their homeostasis and role in immune responses, and for the development of DC subset-specific therapeutic interventions. Murine DC arise from unique DC-restricted bone-marrow (BM) progenitors known as common DC progenitors (CDP), which differentiate into pDC and DC precursors (pre-DC) and migrate out of the BM into peripheral tissues (*4-7*). Human equivalents of murine CDP and pre-DC have recently been described (*8*, *9*); human pre-DC comprise ~0.001% of peripheral blood mononuclear cells (PBMC) and were identified by their expression of cytokine receptors that mark and drive DC differentiation in mice, including CD117 (c-kit, SCF), CD116 (GMCSF), CD135 (FLT3) and CD123 (IL3-Rα) (*9*). Previous studies have observed similar receptor expression patterns within human pDC populations, which can differentiate into cDC-like cells when stimulated with IL-3 and CD40L (*10*, *11*). Therefore, either pDC are precursors of cDC, as proposed (*11*), or the conventionally-defined pDC population is heterogeneous, incorporating an independent pre-DC sub-population.

To answer this question, we interrogated the blood $CD135^{+}HLA\text{-}DR^{+}$ fraction, that should contain DC and their precursors (*12-14*), using several integrated, high-dimensional analysis techniques, including single-cell mRNA sequencing (scmRNAseq) and Cytometry by

Time-of-Flight (CyTOF). Using these approaches, we identified a novel population of pre-DC within the conventionally-defined pDC population. These pre-DC exhibit a unique phenotype and distinct functional properties, which were previously attributed to pDC. Extending our analysis to all DC populations in human blood and BM, we identified the entire DC lineage arising from the BM, and revealed the transcriptional priming of pre-DC towards distinct DC subsets. These data offer new insights into DC heterogeneity and ontogeny, and highlight unexplored avenues for investigation of the therapeutic potential of DC subset-specific targeting.

## Results

### *Unbiased identification of DC precursors by unsupervised single-cell RNAseq and CyTOF*

Using PBMC isolated from human blood, we employed massively-parallel single-cell mRNA sequencing (MARS-seq) (*15*) to assess the transcriptional profile of 710 individual cells within the lineage marker (Lin)(CD3/CD14/CD16/CD20/CD34)$^{-}$, HLA-DR$^{+}$CD135$^{+}$ population (Fig. 1, A to G, and fig. S1a (*16*), fig. S1, b to j (*17*), Table S1 (*18*)). The MARS-seq data were processed using non-linear dimensionality reduction via t-stochastic neighbor embedding (tSNE), which enables unbiased visualization of high-dimensional similarities between cells in the form of a two-dimensional map (*19-21*). Density-based spatial clustering of applications with noise (DBSCAN) (*22*) on the tSNE dimensions identified five distinct clusters of transcriptionally-related cells within the selected PBMC population (Fig. 1A, and fig. S1g). To define the nature of these clusters, we calculated gene signature scores for pDC, cDC1 and cDC2 (as described in (*23*), Table S2 (*24*)), and overlaid the expression of the signatures attributed to each cell onto the tSNE visualization. Clusters #1 and #2 (containing 308 and 72 cells, respectively) were identified as pDC, cluster #3 (containing 160 cells) as cDC1, and cluster #5 (containing 120 cells) as cDC2. Cluster #4 (containing 50 cells) lay in between the cDC1 (#3)

and cDC2 (#5) clusters and possessed a weak, mixed pDC/cDC signature (Fig. 1A). We then performed a connectivity MAP (cMAP) analysis (*25*) to calculate the degree of enrichment of pDC or cDC signature gene transcripts in each individual cell. This approach confirmed the signatures of pDC (#1 and #2) and cDC (#3 and #5) clusters, and showed that most cells in cluster #4 expressed a cDC signature (Fig. 1B). The Mpath algorithm (*26*) was then applied to the five clusters to identify hypothetical developmental relationships based on these transcriptional similarities between cells (Fig. 1C, and fig. S2, a and b). Mpath revealed that the five clusters were grouped into three distinct branches with one central cluster (cluster #4) at the intersection of the three branches (Fig. 1C, and fig. S2a). The Mpath edges connecting cluster #4 to cDC1 cluster #3 and cDC2 cluster #5 have a high cell count (159 and 137 cells, respectively), suggesting that the transition from cluster #4 to clusters #3 and #5 is likely valid, and indicates that cluster #4 could contain putative cDC precursors (Fig. 1C). In contrast, the edge connecting cluster #4 and pDC cluster #2 has a cell count of only 7 (Fig. 1C, and fig. S2b), which suggests that this connection is very weak. The edge connecting cluster #4 and #2 was retained when Mpath trimmed the weighted neighborhood network (fig. S2b), simply due to the feature of the Mpath algorithm that requires all clusters to be connected (*26*). To confirm these findings, we tested Monocle (*27*), principal component analyses (PCA), Wishbone (*28*) and Diffusion Map algorithms (*29*). Monocle and PCA resolved the cells into the same three branches as the original Mpath analysis, with the cells from the tSNE cluster #4 again falling at the intersection (Fig. 1, D and E). Diffusion Map and Wishbone analyses indicated that there was a continuum between clusters #3 (cDC1), #4 and #5 (cDC2): cells from cluster #4 were predominantly found in the DiffMap_dim2$^{low}$ region, and cells from clusters #3 and #5 were progressively drifting away from the DiffMap_dim2$^{low}$ region towards the left and right, respectively. The pDC clusters (#1 and #2) were clearly separated from all other clusters (Fig. 1F, and fig. S2c). In support of this observation, cells from these pDC clusters had a higher expression of pDC-specific markers and

transcription factors (TF) than the cDC clusters (#3 and #5) and central cluster #4. Conversely, cells in cluster #4 expressed higher levels of markers and TF associated with all cDC lineage than the pDC clusters (Fig. 1G). This phenotype led us to hypothesize that cluster #4 represented a population of putative uncommitted cDC precursors.

We next employed CyTOF, which simultaneously measures the intensity of expression of up to 38 different molecules at the single cell level, to further understand the composition of the delineated sub-populations. We designed a panel of 38 labeled antibodies to recognize DC lineage and/or progenitor-associated surface molecules (Table S3, Fig. 1, H to J, and fig. S3), and the molecules identified in cluster #4 by MARS-seq, such as CD2, CX3CR1, CD11c and HLA-DR (Fig. 1I). Using the tSNE algorithm, the $CD45^{+}Lin(CD7/CD14/CD15/CD16/CD19/CD34)^{-}HLA\text{-}DR^{+}$ PBMC fraction (fig. S3a) resolved into three distinct clusters representing cDC1, cDC2 and pDC (Fig. 1H). An intermediate cluster at the intersection of the cDC and pDC clusters that expressed both cDC-associated markers (CD11c/CX3CR1/CD2/CD33/CD141/BTLA) and pDC-associated markers (CD45RA/CD123/CD303) (Fig. 1, I to J, and fig. S3b) corresponded to the MARS-seq cluster #4. The delineation of these clusters was confirmed when applying the phenograph unsupervised clustering algorithm (*30*) (fig. S3c). The position of the intermediate $CD123^{+}CD33^{+}$ cell cluster was distinct, and the cells exhibited high expression of CD5, CD327, CD85j, together with high levels of HLA-DR and the cDC-associated molecule CD86 (Fig. 1, I to J). Taken together, these characteristics led us to consider whether $CD123^{+}CD33^{+}$ cells might represent circulating human pre-DC.

***Pre-DC exist within the pDC fraction and give rise to cDC***

We analyzed the $CD123^{+}CD33^{+}$ cell cluster within the $Lin^{-}HLA\text{-}DR^{+}$ fraction of the PBMC by flow cytometry. Here, we identified $CD123^{+}CD33^{-}$ pDC, $CD45RA^{+/-}CD123^{-}$ cDC1 and cDC2, and $CD33^{+}CD45RA^{+}CD123^{+}$ putative pre-DC (Fig. 2A, and fig. S4a). The putative pre-DC expressed CX3CR1, CD2, CD303 and CD304, with low CD11c expression, whereas $CD123^{+}CD33^{-}$ pDC exhibited variable CD2 expression (Fig. 2, A and B, and fig. S4, b and c). We then extended our analysis to immune cells from the spleen and identified a similar putative pre-DC population, which was more abundant than in blood and expressed higher levels of CD11c (Fig. 2, A and C, and fig. S4d). Both putative pre-DC populations in the blood and spleen expressed CD135 and intermediate levels of CD141 (fig. S4c). Wright-Giemsa staining of putative pre-DC sorted from the blood revealed an indented nuclear pattern reminiscent of classical cDC, a region of perinuclear clearing, and a basophilic cytoplasm reminiscent of pDC (Fig. 2D). At the ultra-structural level, putative pre-DC and pDC exhibited distinct features, despite their morphological similarities (Fig. 2E, and fig. S4e): putative pre-DC possessed a thinner cytoplasm, homogeneously-distributed mitochondria (m), less rough endoplasmic reticulum (RER), an indented nuclear pattern, a large nucleus and limited cytosol, compared to pDC; pDC contained a smaller nucleus, abundant cytosol, packed mitochondria, well-developed and polarized cortical RER organized in parallel cisterna alongside numerous stacks of rough ER membranes, suggesting a developed secretory apparatus, in agreement with previously-published data (*31*).

We then compared the differentiation capacity of pre-DC to that of cDC and pDC, through stromal culture in the presence of FLT3L, GMCSF and SCF, as previously described (*8*). After 5 days, the pDC, cDC1 and cDC2 populations remained predominantly in their initial states, whereas the putative pre-DC population had differentiated into cDC1 and cDC2 in the known proportions found *in vivo* (*14*, *23*, *32*, *33*) (Fig. 2F, fig. S4f, and fig. S5). Altogether, these data suggest that $CD123^{+}CD33^{+}CD45RA^{+}CX3CR1^{+}CD2^{+}$ cells are circulating pre-DC

with cDC differentiation potential. Breton and colleagues (*9*) recently reported a minor population of human pre-DC (highlighted in fig. S6a), which shares a similar phenotype with the $Lin^{-}CD123^{+}CD33^{+}CD45RA^{+}$ pre-DC defined here (fig. S6, a and b). Our results reveal that the pre-DC population in blood and spleen is markedly larger than the one identified within the minor $CD303^{-}CD141^{-}CD117^{+}$ fraction considered previously (fig. S6, c and d).

***Pre-DC are functionally distinct from pDC***

IFNα-secreting pDC can differentiate into cells resembling cDC when cultured with IL-3 and CD40L (*10*, *11*), and have been considered DC precursors (*11*). However, when we used traditional $ILT3^{+}ILT1^{-}$ (*10*) or $CD4^{+}CD11c^{-}$ (*11*) pDC gating strategies, we detected a "contaminating" $CD123^{+}CD33^{+}CD45RA^{+}$ pre-DC sub-population in both groups (fig. S6, e and f). We questioned, therefore, whether other properties of traditionally-classified "pDC populations" might be attributed to pre-DC. TLR7/8 (CL097) or TLR9 (CpG ODN 2216) stimulation of pure pDC cultures resulted in abundant secretion of IFNα, but not IL-12p40, whereas pre-DC readily secreted IL-12p40 but not IFNα (Fig. 2G, and fig. S7). Furthermore, while pDC were previously thought to induce proliferation of naïve $CD4^{+}$ T cells (*10*, *34*), here we found that only the pre-DC sub-population exhibited this attribute (Fig. 2H). Reports of potent allostimulatory capacity and IL-12p40 production by $CD2^{+}$ pDC (*34*) might then be explained by $CD2^{+}$ pre-DC "contamination" (*35*) (fig. S8).

Pitt-Hopkins Syndrome (PHS) is characterized by abnormal craniofacial and neural development, severe mental retardation, and motor dysfunction, and is caused by haplo-insufficiency of *TCF4*, which encodes the E2-2 transcription factor — a central regulator of pDC development (*36*). We confirmed that patients with PHS had a marked reduction in their blood pDC numbers compared to healthy individuals, but that they retained a population of pre-DC

(Fig. 2I, and fig. S9), which likely accounts for the unexpected $CD45RA^{+}CD123^{+}CD303^{lo}$ cell population reported in these patients (*37*). Taken together, our data indicate that, while pre-DC and pDC share some phenotypic features, they can be separated by their differential expression of several markers, including CD33, CX3CR1, CD2, CD5 and CD327. pDC are *bona fide* IFNα-producing cells, but the reported IL-12 production and $CD4^{+}$ T-cell allostimulatory capacity of pDC can likely be attributed to "contaminating" pre-DC, which can give rise to both cDC1 and cDC2.

***Identification and characterization of committed pre-DC subsets***

The murine pre-DC population contains both uncommitted and committed pre-cDC1 and pre-cDC2 precursors (*7*). We asked whether the same was true for human blood pre-DC using microfluidic scmRNAseq (fig. S10a (*38*), fig. S10, b and c (*39*), Table S4 (*40*)). The additional single cell gene expression data relative to the MARS-seq strategy used for Fig. 1, A to G (2.5 million reads/cell and an average of 4,742 genes detected per cell *vs* 60,000 reads/cell and an average of 749 genes detected per cell, respectively) was subjected to cMAP analysis, which calculated the degree of enrichment for cDC1 or cDC2 signature gene transcripts (*23*) for each single cell (Fig. 3A). Among the 92 analyzed pre-DC, 25 cells exhibited enrichment for cDC1 gene expression signatures, 12 cells for cDC2 gene expression signatures, and 55 cells showed no transcriptional similarity to either cDC subset. Further Mpath analysis showed that these 55 "unprimed" pre-DC were developmentally related to cDC1-primed and cDC2-primed pre-DC, and thus their patterns of gene expression fell between the cDC1 and cDC2 signature scores by cMAP (Fig. 3B, and fig. S11). These data suggest that the human pre-DC population contains cells exhibiting transcriptomic priming towards cDC1 and cDC2 lineages, as observed in mice (*7*).

We next asked whether we could identify this heterogeneity within the pre-DC population by flow cytometry, using either pre-DC-specific markers (CD45RA, CD327, CD5) or markers expressed more intensely by pre-DC compared to cDC2 (BTLA, CD141). 3D-PCA analysis of the $Lin^{-}HLA\text{-}DR^{+}CD33^{+}$ population (containing both differentiated cDC and pre-DC) identified three major cell clusters: $CADM1^{+}$ cDC1, $CD1c^{+}$ cDC2 and $CD123^{+}$ pre-DC (Fig. 3C, and fig. S12a). Interestingly, while cells located at the intersection of these three clusters (Fig. 3D) expressed lower levels of CD123 than pre-DC, but higher levels than differentiated cDC (Fig. 3C), they also expressed high levels of pre-DC markers (Fig. 3D, and fig. S12a). We reasoned that these $CD45RA^{+}CD123^{lo}$ cells might be committed pre-DC that are differentiating into either cDC1 or cDC2 (Fig. 3E). The Wanderlust algorithm (*41*), which orders cells into a constructed trajectory according to their maturity, confirmed the developmental relationship between pre-DC (early events), $CD45RA^{+}CD123^{lo}$ cells (intermediate events) and mature cDC (late events) (Fig. 3F). Flow cytometry of PBMC identified $CD123^{+}CADM1^{-}CD1c^{-}$ putative uncommitted pre-DC, alongside putative $CADM1^{+}CD1c^{-}$ pre-cDC1 and $CADM1^{-}CD1c^{+}$ pre-cDC2 within the remaining $CD45RA^{+}$ cells (Fig. 3G, and fig. S12b). These three populations were also present, and more abundant, in the spleen (fig. S12c). Importantly, *in vitro* culture of pre-DC subsets sorted from PBMC did not give rise to any $CD303^{+}$ cells (which would be either undifferentiated pre-DC or differentiated pDC), whereas early pre-DC gave rise to both cDC subsets, and pre-cDC1 and pre-cDC2 differentiated exclusively into cDC1 and cDC2 subsets, respectively (Fig. 3H, fig. S12d, and fig. S13).

Scanning electron microscopy confirmed that early pre-DC are larger and rougher in appearance than pDC, and that committed pre-DC subsets closely resemble their mature cDC counterparts (Fig. 3I, and fig. S14a). Phenotyping of blood pre-DC by flow cytometry (fig. S14b) identified patterns of transitional marker expression throughout the development of early pre-DC towards pre-cDC1/2 and differentiated cDC1/2. Specifically, CD45RO and CD33 were acquired

in parallel with the loss of CD45RA; CD5, CD123, CD304 and CD327 were expressed abundantly by early pre-DC, intermediately by pre-cDC1 and pre-cDC2, and rarely if at all by mature cDC and pDC; FcεRI and CD1c were acquired as early pre-DC commit towards the cDC2 lineage, concurrent with the loss of BTLA and CD319 expression; early pre-DC had an intermediate expression of CD141 that dropped along cDC2 differentiation but was increasingly expressed during commitment towards cDC1, with a few pre-cDC1 already starting to express Clec9A; and IRF8 and IRF4 - transcription factors regulating cDC lineage development (*2*, *3*) - were expressed by early pre-DC and pre-cDC1, while pre-cDC2 maintained only IRF4 expression (fig. S14c).

We next sorted pre-DC and DC subsets from blood and performed microarray analyses to define their entire transcriptome. 3D-PCA analysis of the microarray data showed that pDC were clearly separated from other pre-DC and DC subsets along the horizontal PC1 axis (Fig. 4A, and fig. S15). The combination of the PC2 and PC3 axes indicated that pre-cDC1 occupied a position between early pre-DC and cDC1 and, although cDC2 and pre-cDC2 exhibited similar transcriptomes, pre-cDC2 were positioned between cDC2 and early pre-DC along the PC3 axis (Fig. 4A). Hierarchical clustering of differentially-expressed genes (DEG) confirmed the similarities between committed pre-DC and their corresponding mature subset (fig. S16). The greatest number of DEG was between early pre-DC and pDC (1249 genes) among which CD86, CD2, CD22, CD5, ITGAX (CD11c), CD33, CLEC10A, SIGLEC6 (CD327), THBD, CLEC12A, KLF4 and ZBTB46 were more highly expressed by early pre-DC, while pDC showed higher expression of CD68, CLEC4C, TCF4, PACSIN1, IRF7 and TLR7 (Fig. 4B). An evolution in the gene expression pattern was evident from early pre-DC, to pre-cDC1 and then cDC1 (Fig. 4C), whereas pre-cDC2 were similar to cDC2 (Fig. 4D, and fig. S16). The union of DEGs comparing pre-cDC1versus early pre-DC and cDC1 versus pre-cDC1 has 62 genes in common with the union of DEGs from comparing pre-cDC2 versus early pre-DC and cDC2 versus pre-cDC2.

These 62 common genes include the transcription factors BATF3, ID2 and TCF4 (E2-2), and the pre-DC markers CLEC4C (CD303), SIGLEC6 (CD327), and IL3RA (CD123) (Fig. 4E, fig. S17 and Table S5). The progressive reduction in transcript abundance of SIGLEC6 (CD327), CD22 and AXL during early pre-DC to cDC differentiation was also mirrored at the protein level (Fig. 4F). Key transcription factors involved in the differentiation and/or maturation of DC subsets showed a progressive change in their expression along the differentiation path from pre-DC to mature cDC (Fig. 4G). Finally, pathway analyses revealed that pre-DC exhibited an enrichment of cDC functions relative to pDC, and were maintained in a relatively immature state compared to mature cDC (fig. S18).

***Committed pre-DC subsets are functional***

We then asked to what extent the functional specializations of DC (*1*, *42*) were acquired at the precursor level by stimulating PBMC with TLR agonists and measuring their cytokine production (Fig. 5A). Pre-DC produced significantly more TNFα and IL-12p40 when exposed to CpG ODN 2216 (TLR9 agonist), than to either LPS (TLR4 agonist) or polyI:C (TLR3 agonist) ($p=0.03$, Mann-Witney test). We confirmed that pDC were uniquely capable of robust IFNα production in response to CL097 and CpG ODN 2216. CpG ODN 2216 stimulation also triggered IL-12p40 and TNFα production by early pre-DC, pre-cDC1, and to a lesser extent pre-cDC2. Although TLR9 transcripts were detected only in early pre-DC (fig. S19a), these data indicate that, contrary to differentiated cDC1 and cDC2, pre-cDC1 and pre-cDC2 do express functional TLR9. Interestingly, while pre-cDC2 resembled cDC2 at the gene expression level, their responsiveness to TLR ligands was intermediate between that of early pre-DC and cDC2. Pre-DC subsets also expressed T-cell co-stimulatory molecules (Fig. 5B) and induced proliferation and polarization of naïve $CD4^+$ T cells to a similar level as did mature cDC (Fig. 5C, and fig. S19b).

***Unsupervised mapping of DC ontogeny***

To understand the relatedness of the cell subsets, we performed an unsupervised isoMAP analysis (*43*) of human BM cells, obtained from CyTOF analysis, for non-linear dimensionality reduction (Fig. 6A, and fig. S20a). This analysis focused on the $Lin^{-}CD123^{hi}$ fraction and identified $CD123^{hi}CD34^{+}$ CDP (phenograph cluster #5), from which branched $CD34^{-}CD123^{+}CD327^{+}CD33^{+}$ pre-DC (clusters #1 and #2) and $CD34^{-}CD123^{+}CD303^{+}CD68^{+}$ pDC (clusters #3 and #4) which both progressively acquired their respective phenotypes. Cells in the pre-DC branch increasingly expressed CD2, CD11c, CD116 and, at a later stage, CD1c. IsoMAP analysis of $Lin^{-}CD123^{+}$ cells in the peripheral blood identified two parallel lineages, corresponding to pre-DC and pDC, in which a CDP population was not detected (Fig. 6B). IsoMAP and phenograph analysis of pre-DC extracted from the isoMAP analysis of Fig. 6A (BM, clusters #1 and #2) and Fig. 6B (blood, cluster #6) revealed the three distinct pre-DC subsets (Fig. 6C) as defined by their unique marker expression patterns (fig. S20, b and c).

In summary, we traced the developmental stages of DC from the BM to the peripheral blood through CyTOF, showing that the CDP population in the BM bifurcates into two pathways, developing into either pre-DC or pDC in the blood (Fig. 6, A to C). This pre-DC population is heterogeneous and exists as distinct subsets detectable in both the blood and BM (Fig. 6C, and fig. S20, b and c). Furthermore, we uncovered an intriguing heterogeneity in blood and BM pDC that warrants further investigation (Fig. 6C, and fig. S20, d and e).

**Discussion**

Using unsupervised scmRNAseq and CyTOF analyses, we have unraveled the complexity of the human DC lineage at the single cell level, revealing a continuous process of differentiation that starts in the BM with CDP, and diverges at the point of emergence of pre-DC

and pDC potentials, culminating in maturation of both lineages in the blood. A previous study using traditional surface marker-based approaches had suggested the presence of a minor pre-DC population in PBMC (*9*), but the combination of high-dimensional techniques and unbiased analyses employed here shows that this minor population had been markedly underestimated: we reveal a population of pre-DC that overlaps with that observed by Breton and colleagues (*9*) within the $CD117^{+}CD303^{-}CD141^{-}$ fraction of PBMC, but accounts for >10 fold the number of cells in peripheral blood than was originally estimated, and is considerably more diverse fig. S6c).

Recent work in mice found uncommitted and subset-committed pre-DC subsets in the BM (*7*, *44*). Here, we similarly identified three functionally- and phenotypically- distinct pre-DC populations in human PBMC, spleen and BM: uncommitted pre-DC and two populations of subset-committed pre-DC (fig. S21 and fig. S22). In line with the concept of continuous differentiation from the BM to the periphery, the proportion of uncommitted cells was higher in the pre-DC population in the BM than in the blood. Altogether, these findings support a two-step model of DC development whereby a central transcriptomic subset-specific program is imprinted on DC precursors from the CDP stage onwards, conferring a core subset identity irrespective of the final tissue destination; in the second step of the process, peripheral tissue-dependent programming occurs to ensure site-specific functionality and adaptation (*7*, *44*). Future studies will be required to reveal the molecular events underlying DC subset lineage priming and the tissue-specific cues that regulate their peripheral programming, and to design strategies that specifically target DC subsets at the precursor level. In addition, how the proportions of uncommitted pre-DC versus committed pre-DC are modified in acute and chronic inflammatory settings warrants further investigation.

An important aspect of unbiased analyses is that cells are not excluded from consideration on the basis of preconceptions concerning their surface phenotype. We found that

pre-DC express most of the markers that classically defined pDC, such as CD123, CD303 and CD304. Thus, any strategy relying on these markers to identify and isolate pDC will have inadvertently included $CD123^+CD33^+$ pre-DC as well. While this calls us to urgently reconsider some aspects of pDC population biology, it may also explain earlier findings including that: pDC cultures possess cDC potential and acquire cDC-like morphology (*10*, *11*), as recently observed in murine BM pDC (*45*); pDC mediate Th1 immunity through production of IFNα and IL-12 (*10*, *46-50*); pDC exhibit naïve T-cell allostimulatory capacity (*34*, *48*); and pDC express co-stimulatory molecules and exhibit antigen-presentation/cross-presentation capabilities at the expense of IFNα secretion (*46*, *51*). These observations could be attributed to the undetected pre-DC in the pDC populations described by these studies, and indeed it has been speculated that the IL-12 production observed in these early studies might be due to the presence of contaminating $CD11c^+$ cDC (*52*). We directly addressed this possibility by separating $CX3CR1^+CD33^+CD123^+CD303^+CD304^+$ pre-DC from $CX3CR1^-CD33^-CD123^+CD303^+CD304^+$ "pure" pDC and showing that pDC could not polarize or induce proliferation of naïve CD4 T cells, whereas pre-DC had this capacity; and that pDC were unable to produce IL-12, unlike pre-DC, but were the only cells capable of strongly producing IFNα in response to TRL7/8/9 agonists, as initially described (*53*). Thus, it is of paramount importance that pre-DC be excluded from pDC populations in future studies, particularly when using commercial pDC isolation kits. Finally, if pDC are stripped of all their cDC properties, it raises the question as to whether they truly belong to the DC lineage, or rather are a distinct type of innate IFN-I-producing lymphoid cell. It also remains to be shown whether the BM $CD34^+CD123^{hi}$ CDP population is also a mixture of independent *bona fide* cDC progenitors and pDC progenitors.

Despite their classification as precursors, human pre-DC appear functional in their own right, being equipped with some T-cell co-stimulatory molecules, and with a strong capacity for naïve T-cell allostimulation and cytokine secretion in response to TLR stimulation (Fig. 2, Fig. 5,

fig. S7, and fig. S19). Pre-DC produced low levels of IFN-α in response to CpG ODN 2216 exposure, and secreted IL-12 and TNF-α in response to various TLR ligands. Hence, it is reasonable to propose that pre-DC have the potential to contribute to both homeostasis and various pathological processes, particularly inflammatory and autoimmune diseases where dysregulation of their differentiation continuum or their arrested development could render them a potent source of inflammatory DC ready for rapid recruitment and mobilization.

Beyond the identification of pre-DC, our data revealed previously-unappreciated transcriptional and phenotypic heterogeneity within the circulating mature DC populations. This was particularly clear in the case of cDC2 and pDC, which were grouped into multiple Mpath clusters in the single-cell RNAseq analysis, and showed marked dispersion in the tSNE analysis of the CyTOF data with phenotypic heterogeneity. IsoMAP analysis of the CyTOF data also revealed another level of pDC heterogeneity by illustrating the progressive phenotypic transition from CDP into $CD2^+$ pDC in the BM, involving intermediate cells that could be pre-pDC. Whether a circulating pre-pDC population exists remains to be concluded. Finally, defining the mechanisms that direct the differentiation of uncommitted pre-DC into cDC1 or cDC2, or that maintain these cells in their initial uncommitted state in health and disease could lead to the development of new therapeutic strategies to modulate this differentiation process.

## Materials and Methods

### Blood, bone marrow and spleen samples

Human samples were obtained in accordance with a favorable ethical opinion from Singapore SingHealth and National Health Care Group Research Ethics Committees. Written informed consent was obtained from all donors according to the procedures approved by the National University of Singapore Institutional Review Board and SingHealth Centralised Institutional Review Board. Peripheral blood mononuclear cells (PBMC) were isolated by Ficoll-

Paque (GE Healthcare) density gradient centrifugation of apheresis residue samples obtained from volunteer donors through the Health Sciences Authorities (HSA, Singapore). Blood samples were obtained from 4 patients with molecularly confirmed Pitt-Hopkins syndrome (PHS), who all showed the classical phenotype (*54*). Spleen tissue was obtained from patients with tumors in the pancreas who underwent distal pancreatomy (Singapore General Hospital, Singapore). Spleen tissue was processed as previously described (*23*). Bone marrow mononuclear cells were purchased from Lonza.

Generation of single cell transcriptomes using MARS-seq

MARS-Seq using the Biomek FXP system (Beckman Coulter) as previously described (*15*) was performed for scmRNAseq of the DC compartment of the human peripheral blood. In brief, Lineage marker (Lin)(CD3/14/16/19/20/34)$^-$CD45$^+$CD135$^+$HLA-DR$^+$CD123$^+$CD33$^+$ single cells were sorted into individual wells of 384-well plates filled with 2 µl lysis buffer (Triton 0.2% (Sigma Aldrich) in molecular biology grade $H_2O$ (Sigma Aldrich), supplemented with 0.4 U/µl protein-based RNase inhibitor (Takara Bio Inc.), and barcoded using 400 nM IDT. Details regarding the barcoding procedure with poly-T primers were previously described (*15*). Samples were pre-incubated for 3 min at 80°C and reverse transcriptase mix consisting of 10 mM DTT (Invitrogen), 4 mM dNTPs (NEB), 2.5 U/µl SuperScript III Reverse Transcriptase (Invitrogen) in 50 mM Tris-HCl (pH 8.3; Sigma), 75 mM KCl (Sigma), 3 mM MgCl2 (Sigma), ERCC RNA Spike-In mix (Life Technologies), at a dilution of 1:80*$10^7$ per cell was added to each well. The mRNA was reverse-transcribed to cDNA with one cycle of 2 min at 42°C, 50 min at 50°C, and 5 min at 85°C. Excess primers were digested with *Exo*I (NEB) at 37°C for 30 min then 10 min at 80°C, followed by cleanup using SPRIselect beads at a 1.2x ratio (Beckman Coulter). Samples

were pooled and second strands were synthesized using a Second Strand Synthesis kit (NEB) for 2.5 h at 16°C, followed by a cleanup using SPRIselect beads at a 1.4x ratio (Beckman Coulter). Samples were linearly amplified by T7-promoter guided *in vitro* transcription using the T7 High Yield RNA polymerase IVT kit (NEB) at 37°C for 12 h. DNA templates were digested with Turbo DNase I (Ambion) for 15 min at 37°C, followed by a cleanup with SPRIselect beads at a 1.2x ratio (Beckman Coulter). The RNA was then fragmented in $Zn^{2+}$ RNA Fragmentation Solution (Ambion) for 1.5 min at 70°C, followed by cleanup with SPRIselect beads at a 2.0 ratio (Beckman Coulter). Barcoded ssDNA adapters (IDT; details of barcode see (*15*)) were then ligated to the fragmented RNAs in 9.5% DMSO (Sigma Aldrich), 1 mM ATP, 20% PEG8000 and 1 U/µl T4 RNA ligase I (NEB) solution in 50 mM Tris HCl pH7.5 (Sigma Aldrich), 10 mM MgCl2 and 1mM DTT for 2 h at 22°C. A second reverse transcription reaction was then performed using Affinity Script Reverse Transcription buffer, 10 mM DTT, 4 mM dNTP, 2.5 U/µl Affinity Script Reverse Transcriptase (Agilent) for one cycle of 2 min at 42°C, 45 min at 50°C, and 5 min at 85°C, followed by a cleanup on SPRIselect beads at a 1.5x ratio (Beckman Coulter). The final libraries were generated by subsequent nested PCR reactions using 0.5 µM of each Illumina primer (IDT; details of primers see (*15*)) and KAPA HiFi HotStart Ready Mix (Kapa Biosystems) for 15 cycles according to manufacturer's protocol, followed by a final cleanup with SPRIselect beads at a 0.7x ratio (Beckman Coulter). The quality and quantity of the resulting libraries was assessed using an Agilent 2200 TapeStation instrument (Agilent), and libraries were subjected to next generation sequencing using an Illumina HiSeq1500 instrument (PE no index; read1: 61 reads (3 reads random nucleotides, 4 reads pool barcode, 53 reads sequence), read2: 13 reads (6 reads cell barcode, 6 reads unique molecular identifier)).

Pre-processing, quality assessment and control of MARS-seq single cell transcriptome data

Cell specific tags and Unique Molecular Identifiers (UMIs) were extracted (2,496 cells sequenced) from sequenced data-pool barcodes. Sequencing reads with ambiguous plate and/or cell-specific tags, UMI sequences of low quality (Phred <27), or reads that mapped to *E. coli* were eliminated using Bowtie1 sequence analysis software (*55*), with parameters "-M 1 -t --best --chunkmbs 64 –strata". Fastq files were demultiplexed using the fastx_barcode_splitter from fastx_toolkit, and R1 reads (with trimming of pooled barcode sequences) were mapped to the human hg38 + ERCC pseudo genome assembly using Bowtie "-m 1 -t --best --chunkmbs 64 –strata". Valid reads were then counted using UMIs if they mapped to the exon-based gene model derived from the BiomaRt HG38 data mining tool provided by Ensembl (*56*). A gene expression matrix was then generated containing the number of UMIs for every cell and gene. Additionally, UMIs and cell barcode errors were corrected and filtered as previously described (*15*).

Normalization and filtering of MARS-seq single cell transcriptome data

In order to account for differences in total molecule counts per cell, we performed a down-sampling normalization as suggested by several studies (*15*, *57*). Here, we randomly down-sampled every cell to a molecule count of 1,050 unique molecules per cell (threshold details discussed below). Cells with molecule counts <1,050 were excluded from the analysis (Table S1 (*18*)). Additionally, cells with a ratio of mitochondrial versus endogenous genes exceeding 0.2, and cells with <90 unique genes, were removed from the analysis. Prior to Seurat analysis (*58*), expression tables were filtered to exclude mitochondrial and ribosomal genes to remove noise. The validation of down sampling threshold for normalization of MARS-seq single cell transcriptome data is detailed in the Supplementary Methods.

Analysis of MARS-seq single cell transcriptome data

Analysis of the normalized and filtered single-cell gene expression data (8,657 genes across 710 single cell transcriptomes used in the final expression table) was achieved using Mpath (*26*), PCA, tSNE, connectivity MAP (cMAP) (*25*) and several functions of the Seurat single cell analysis package. cMAP analysis was performed using DEGs between pDC and cDC derived from the gene expression omnibu data series GSE35457 (*23*). For individual cells, cMAP generated enrichment scores that quantified the degree of enrichment (or "closeness") to the given gene signatures. The enrichment scores were scaled and assigned positive or negative values to indicate enrichment for pDC or cDC signature genes, respectively. A permutation test (n=1,000) between gene signatures was performed on each enrichment score to determine statistical significance. For the tSNE/Seurat analysis, a Seurat filter was used to include genes that were detected in at least one cell (molecule count = 1), and excluded cells with <90 unique genes. To infer the structure of the single-cell gene expression data, a PCA was performed on the highly variable genes determined as genes exceeding the dispersion threshold of 0.75. The first two principle components were used to perform a tSNE that was combined with a DBSCAN clustering algorithm (*22*) to identify cells with similar expression profiles. DBSCAN was performed by setting 10 as the minimum number of reachable points and 4.1 as the reachable epsilon neighbourhood parameter; the latter was determined using a KNN plot integrated in the DBSCAN R package (*59*) (https://cran.r-project.org/web/packages/dbscan/). The clustering did not change when using the default minimal number of reachable points.

To annotate the clusters, we used the gene signatures of blood pDC, cDC1 and cDC2 derived from the Gene Expression Omnibus data series GSE35457 (*23*) (Table S2 (*24*), data processing described below) to calculate the signature gene expression scores of cell type-specific gene signatures, and then overlaid these signature scores onto the tSNE plots. Raw expression data of $CD141^+$ (cDC1), $CD1c^+$ (cDC2) DCs and pDC samples from blood of up to

four donors (I, II, V and VI) was imported into Partek® Genomics Suite® software, version 6.6 Copyright©; 2017 (PGS), where they were further processed. Data were quantile-normalized and log2-transformed, and a batch-correction was performed for the donor using PGS. Differential probe expression was calculated from the normalized data (ANOVA, Fold-Change ≥ 2 and FDR-adj. p-value < 0.05) for the three comparisons of every cell type against the remaining cell types. The three lists of differentially-expressed (DE) probes were intersected and only exclusively-expressed probes were used for the cell-type specific gene signatures. The probes were then reduced to single genes, by keeping the probe for a corresponding gene with the highest mean expression across the dataset. Resulting gene signatures for blood pDCs, $CD1c^+$ and $CD141^+$ DCs contained 725, 457 and 368 genes, respectively. The signature gene expression score was calculated as the mean expression of all signature genes in a cluster. In order to avoid bias due to outliers, we calculated the trimmed mean (trim = 0.08).

Monocle analysis was performed using the latest pre-published version of Monocle v.2.1.0 (*27*). The data were loaded into a monocle object and then log-transformed. Ordering of the genes was determined by dispersion analysis if they had an average expression of ≥0.5 and at least a dispersion of two times the dispersion fit. The dimensionality reduction was performed using the reduceDimension command with parameters max_components=2, reduction_method = "DDRTree" and norm_method = "log". The trajectory was then built using the plot_cell_trajectory command with standard parameters.

Wishbone analysis (*28*) was performed using the Python toolkit downloaded from https://github.com/ManuSetty/wishbone. MARS-seq data were loaded using the wishbone.wb.SCData.from_csv function with the parameters data_type='sc-seq' and normalize=True. Wishbone was then performed using wb.run_wishbone function with parameter start_cell= "run1_CATG_AAGACA", components_list=[1, 2, 3, 4], num_waypoints=150, branch = True. Start_cell was randomly selected from the central cluster #4. Diffusion map

analysis was performed using the scdata.run_diffusion_map function with default parameters (*29*). Wishbone revealed three trajectories giving rise to pDC, cDC1 and cDC2 respectively. Along each trajectory, the respective signature gene shows increasing expression (fig. S2c). Although Wishbone results might be interpreted to suggest that cDC2 are early cells and differentiate into pDC and cDC1 on two separate branches, this is simply because Wishbone allows a maximum of two branches and assumes all cells fall on continuous trajectories. Nevertheless, it is able to delineate the three trajectories that are in concordance with Mpath, monocle, and diffusion map analysis.

C1 Single cell mRNA sequencing

Lin(CD3/14/16/19/20)$^-$HLA-DR$^+$CD33$^+$CD123$^+$ cells at 300 cells/µl were loaded onto two 5–10 µm C1 Single-Cell Auto Prep integrated fluidic circuits (Fluidigm) and cell capture was performed according to the manufacturer's instructions. Individual capture sites were inspected under a light microscope to confirm the presence of single, live cells. Empty capture wells and wells containing multiple cells or cell debris were discarded for quality control. A SMARTer Ultra Low RNA kit (Clontech) and Advantage 2 PCR Kit (Clontech) was used for cDNA generation. An ArrayControl™ RNA Spots and Spikes kit (with spike numbers 1, 4 and 7) (Ambion) was used to monitor technical variability, and the dilutions used were as recommended by the manufacturer. The concentration of cDNA for each single cell was determined by Quant-iT™ PicoGreen® dsDNA Reagent, and the correct size and profile was confirmed using DNA High Sensitivity Reagent Kit and DNA Extended Range LabChip (Perkin Elmer). Multiplex sequencing libraries were generated using the Nextera XT DNA Library Preparation Kit and the Nextera XT Index Kit (Illumina). Libraries were pooled and subjected to an indexed PE sequencing run of 2x51 cycles on an Illumina HiSeq 2000 (Illumina) at an average depth of 2.5-million row reads/cell.

C1 Single cell analysis

Raw reads were aligned to the human reference genome GRCh38 from GENCODE (*60*) using RSEM program version 1.2.19 with default parameters (*61*). Gene expression values in transcripts per million were calculated using the RSEM program and the human GENCODE annotation version 22. Quality control and outlier cell detection was performed using the *SINGuLAR* (Fluidigm) analysis *toolset.* cMAP analysis was performed using cDC1 versus cDC2 DEGs identified from Gene Expression Omnibus data series GSE35457 (*23*), and the enrichment scores were obtained. Similar to the gene set enrichment analyses, cMAP was used to identify associations of transcriptomic profiles with cell-type characteristic gene signatures.

Mpath analysis of MARS- or C1 single cell mRNA sequencing data

Developmental trajectories were defined using the Mpath algorithm (*26*), which constructs multi-branching cell lineages and re-orders individual cells along the branches. In the analysis of the MARS-seq single cell transcriptomic data, we first used the Seurat R package to identify five clusters: for each cluster, Mpath calculated the centroid and used it as a landmark to represent a canonical cellular state; subsequently, for each single cell, Mpath calculated its Euclidean distance to all the landmarks, and identified the two nearest landmarks. Each individual cell was thus assigned to the neighborhood of its two nearest landmarks. For every pair of landmarks, Mpath then counted the number of cells that were assigned to the neighborhood, and used the determined cell counts to estimate the possibility of the transition between landmarks to be true. A high cell count implied a high possibility that the transition was valid. Mpath then constructed a weighted neighborhood network whereby nodes represented landmarks, edges represented a putative transition between landmarks, and numbers allocated to the edges represented the cell-count support for the transition. Given that single cell transcriptomic data tend to be noisy, edges

with low cell-count support were considered likely artifacts. Mpath therefore removed the edges with a low cell support by using $(1 - n)$ ($n$ represents cell count) to quantify the distance between nodes followed by applying a minimum spanning tree algorithm to find the shortest path that could connect all nodes with the minimum sum of distance. Consequently, the resulting trimmed network is the one that connects all landmarks with the minimum number of edges and the maximum total number of cells on the edges. Mpath was then used to project the individual cells onto the edge connecting its two nearest landmarks, and assigned a pseudo-time ordering to the cells according to the location of their projection points on the edge. In the analysis of the C1 single cell transcriptome data, we first used the cMAP analysis to identify cDC1-primed, un-primed, and cDC2-primed clusters, and then used Mpath to construct the lineage between these three clusters. The Mpath analysis was carried out in an un-supervised manner without prior knowledge of the starting cells or number of branches. This method can be used for situations of non-branching networks, bifurcations, and multi-branching networks with three or more branches.

Mass cytometry staining, barcoding, acquisition and data analysis

For mass cytometry, pre-conjugated or purified antibodies were obtained from Invitrogen, Fluidigm (pre-conjugated antibodies), Biolegend, eBioscience, Becton Dickinson or R&D Systems as listed in Table S3. For some markers, fluorophore- or biotin- conjugated antibodies were used as primary antibodies, followed by secondary labeling with anti-fluorophore metal-conjugated antibodies (such as the anti-FITC clone FIT-22) or metal-conjugated streptavidin, produced as previously described (*19*). Briefly, 3 x $10^6$ cells/well in a U-bottom 96 well plate (BD Falcon, Cat# 3077) were washed once with 200 µL FACS buffer (4% FBS, 2mM EDTA, 0.05% Azide in 1X PBS), then stained with 100 µL 200 µM cisplatin (Sigma-Aldrich, Cat# 479306-1G) for 5 min on ice to exclude dead cells. Cells were then incubated with anti-CADM1-

biotin and anti-CD19-FITC primary antibodies in a 50 µL reaction for 30 min on ice. Cells were washed twice with FACS buffer and incubated with 50 µL heavy-metal isotope-conjugated secondary mAb cocktail for 30 min on ice. Cells were then washed twice with FACS buffer and once with PBS before fixation with 200 µL 2% paraformaldehyde (PFA; Electron Microscopy Sciences, Cat# 15710) in PBS overnight or longer. Following fixation, the cells were pelleted and resuspended in 200uL 1X permeabilization buffer (Biolegend, Cat# 421002) for 5 mins at room temperature to enable intracellular labeling. Cells were then incubated with metal-conjugated anti-CD68 in a 50 µL reaction for 30 min on ice. Finally, the cells were washed once with permeabilization buffer and then with PBS before barcoding.

Bromoacetamidobenzyl-EDTA (BABE)-linked metal barcodes were prepared by dissolving BABE (Dojindo, Cat# B437) in 100mM HEPES buffer (Gibco, Cat# 15630) to a final concentration of 2 mM. Isotopically-purified $PdCl_2$ (Trace Sciences Inc.) was then added to the 2 mM BABE solution to a final concentration of 0.5 mM. Similarly, DOTA-maleimide (DM)-linked metal barcodes were prepared by dissolving DM (Macrocyclics, Cat# B-272) in L buffer (MAXPAR, Cat# PN00008) to a final concentration of 1 mM. $RhCl_3$ (Sigma) and isotopically-purified $LnCl_3$ was then added to the DM solution at 0.5 mM final concentration. Six metal barcodes were used: BABE-Pd-102, BABE-Pd-104, BABE-Pd-106, BABE-Pd-108, BABE-Pd-110 and DM-Ln-113.

All BABE and DM-metal solution mixtures were immediately snap-frozen in liquid nitrogen and stored at -80$^{o}$C. A unique dual combination of barcodes was chosen to stain each tissue sample. Barcode Pd-102 was used at 1:4000 dilution, Pd-104 at 1:2000, Pd-106 and Pd-108 at 1:1000, Pd-110 and Ln-113 at 1:500. Cells were incubated with 100 µL barcode in PBS for 30 min on ice, washed in permeabilization buffer and then incubated in FACS buffer for 10 min on ice. Cells were then pelleted and resuspended in 100 µL nucleic acid Ir-Intercalator (MAXPAR, Cat# 201192B) in 2% PFA/PBS (1:2000), at room temperature. After 20 min, cells

were washed twice with FACS buffer and twice with water before a final resuspension in water. In each set, the cells were pooled from all tissue types, counted, and diluted to $0.5x10^6$ cells/mL. EQ Four Element Calibration Beads (DVS Science, Fluidigm) were added at a 1% concentration prior to acquisition. Cell data were acquired and analyzed using a CyTOF Mass cytometer (Fluidigm).

The CyTOF data were exported in a conventional flow-cytometry file (.fcs) format and normalized using previously-described software (*62*). Events with zero values were randomly assigned a value between 0 and –1 using a custom R script employed in a previous version of mass cytometry software (*63*). Cells for each barcode were deconvolved using the Boolean gating algorithm within FlowJo. The $CD45^+$Lin (CD7/CD14/CD15/CD16/CD19/CD34)$^-$HLA-$DR^+$ population of PBMC were gated using FlowJo and exported as a .fcs file. Marker expression values were transformed using the logicle transformation function (*64*). Random sub-sampling without replacement was performed to select 20,000 cell events. The transformed values of sub-sampled cell events were then subjected to t-distributed Stochastic Neighbor Embedding (tSNE) dimension reduction (*21*) using the markers listed in supplementary Table S3, and the Rtsne function in the Rtsne R package with default parameters. Similarly, isometric feature mapping (isoMAP) (*43*) dimension reduction was performed using vegdist, spantree and isomap functions in the vegan R package (*65*).

The vegdist function was run with method="euclidean". The spantree function was run with default parameters. The isoMAP function was run with ndim equal to the number of original dimensions of input data, and k=5. Phenograph clustering (*30*) was performed using the markers listed in supplementary Table S3 before dimension reduction, and with the number of nearest neighbors equal to 30. The results obtained from the tSNE, isoMAP and Phenograph analyses were incorporated as additional parameters in the .fcs files, which were then loaded into FlowJo to generate heat plots of marker expression on the reduced dimensions. The above analyses were

performed using the cytofkit R package which provides a wrapper of existing state-of-the-art methods for cytometry data analysis (*66*).

Human cell flow cytometry: Labeling, staining, analysis and cell sorting

All antibodies used for fluorescence-activated cell sorting (FACS) and flow cytometry were mouse anti-human monoclonal antibodies (mAbs), except for chicken anti-human CADM1 IgY primary mAb. The mAbs and secondary reagents used for flow cytometry are listed in Table S6. Briefly, 5 x $10^6$ cells/tube were washed and incubated with Live/Dead blue dye (Invitrogen) for 30 min at 4℃ in phosphate buffered saline (PBS) and then incubated in 5% heat-inactivated fetal calf serum (FCS) for 15 min at 4℃ (Sigma Aldrich). The appropriate antibodies diluted in PBS with 2% fetal calf serum (FCS) and 2 mM EDTA were added to the cells and incubated for 30 min at 4℃, and then washed and detected with the secondary reagents. For intra-cytoplasmic or intra-nuclear labeling or staining, cells were fixed and permeabilized with BD Cytofix/Cytoperm (BD Biosciences) or with eBioscience FoxP3/Transcription Factor Staining Buffer Set (eBioscience/Affimetrix), respectively according to the manufacturer's instructions. Flow cytometry was performed using a BD LSRII or a BD FACSFortessa (BD Biosciences) and the data analyzed using BD FACSDiva 6.0 (BD Biosciences) or FlowJo v.10 (Tree Star Inc.). For isolation of precursor dendritic cells (pre-DC), PBMC were first depleted of T cells, monocytes and B cells with anti-CD3, anti-CD14 and anti-CD20 microbeads (Miltenyi Biotec) using an AutoMACS Pro Separator (Miltenyi Biotec) according to the manufacturer's instructions. FACS was performed using a BD FACSAriaII or BD FACSAriaIII (BD Biosciences). Wanderlust analysis (*41*) of flow cytometry data was performed using the CYT tool downloaded from https://www.c2b2.columbia.edu/danapeerlab/html/cyt-download.html. As Wanderlust requires users to specify a starting cell, we selected one cell at random from the $CD45RA^+$ $CD123^+$ population.

Cytospin and Scanning Electron Microscopy

Cytospins were prepared from purified cells and stained with the Hema 3 system according to the manufacturer's protocol (Fisher Diagnostics). Images were analyzed at 100X magnification with an Olympus BX43 upright microscope (Olympus). Scanning electron microscopy was performed as previously described (*23*).

Dendritic cell (DC) differentiation co-culture assay on MS-5 stromal cells

MS-5 stromal cells were maintained and passaged as previously described (*8*). MS-5 cells were seeded in 96-well round-bottom plates (Corning) at 3,000 cells per well in complete alpha-Minimum Essential Media (α-MEM) (Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Serana) and 1% penicillin/streptomycin (Nacalai Tesque). A total of 5,000 sorted purified cells were added 18-24 h later, in medium containing 200 ng/mL Flt3L (Miltenyi Biotec), 20 ng/mL SCF (Miltenyi Biotec), and 20 ng/mL GM-CSF (Miltenyi Biotec), and cultured for up to 5 days. The cells were then resuspended in their wells by physical dissociation and filtered through a cell strainer into a polystyrene FACS tube.

Intracellular cytokine detection following stimulation with TLR ligands

A total of $5x10^6$ PBMC were cultured in Roswell Park Memorial Institute (RPMI)-1640 Glutmax media (Life Technologies) supplemented with 10% FBS, 1% penicillin/streptomycin and stimulated with either lipopolysaccharide (LPS, 100ng/mL; InvivoGen), LPS (100ng/mL) + interferon gamma (IFNγ, 1,000U/mL; R&D Systems), Flagellin (100 ng/mL, Invivogen), polyI:C (10 μg/mL; InvivoGen), Imidazoquinoline (CL097; Invivogen) or CpG oligodeoxynucleotides 2216 (ODN, 5 μM; InvivoGen) for 2 h, after which 10 μg/ml Brefeldin A solution (eBioscience) was added and the cells were again stimulated for an additional 4 h. After the 6 h stimulation, the

cells were labeled with cytokine-specific antibodies and analyzed by flow cytometry, as described above.

Mixed lymphocyte reaction

Naïve T cells were isolated from PBMC using Naïve Pan T-Cell Isolation Kit (Miltenyi Biotec) according to the manufacturer's instructions, and labeled with 0.2μM carboxyfluorescein succinimidyl ester (CFSE) (Life Technologies) for 5 min at 37 °C. A total of 5,000 cells from sorted DC subsets were co-cultured with 100,000 CFSE-labeled naïve T cells for 7 days in Iscove's Modified Dulbecco's Medium (IMDM; Life Technologies) supplemented with 10% KnockOut™ Serum Replacement (Life Technologies). On day 7, the T cells were stimulated with 10 μg/ml phorbol myristate acetate (InvivoGen) and 500 μg/ml ionomycin (Sigma Aldrich) for 1 h at 37 °C. 10 μg/ml Brefeldin A solution was added for 4 h, after which the cells were labeled with cytokine-specific antibodies and analyzed by flow cytometry, as described above.

Electron microscopy

Sorted cells were seeded on poly-lysine-coated coverslips for 1 h at 37°C. The cells were then fixed in 2 % glutaraldehyde in 0.1 M cacoldylate buffer, pH 7.4 for 1 h, post fixed for 1 h with 2% buffered osmium tetroxide, then dehydrated in a graded series of ethanol solutions, before embedding in epoxy resin. Images were acquired with a Quemesa (SIS) digital camera mounted on a Tecnai 12 transmission electron microscope (FEI Company) operated at 80kV.

Microarray analysis

Total RNA was isolated from FACS-sorted blood pre-DC and DC subsets using a RNeasy® Micro kit (Qiagen). Total RNA integrity was assessed using an Agilent Bioanalyzer (Agilent) and the RNA Integrity Number (RIN) was calculated. All RNA samples had a RIN ≥7.1.

Biotinylated cRNA was prepared using an Epicentre TargetAmp™ 2-Round Biotin-aRNA Amplification Kit 3.0 according to the manufacturer's instructions, using 500 pg of total RNA starting material. Hybridization of the cRNA was performed on an Illumina Human-HT12 Version 4 chip set (Illumina). Microrarray data were exported from GenomeStudio (Illumina) without background subtraction. Probes with detection P-values > 0.05 were considered as not being detected in the sample, and were filtered out. Expression values for the remaining probes were log2 transformed and quantile normalized. For differentially-expressed gene (DEG) analysis, comparison of one cell subset with another was carried out using the limma R software package (*67*) with samples paired by donor identifiers. DEGs were selected with Benjamini-Hochberg multiple testing (*68*) corrected P-value <0.05. In this way, limma was used to select up and down-regulated signature genes for each of the cell subsets in the pre-DC data by comparing one subset with all other subsets pooled as a group. Expression profiles shown in Fig. 4E were from 62 common genes identified from the union of DEGs from comparing pre-cDC1 versus early pre-DC and cDC1 versus pre-cDC1, and the union of DEGs from comparing pre-cDC2 versus early pre-DC and cDC2 versus pre-cDC2 (Table S5 (*69*), fig. S17 (*70*)).

Luminex® Drop Array™ assay on sorted and stimulated pre-DC and DC populations

A total of 2,000 cells/well of sorted pre-DC and DC subsets were seeded in V-bottom 96 well plates and then incubated for 18 h in 50 µL complete RPMI-1640 Glutmax media (Life Technologies) supplemented with 10% FBS and 1% penicillin/streptomycin, and stimulated with either LPS, LPS + IFNγ, Flagellin, polyI:C, Imidazoquinoline or CpG oligodeoxynucleotides (ODN) 2216. Cells were then pelleted and 30 µL supernatant was collected. A Luminex® Drop Array™ was performed using 5 µL of the supernatant. Human G-CSF, GM-CSF, IFN-a2, IL-10, IL-12p40, IL-12p70, IL-15, IL-1RA, IL-1a, IL-1b, IL-6, IL-7, IL-8, MIP-1b, TNF-α, TNF-β were tested by multiplexing (EMD Millipore) with DropArray-bead plates (Curiox) according to

the manufacturer's instructions. Acquisition was performed using xPONENT 4.0 (Luminex) acquisition software, and data analysis was performed using Bio-Plex Manager 6.1.1 (Bio-Rad).

Statistical analyses

The Mann-Whitney test was used to compare data derived from patients with Pitt-Hopkins Syndrome and controls and the intracellular detection of IL-12p40 and TNFα in pre-DC stimulated with LPS or poly I:C versus CpG ODN 2216. The Kruskal-Wallis test, followed by the Dunn's multiple comparison test, was used to compare the expression level of individual genes in single cells in the MARS-seq single cell RNAseq dataset. Differences were defined as statistically significant when adjusted $P<0.05$. All statistical tests were performed using GraphPad Prism 6.00 for Windows (GraphPad Software). Correlation coefficients were calculated as Pearson's correlation coefficient.

## REFERENCES


1. A. Schlitzer, N. McGovern, F. Ginhoux, Dendritic cells and monocyte-derived cells: Two complementary and integrated functional systems. Semin. Cell Dev. Biol. 41, 9–22 (2015).

2. M. Merad, P. Sathe, J. Helft, J. Miller, A. Mortha, The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. Annu. Rev. Immunol. 31, 563–604 (2013).

3. M. Guilliams et al., Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. Nat. Rev. Immunol. 14, 571–578 (2014).

4. K. Liu et al., In vivo analysis of dendritic cell development and homeostasis. Science. 324, 392–397 (2009).

5. F. Ginhoux et al., The origin and development of nonlymphoid tissue CD103+ DCs. J. Exp. Med. 206, 3115–3130 (2009).

6. N. Onai et al., A clonogenic progenitor with prominent plasmacytoid dendritic cell developmental potential. Immunity. 38, 943–957 (2013).

7. A. Schlitzer et al., Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. Nat. Immunol. 16, 718–728 (2015).

8. J. Lee et al., Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow. J. Exp. Med. 212, 385–399 (2015).

9. G. Breton et al., Circulating precursors of human CD1c+ and CD141+ dendritic cells. J. Exp. Med. 212, 401–413 (2015).

10. M. Cella et al., Plasmacytoid monocytes migrate to inflamed lymph nodes and produce large amounts of type I interferon. Nat. Med. 5, 919–923 (1999).

11. G. Grouard et al., The enigmatic plasmacytoid T cells develop into dendritic cells with interleukin (IL)-3 and CD40-ligand. J. Exp. Med. 185, 1101–1111 (1997).

12. S. Doulatov et al., Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nat. Immunol. 11, 585–593 (2010).

13. J. D. Griffin et al., Differential expression of HLA-DR antigens in subsets of human CFU-GM. Blood. 66, 788–795 (1985).

14. A. Dzionek et al., BDCA-2, BDCA-3, and BDCA-4: Three Markers for Distinct Subsets of Dendritic Cells in Human Peripheral Blood. The Journal of Immunology. 165, 6037–6046 (2000).

15. D. A. Jaitin et al., Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science. 343, 776–779 (2014).

16. Sorting strategy from total $Lin^{-}HLA\text{-}DR^{+}CD135^{+}$ cells for MARS-seq.

17. Workflow & quality control of MARS-seq single cell data analysis.

18. Number of detected genes per cell in the total DC MARS-seq experiment.

19. B. Becher et al., High-dimensional analysis of the murine myeloid cell system. Nat. Immunol. 15, 1181–1189 (2014).

20. E.-A. D. Amir et al., viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat. Biotechnol. 31, 545–552 (2013).

21. L. Van der Maaten, Visualizing data using t-SNE. Journal of Machine Learning Research. 9, 2579–2625 (2008).

22. M. Ester, H. P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd (1996).

23. M. Haniffa et al., Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells. Immunity. 37, 60–73 (2012).

24. DC subset signature genes derived from Gene Expression Omnibus data series GSE35457 and used for MARS-seq and C1 data analyses.

25. J. Lamb, The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. Science. 313, 1929–1935 (2006).

26. J. Chen, A. Schlitzer, S. Chakarov, F. Ginhoux, M. Poidinger, Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. Nat Commun. 7, 11988 (2016).

27. C. Trapnell et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386 (2014).

28. M. Setty et al., Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat. Biotechnol. 34, 637–645 (2016).

29. R. R. Coifman et al., Geometric diffusions as a tool for harmonic analysis and structure definition of data: multiscale methods. Proceedings of the National Academy of Sciences. 102, 7432–7437 (2005).

30. J. H. Levine et al., Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell. 162, 184–197 (2015).

31. C. Sadaka, M.-A. Marloie-Provost, V. Soumelis, P. Benaroch, Developmental regulation of MHC II expression and transport in human plasmacytoid-derived dendritic cells. Blood. 113, 2127–2135 (2009).

32. S. L. Jongbloed et al., Human CD141+ (BDCA-3)+ dendritic cells (DCs) represent a unique myeloid DC subset that cross-presents necrotic cell antigens. J. Exp. Med. 207, 1247–1260 (2010).

33. K. P. A. MacDonald et al., Characterization of human blood dendritic cell subsets. Blood. 100, 4512–4520 (2002).

34. T. Matsui et al., CD2 distinguishes two subsets of human plasmacytoid dendritic cells with distinct phenotype and functions. J. Immunol. 182, 6815–6823 (2009).

35. H. Yu et al., Human BDCA2(+)CD123 (+)CD56 (+) dendritic cells (DCs) related to blastic plasmacytoid dendritic cell neoplasm represent a unique myeloid DC subset. Protein Cell. 6, 297–306 (2015).

36. B. Reizis, A. Bunin, H. S. Ghosh, K. L. Lewis, V. Sisirak, Plasmacytoid dendritic cells: recent progress and open questions. Annu. Rev. Immunol. 29, 163–183 (2011).

37. B. Cisse et al., Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. Cell. 135, 37–48 (2008).

38. Sorting strategy of $Lin^{-}HLA\text{-}DR^{+}CD33^{+}CD45RA^{+}CD1c^{lo/-}CD2^{+}CADM1^{lo/-}CD123^{+}$ pre-DC for C1 scmRNAseq.

39. Workflow & quality control of C1 scmRNAseq data analysis.

40. Number of expressed genes detected per cell in the pre-DC C1 scmRNAseq experiment.

41. S. C. Bendall et al., Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell. 157, 714–725 (2014).

42. M. Swiecki, M. Colonna, The multifaceted biology of plasmacytoid dendritic cells. Nat. Rev. Immunol. 15, 471–485 (2015).

43. J. B. Tenenbaum, V. de Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction. Science. 290, 2319–2323 (2000).

44. G. E. Grajales-Reyes et al., Batf3 maintains autoactivation of Irf8 for commitment of a CD8α(+) conventional DC clonogenic progenitor. Nat. Immunol. 16, 708–717 (2015).

45. A. Schlitzer et al., Identification of CCR9- murine plasmacytoid DC precursors with plasticity to differentiate into conventional DCs. Blood. 117, 6562–6570 (2011).

46. A. Krug et al., Toll-like receptor expression reveals CpG DNA as a unique microbial stimulus for plasmacytoid dendritic cells which synergizes with CD40 ligand to induce high amounts of IL-12. Eur. J. Immunol. 31, 3026–3037 (2001).

47. A. Dzionek et al., Plasmacytoid dendritic cells: from specific surface markers to specific cellular functions. Hum. Immunol. 63, 1133–1148 (2002).

48. M. Cella, F. Facchetti, A. Lanzavecchia, M. Colonna, Plasmacytoid dendritic cells activated by influenza virus and CD40L drive a potent TH1 polarization. Nat. Immunol. 1, 305–310 (2000).

49. T. Ito et al., Plasmacytoid dendritic cells prime IL-10-producing T regulatory cells by inducible costimulator ligand. J. Exp. Med. 204, 105–115 (2007).

50. J.-F. Fonteneau et al., Activation of influenza virus-specific CD4+ and CD8+ T cells: a new role for plasmacytoid dendritic cells in adaptive immunity. Blood. 101, 3520–3526 (2003).

51. G. Hoeffel et al., Antigen crosspresentation by human plasmacytoid dendritic cells. Immunity. 27, 481–492 (2007).

52. Y.-J. Liu, IPC: professional type 1 interferon-producing cells and plasmacytoid dendritic cell precursors. Annu. Rev. Immunol. 23, 275–306 (2005).

53. F. P. Siegal et al., The Nature of the Principal Type 1 Interferon-Producing Cells in Human Blood. Science. 284, 1835–1837 (1999).

54. C. F. de Winter et al., Phenotype and natural history in 101 individuals with Pitt-Hopkins syndrome through an internet questionnaire system. Orphanet J Rare Dis. 11, 37 (2016).

55. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).

56. A. Yates et al., Ensembl 2016. Nucleic Acids Res. 44, D710–6 (2016).

57. D. Grün et al., Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature. 525, 251–255 (2015).

58. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502 (2015).

59. M. Hahsler, M. Piekenbrock, dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms. R package version 1.0-0. https://CRAN.R-project.org/package=dbscan (2017).

60. J. Harrow et al., GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22, 1760–1774 (2012).

61. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12, 323 (2011).

62. R. Finck et al., Normalization of mass cytometry data with bead standards. Cytometry A. 83, 483–494 (2013).

63. E. W. Newell, N. Sigal, S. C. Bendall, G. P. Nolan, M. M. Davis, Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes. Immunity. 36, 142–152 (2012).

64. D. R. Parks, M. Roederer, W. A. Moore, A new “Logicle” display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. Cytometry A. 69, 541–551 (2006).

65. J. Oksanen et al., vegan: Community Ecology Package. R package version 2.4-2.

https://CRAN.R-project.org/package=vegan (2017).

66. H. Chen et al., Cytofkit: A Bioconductor Package for an Integrated Mass Cytometry Data Analysis Pipeline. PLoS Comput Biol. 12, e1005112 (2016).

67. G. K. Smyth, Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 3, Article3 (2004).

68. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society Series B ( … (1995).

69. Lists of genes identified from the microarray DEG analysis comparisons along the lineage progression from early pre-DC to mature cDC, for cDC1 and cDC2 respectively, and the list of the 62 common genes.

70. Venn diagram comparison of the two lists of DEGs and identification of the 62 common genes.

71. E. Mass et al., Specification of tissue-resident macrophages during organogenesis. Science. 353 (2016), doi:10.1126/science.aaf4238.

72. G. X. Y. Zheng et al., Massively parallel digital transcriptional profiling of single cells. Nat Commun. 8, 14049 (2017).

73. F. Paul et al., Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell. 163, 1663–1677 (2015).

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

75. J. A. Donovan, G. A. Koretzky, CD45 and the immune response. J. Am. Soc. Nephrol. 4, 976–985 (1993).

76. L. Ziegler-Heitbrock et al., Nomenclature of monocytes and dendritic cells in blood. Blood. 116, e74–80 (2010).

77. F. Nakayama et al., CD15 expression in mature granulocytes is determined by alpha 1,3-fucosyltransferase IX, but in promyelocytes and monocytes by alpha 1,3-fucosyltransferase IV. J. Biol. Chem. 276, 16100–16106 (2001).

78. J. M. Milush et al., Functionally distinct subsets of human NK cells and monocyte/DC-like cells identified by coexpression of CD56, CD7, and CD4. Blood. 114, 4823–4831 (2009).

79. F. Schütz, H. Hackstein, Identification of novel dendritic cell subset markers in human blood. Biochem. Biophys. Res. Commun. 443, 453–457 (2014).

80. B. Morandi et al., Distinctive Lack of CD48 Expression in Subsets of Human Dendritic Cells Tunes NK Cell Activation. The Journal of Immunology. 175, 3690–3697 (2005).

81. H. Strobl et al., Identification of CD68+lin- peripheral blood cells with dendritic precursor characteristics. The Journal of Immunology. 161, 740–748 (1998).

82. L. Galibert et al., Nectin-like protein 2 defines a subset of T-cell zone dendritic cells and is a ligand for class-I-restricted T-cell-associated molecule. J. Biol. Chem. 280, 21955–21964 (2005).

83. M. Cheng et al., Characterization of species-specific genes regulated by E2-2 in human plasmacytoid dendritic cells. Sci Rep. 5, 10752 (2015).

84. M. Haniffa, V. Bigley, M. Collin, Human mononuclear phagocyte system reunited. Semin. Cell Dev. Biol. 41, 59–69 (2015).

85. L. F. Poulin et al., Characterization of human DNGR-1+ BDCA3+ leukocytes as putative equivalents of mouse CD8alpha+ dendritic cells. J. Exp. Med. 207, 1261–1271 (2010).

86. P. B. Watchmaker et al., Comparative transcriptional and functional profiling defines conserved programs of intestinal DC differentiation in humans and mice. Nat. Immunol. 15, 98–108 (2014).

87. C. Weber et al., Differential chemokine receptor expression and function in human monocyte subpopulations. J. Leukoc. Biol. 67, 699–704 (2000).

88. U. O'Doherty et al., Human blood contains two subsets of dendritic cells, one immunologically mature and the other immature. Immunology. 82, 487–493 (1994).

**Acknowledgments:**

We would like to thank L. Robinson of Insight Editing London for critical review and editing of the manuscript; P.Y.J Ai from the SingHealth Flow Cytometry Core Platform; M.L.Ng, S.H.Tan, and T.B. Lu from the Electron Microscope Unit of NUS for their assistance. This work was supported by Singapore Immunology Network core funding (F.G, E.N.), Agency for Science, Technology and Research (A*STAR), Singapore, the A*STAR Graduate Scholarship (P.S.), the Wellcome Trust (WT 107931/Z/15/Z) (M.H.) and by the National Research Foundation Singapore under its cooperative basic research grant new investigator grant (NMRC/BNIG/2026/2014) and administered by the Singapore Ministry of Health's National Medical Research Council (C-A.D). This work was supported by the French National Research Agency through the "Investments for the Future" program (France-BioImaging, ANR-10-INSB-

04), LABEX DCBIOL (ANR-10-IDEX-0001-02 PSL and ANR-11-LABX-0043) and by grants from «Agence Nationale de Recherche contre le SIDA et les hépatites virales» (ANRS) (P.B, N.R, M.J and E.G.M). J.L.S, M.B, and A.S are members of the Excellence Cluster ImmunoSensation. J.L.S is funded by Sonderforschungsbereich 645 and 704. A.S is funded by an Emmy-Noether fellowship (SCHL 2116/1-1) of the German research foundation and a Young Investigator Award of the Biomedical Research Council Singapore. F.G and P.S are inventors on patent application (10201607246S) held by A*STAR that covers the methods for the identification, targeting, isolation of human dendritic cell (DC) precursors "preDC" and their use as biomarkers of inflammatory diseases. The MARS-seq, microfluidic scmRNAseq and microarray datasets are deposited in the Genome Expression Omnibus under accession number GSE98052, GSE98011 and GSE80171, respectively.

**Supplementary Materials:**

Materials and Methods

Figs. S1 to S22

Tables S1 to S6

Additional References (71-88)

**Figure legends:**

**Fig. 1. MARS-seq and CyTOF identify rare CD123$^{+}$CD33$^{+}$ putative DC precursors (pre-DC).** (**A-E**) Lin(CD3/CD14/CD16/CD20/CD34)$^{-}$HLA-DR$^{+}$CD135$^{+}$ sorted PBMC were subjected to MARS-seq. (**A**) t-stochastic neighbor embedding (tSNE) plot of 710 cells fulfilling all quality criteria, colored by clusters identified by tSNE plus Seurat clustering, or by the relative signature score for pDC, cDC1 and cDC2. (**B**) Connectivity MAP (cMAP) analysis

showing the degree of enrichment for pDC or cDC signature genes in the tSNE/Seurat clusters. (**C**) Mpath analysis applied to the tSNE/Seurat clusters defining their developmental relationship. Representations of the 710 cells by (**D**) Monocle, (**E**) Principal component analysis (PCA) and (**F**) Diffusion Map, highlighting the tSNE/Seurat clusters identified in (**A**). (**G**) Violin plots of tSNE/Seurat pDC clusters, cluster #4 and cDC clusters showing the expression of pDC and cDC signature genes with differential expression between cluster #4 and pDC clusters. Adjusted P-values calculated by Kruskal-Wallis test followed by Dunn's multiple comparisons procedure. (**H, I**) tSNE plots of CyTOF data from $CD45^+$Lin(CD7/CD14/CD15/CD16/CD19/CD34)$^-$HLA-$DR^+$ PBMC, showing (**H**) gates defining the $CD123^+CD33^+$ cells and DC subsets, and (**I**) relative expression of selected markers. (**J**) Subsets defined in **(H)** were overlaid onto 2D-contour plots for phenotypic comparison. The gating strategy prior to MARS-seq is shown in fig. S1a.

**Fig. 2**. **Characterization of human pre-DC.** (**A**) Flow cytometric identification of pre-DC and pDC within PBMC and spleen cell suspensions. (**B**) Expression of CD303/CD304/CD123/CD11c by blood pre-DC and DC subsets. (**C**) % pre-DC within spleen (n=3) and PBMC (n=6) $CD45^+$ populations. (**D**) Wright-Giemsa staining of sorted blood pre-DC and DC subsets. (**E**) Electron micrographs of pre-DC and pDC [(RER (arrowheads), centriole (C) and microtubules (small arrows), near RER cisterna are indicated). (**F**) DC subsets or pre-DC were co-cultured for 5 days with MS-5 feeder-cells, FLT3L, GM-CSF and SCF. Their capacity to differentiate into cDC1 or cDC2 was measured by flow cytometry. (n=3) (**G**) Intracellular detection of cytokines in DC subsets and pre-DC post-TLR stimulation. IFNα and IL-12p40 production by pDC and pre-DC, alongside mean % cytokine-positive pre-DC and DC subsets exposed to LPS, LPS+IFNγ (L+I), polyI:C (pI:C), CL097 (CL) or CpG-ODN2216 (CpG) (n=4). (**H**) Proliferation of naïve $CD4^+$ T cells cultured for 6 days with allogeneic pDC, total $CD123^+$HLA-$DR^+$ cells or pre-DC (n=2). (**I**) Frequency of pDC and pre-DC from control

subjects (Ctrl, n=11) and Pitt-Hopkins Syndrome (PHS) patients (n=4). P-values calculated by Mann-Whitney test. Error bars represent mean +/- SEM.

**Fig. 3. Identification of committed human pre-DC subsets.** (**A-B**) Single-cell mRNA sequencing (scmRNAseq) of 92 Lin(CD3/14/16/19/20)$^-$HLA-DR$^+$CD33$^+$CD123$^+$ cells (sort gating strategy in fig. S8a). (**A**) Connectivity MAP (cMAP) enrichment score of cells (cDC1- vs cDC2-specific signatures). (**B**) Mpath analysis showing the developmental relationship between "unprimed", cDC1-primed or cDC2-primed cells defined in (**A**). (**C**) Lin$^-$HLA-DR$^+$CD33$^+$ PBMC analyzed by flow cytometry and visualized as 3D-PCA of three cell clusters (pre-DC, cDC1 and cDC2) and the relative expression of CADM1, CD1c and CD123. (**D**) Relative expression of CD45RA, BTLA, CD327, CD141 and CD5 in the same 3D-PCA plot. The dashed black circles indicate the intermediate CD45RA$^+$ population. (**E**) CD45RA/CD123 dot plots showing overlaid cell subsets defined in the 3D-PCA plot (left panel) with the relative expression of BTLA, CD327, CD141 and CD5. (**F**) Overlay of the Wanderlust dimension (progression from early (dark) to late (clear) events is shown) onto the 3D-PCA and CD45RA/CD123 dot plots. (**G**) Gating strategy starting from live CD45$^+$Lin(CD3/14/16/19/20)$^-$CD34$^-$HLA-DR$^+$ PBMC to define pre-DC subsets among CD33$^+$CD45RA$^+$ cDC. (**H**) Pre-DC subsets were co-cultured for 5 days with MS-5 feeder-cells, FLT3L, GM-CSF and SCF (n=3). Their capacity to differentiate into Clec9A$^+$CADM1$^+$ cDC1 (red), or CD1c$^+$CD11c$^+$ cDC2 (beige) was analyzed by flow cytometry. (**I**) Scanning electron microscopy of pre-DC and DC subsets (scale bar: 1µm).

**Fig. 4. DC and pre-DC subset gene expression analysis.** (**A**) Microarray data from sorted DC and pre-DC subsets (shown in Fig. 3) were analyzed by 3D PCA using differentially-expressed genes (DEG). For each PCA dimension (principal component, PC), the variance explained by each component is indicated. (**B-D**) Heat maps of DEG between (**B**) early pre-DC/pDC, (**C**) early pre-DC/pre-cDC1/cDC1 and (**D**) early pre-DC/pre-cDC2/cDC2. (**E**) Expression profiles of

62 common genes identified from DEG analysis comparisons along the lineage progression from early pre-DC to mature cDC, for cDC1 and cDC2 respectively. The profiles were plotted with the log2 fold-change values (versus early pre-DC). (**F**) Expression level of CD327 (SIGLEC6), CD22 and AXL proteins by DC and pre-DC subsets evaluated by flow cytometry. The mean fluorescence intensities are indicated. (**G**) Expression profile of selected transcription factors.

**Fig. 5. Functional analysis of DC and pre-DC subsets.** (**A**) Frequency of cytokine production by pre-DC and DC subsets upon TLR stimulation was measured by intracellular flow cytometry. Dot plots (left panel) show IFNα, IL-12p40 and TNFα production by pDC, early pre-DC, pre-DC2, cDC2, pre-DC1 and cDC1. Bar charts (right panel) show the mean relative numbers of pre-DC and DC subset cells producing IFNα$^+$, IL-12p40$^+$ or TNFα$^+$ in response to LPS, LPS+IFNγ (L+I), pI:C, CL097 (CL) or CpG ODN2216 (CpG) (n=4). (**B**) Expression level (represented as mean fluorescence intensity (MFI)) of costimulatory molecules (CD40, CD80, CD83, CD86) by blood pre-DC and DC subsets (n=4). (**C**) Proliferation of naïve CD4$^+$ T cells after 6 days of culture with allogenic pre-DC and DC subsets (n=3). P-values calculated by Mann-Whitney test. Error bars represent mean +/- SEM.

**Fig. 6. Unsupervised mapping of DC ontogeny using CyTOF.** CyTOF data from bone marrow (BM) and PBMC were analyzed using isoMAP dimensionality reduction to compare overall phenotypic relatedness of cell populations, and were automatically subdivided into clusters using the phenograph algorithm. (**A, B**) IsoMAP1-2 plots showing the expression level of common DC progenitor (CDP), pDC, pre-DC and cDC -specific markers within (**A**) BM and (**B**) blood Lin(CD3/CD7/CD14/CD15/CD19/CD34)$^-$HLA-DR$^+$CD123$^+$ cells. (**C**) Phenotypic association between Lin-HLA-DR$^+$CD123$^{hi}$ BM and CD123$^+$ PBMC, showing progression from CDP towards pDC or pre-DC in the BM, and the clear separation of pDC and pre-DC in the blood. Cells within the pre-DC phenograph clusters (clusters #1 and #2 in the BM, and #6 in the blood)

and cells within the pDC phenograph clusters (clusters #3 and #4 in the BM, and #7 in the blood) were further analyzed by isoMAP to define pre-DC subsets (left panels, and fig. S20, c and d) and heterogeneity among pDC (right panels, and fig. S20, d and e).

Accepted Manuscript: See et al., Science 2017, Vol. 356, Issue 6342, eaag3009, doi:10.1126/science.aag3009

Figure 1

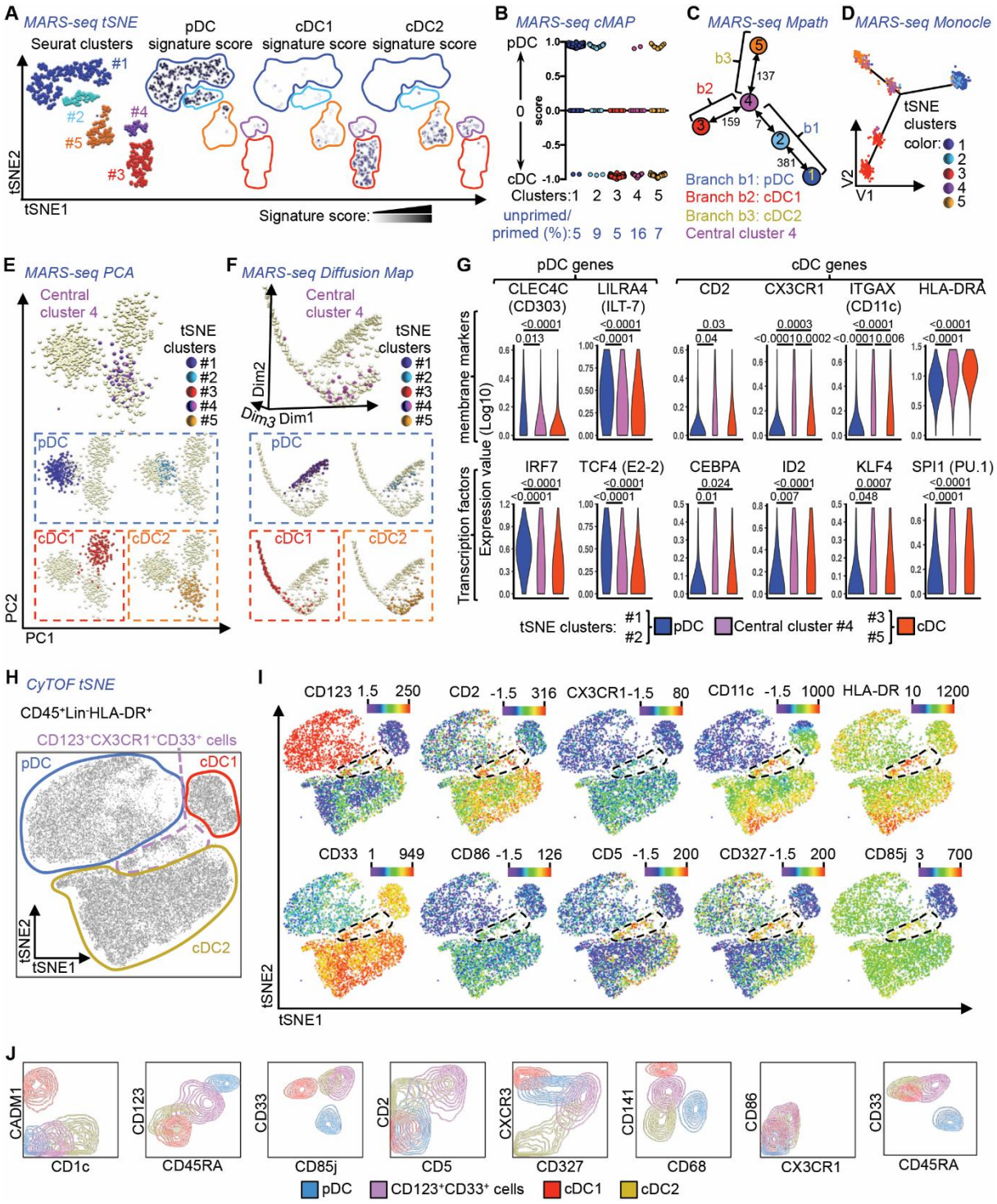


Figure 2

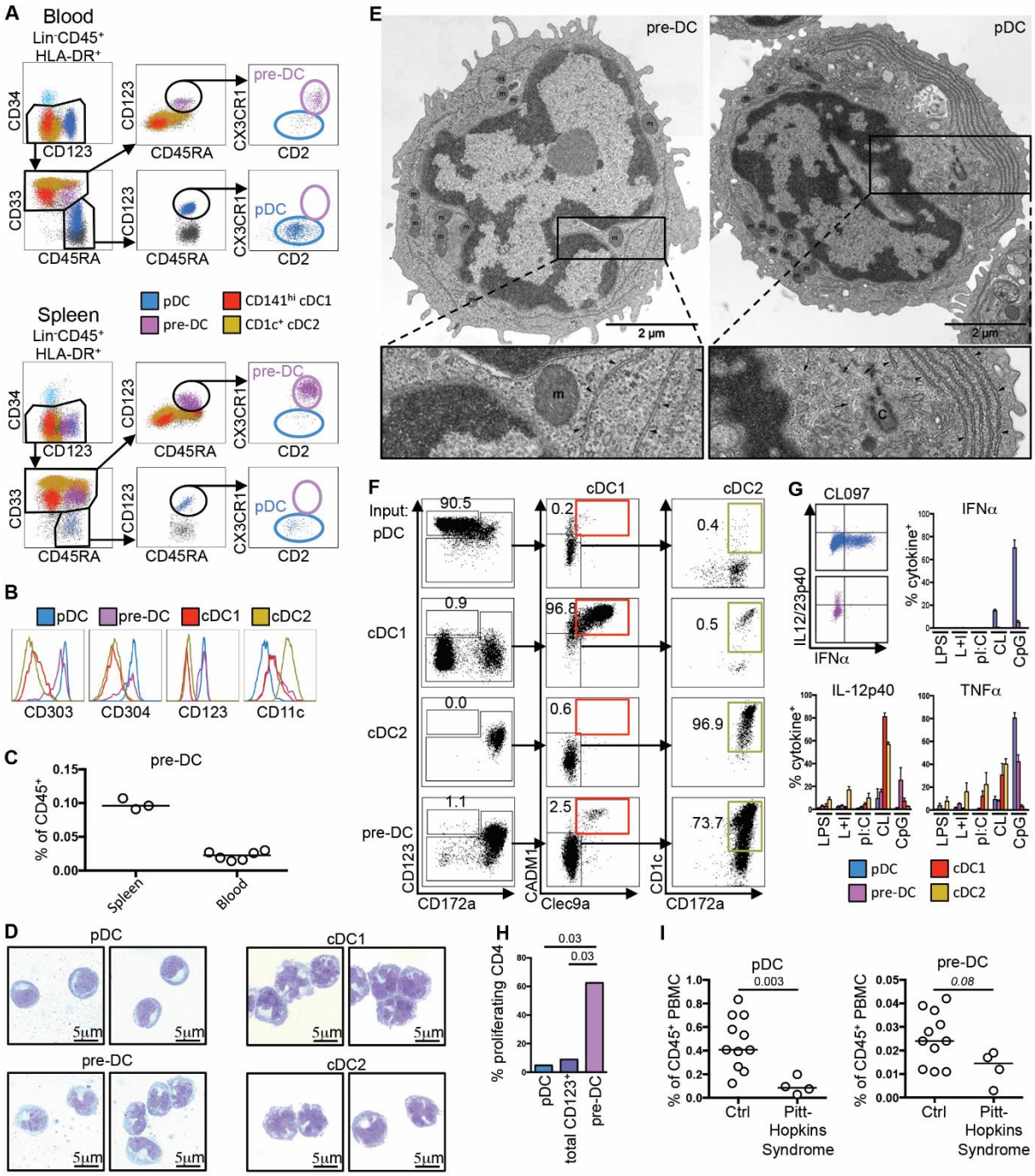

A
Blood
Lin-CD45+
HLA-DR+
CD34
CD123
CD45RA
CD123
CX3CR1
CD2
pre-DC
CD33
CD123
CD45RA
CD123
CX3CR1
CD2
pDC
Spleen
Lin-CD45+
HLA-DR+
pDC
pre-DC
CD141hi cDC1
CD1c+ cDC2
B
pDC
pre-DC
cDC1
cDC2
CD303
CD304
CD123
CD11c
C
pre-DC
% of CD45+
Spleen
Blood
D
pDC
cDC1
pre-DC
cDC2
5µm
E
pre-DC
pDC
2 µm
m
c
F
Input:
pDC
cDC1
cDC2
90.5
0.2
0.4
0.9
96.8
0.5
0.0
0.6
96.9
1.1
2.5
73.7
CD123
CD172a
CADM1
Clec9a
CD1c
CD172a
G
CL097
IL12/23p40
IFNα
IFNα
% cytokine+
LPS
L+I
pI:C
CL
CpG
IL-12p40
TNFα
pDC
cDC1
pre-DC
cDC2
H
% proliferating CD4
0.03
0.03
pDC
pre-DC
total CD123+
I
pDC
0.003
pre-DC
0.08
% of CD45+ PBMC
Ctrl
Pitt-Hopkins Syndrome

Figure 3

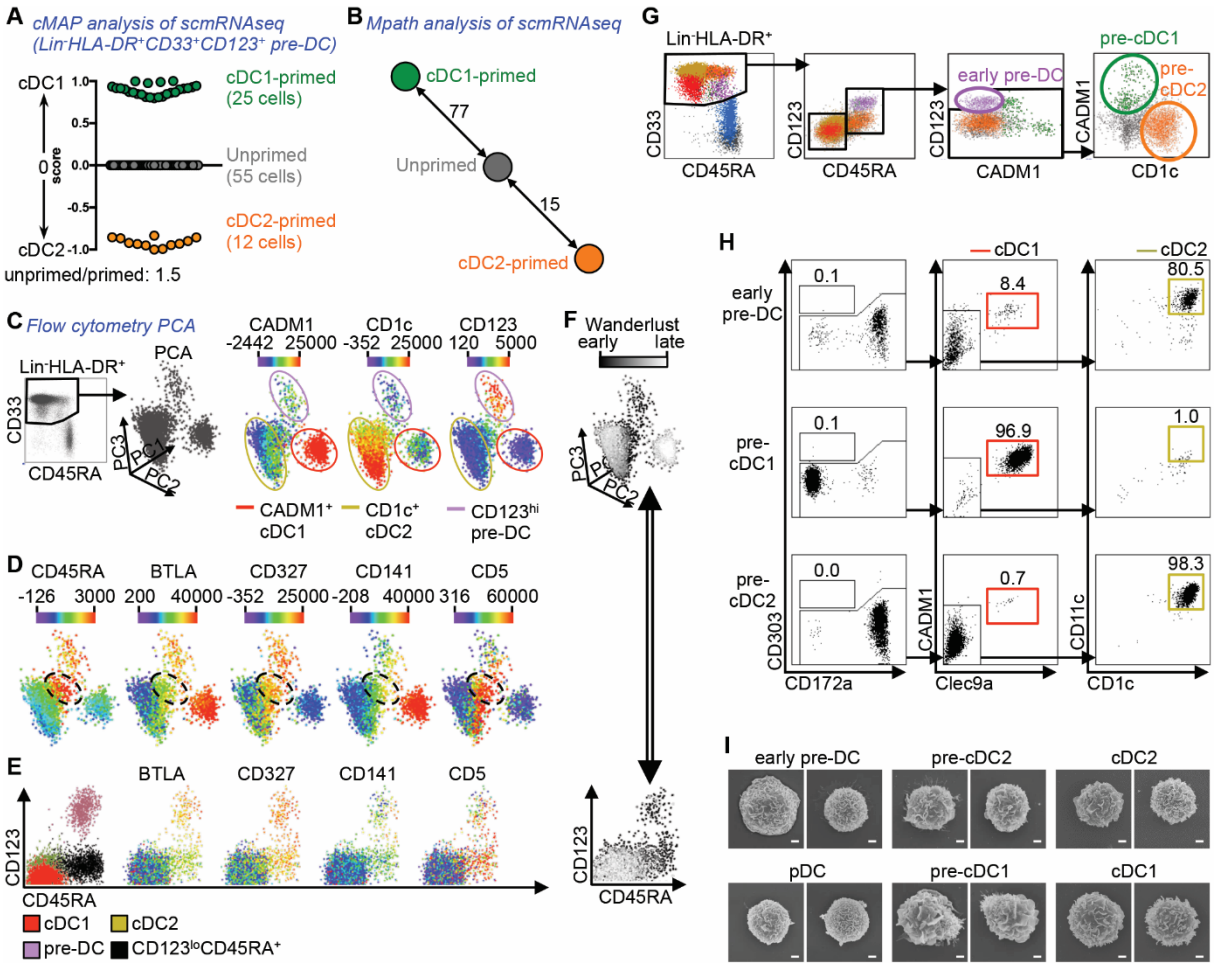

A cMAP analysis of scmRNAseq (Lin-HLA-DR+CD33+CD123+ pre-DC)
cDC1
cDC2
score
cDC1-primed (25 cells)
Unprimed (55 cells)
cDC2-primed (12 cells)
unprimed/primed: 1.5
B Mpath analysis of scmRNAseq
cDC1-primed
77
Unprimed
15
cDC2-primed
G
Lin-HLA-DR+
CD33
CD45RA
CD123
early pre-DC
CADM1
pre-cDC1
pre-cDC2
CD1c
C Flow cytometry PCA
PCA
PC1
PC2
PC3
CADM1 -2442 25000
CD1c -352 25000
CD123 120 5000
CADM1+ cDC1
CD1c+ cDC2
CD123hi pre-DC
F Wanderlust
early
late
D
CD45RA -126 3000
BTLA 200 40000
CD327 -352 25000
CD141 -208 40000
CD5 316 60000
E
BTLA
CD327
CD141
CD5
CD123
CD45RA
cDC1
cDC2
pre-DC
CD123loCD45RA+
H
cDC1
cDC2
early pre-DC
pre-cDC1
pre-cDC2
0.1
8.4
80.5
0.1
96.9
1.0
0.0
0.7
98.3
CD303
CD172a
CADM1
Clec9a
CD11c
CD1c
I
early pre-DC
pre-cDC2
cDC2
pDC
pre-cDC1
cDC1

Figure 4

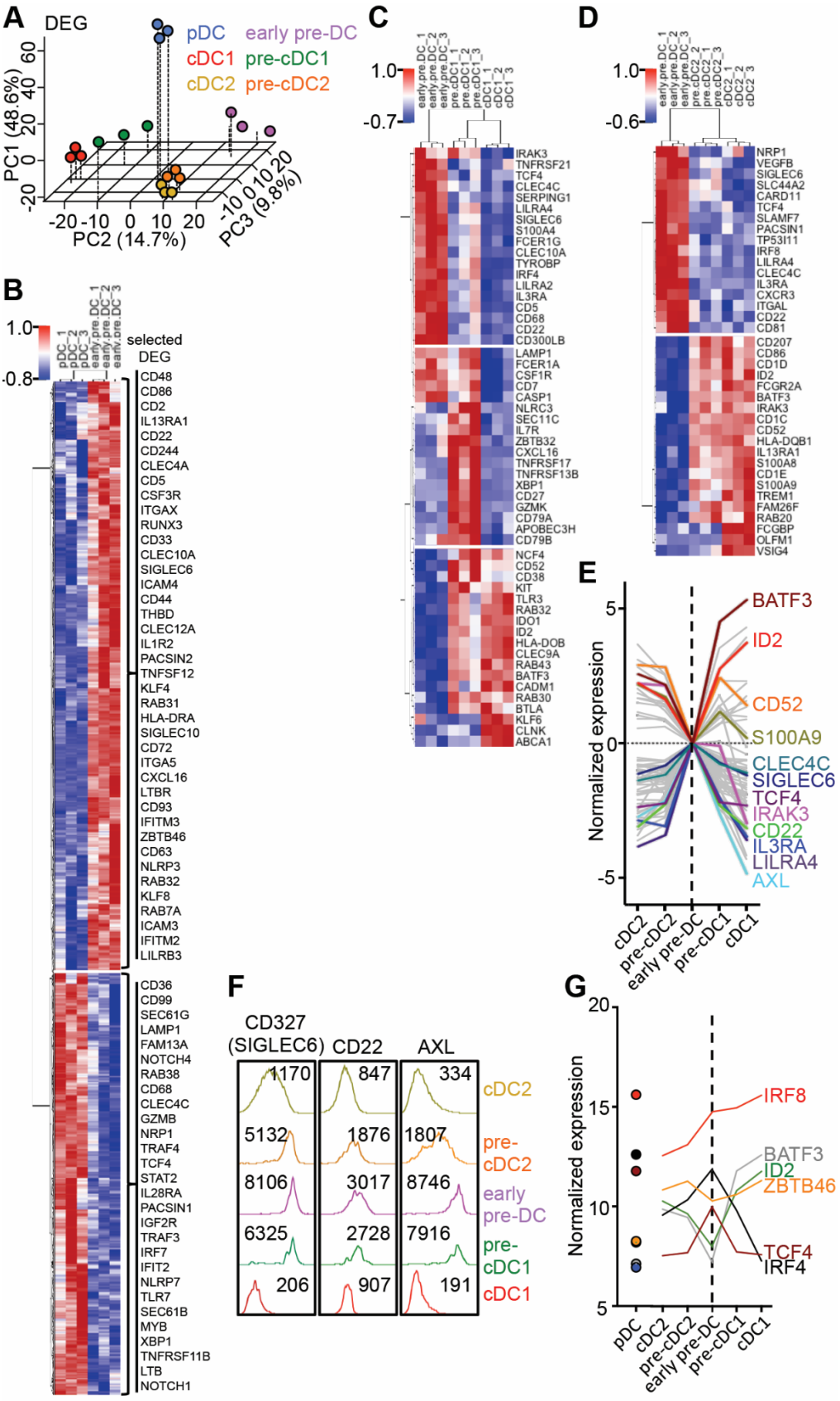

A
DEG
pDC
early pre-DC
cDC1
pre-cDC1
cDC2
pre-cDC2
PC1 (48.6%)
PC2 (14.7%)
PC3 (9.8%)
B
selected DEG
C
D
E
Normalized expression
BATF3
ID2
CD52
S100A9
CLEC4C
SIGLEC6
TCF4
IRAK3
CD22
IL3RA
LILRA4
AXL
cDC2
pre-cDC2
early pre-DC
pre-cDC1
cDC1
F
CD327 (SIGLEC6)
CD22
AXL
1170
847
334
5132
1876
1807
8106
3017
8746
6325
2728
7916
206
907
191
G
Normalized expression
IRF8
BATF3
ID2
ZBTB46
TCF4
IRF4
pDC

Figure 5

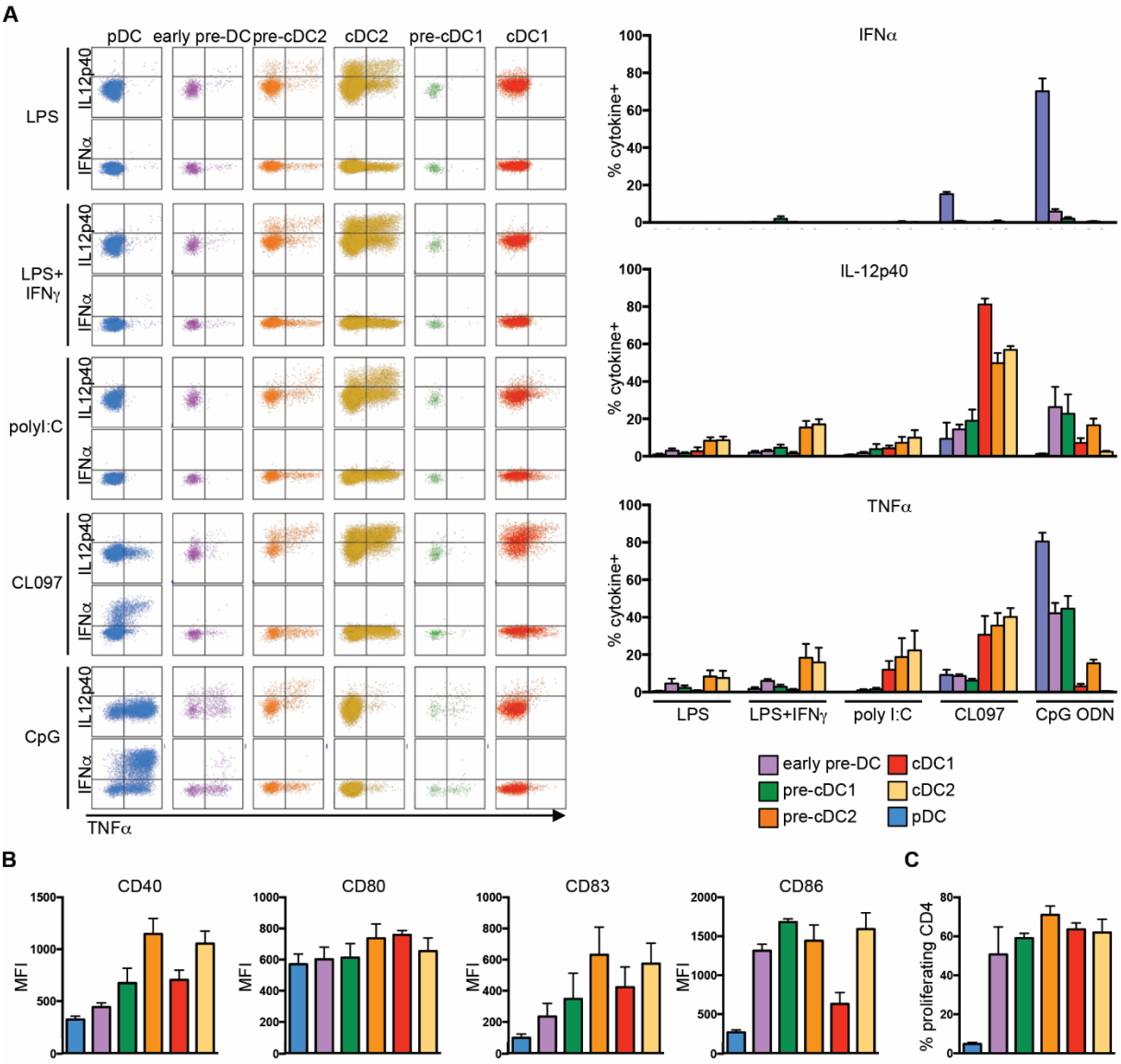
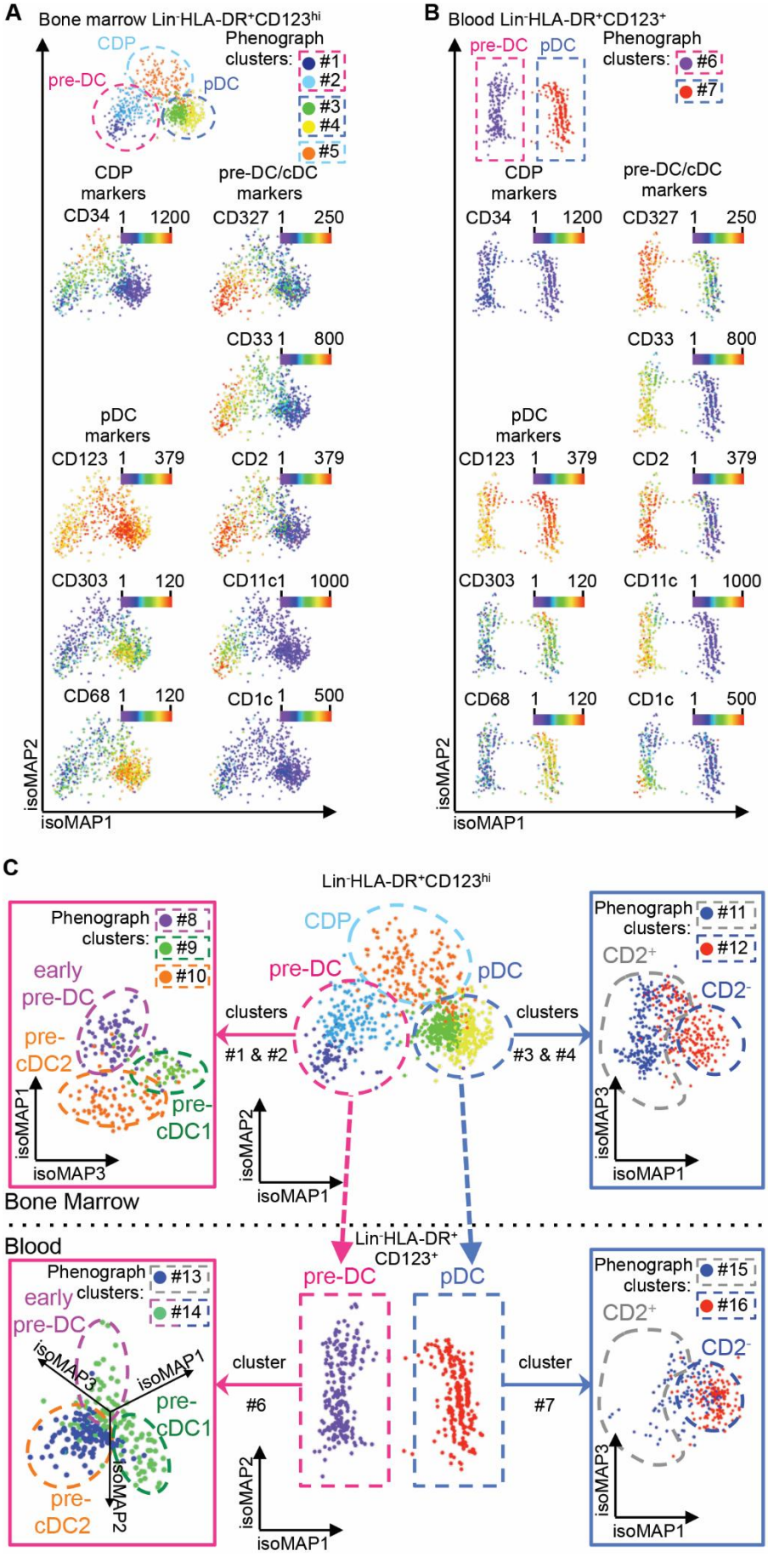

A
pDC
early pre-DC
pre-cDC2
cDC2
pre-cDC1
cDC1
LPS
LPS+ IFNγ
polyI:C
CL097
CpG
IL12p40
IFNα
TNFα
IFNα
IL-12p40
TNFα
% cytokine+
LPS
LPS+IFNγ
poly I:C
CL097
CpG ODN
early pre-DC
cDC1
pre-cDC1
cDC2
pre-cDC2
pDC
B
CD40
CD80
CD83
CD86
MFI
C
% proliferating CD4

Figure 6


A
Bone marrow Lin-HLA-DR+CD123hi
CDP
pre-DC
pDC
Phenograph clusters:
#1
#2
#3
#4
#5
CDP markers
pre-DC/cDC markers
CD34 1 1200
CD327 1 250
CD33 1 800
pDC markers
CD123 1 379
CD2 1 379
CD303 1 120
CD11c1 1000
CD68 1 120
CD1c 1 500
isoMAP2
isoMAP1
B
Blood Lin-HLA-DR+CD123+
pre-DC
pDC
Phenograph clusters:
#6
#7
CDP markers
pre-DC/cDC markers
CD34 1 1200
CD327 1 250
CD33 1 800
pDC markers
CD123 1 379
CD2 1 379
CD303 1 120
CD11c 1 1000
CD68 1 120
CD1c 1 500
isoMAP2
isoMAP1
C
Lin-HLA-DR+CD123hi
CDP
pre-DC
pDC
clusters #1 & #2
clusters #3 & #4
Phenograph clusters:
#8
#9
#10
early pre-DC
pre-cDC2
pre-cDC1
isoMAP1
isoMAP3
isoMAP2
isoMAP1
Phenograph clusters:
#11
#12
CD2+
CD2-
isoMAP3
isoMAP1
Bone Marrow
Blood
Lin-HLA-DR+CD123+
pre-DC
pDC
cluster #6
cluster #7
Phenograph clusters:
#13
#14
early pre-DC
pre-cDC1
pre-cDC2
isoMAP3
isoMAP1
isoMAP2
isoMAP2
isoMAP1
Phenograph clusters:
#15
#16
CD2+
CD2-
isoMAP3
isoMAP1

# Supplementary Materials for

## Mapping the human DC lineage through the integration of high dimensional techniques

Peter See[#], Charles-Antoine Dutertre[#], Jinmiao Chen[#], Patrick Günther, Naomi McGovern, Sergio Erdal Irac, Merry Gunawan, Marc Beyer, Kristian Händler, Kaibo Duan, Hermi Rizal Bin Sumatoh, Nicolas Ruffin, Mabel Jouve, Ester Gea-Mallorquí, Raoul C.M. Hennekam, Tony Lim, Chan Chung Yip, Ming Wen, Benoit Malleret, Ivy Low, Nurhidaya Binte Shadan, Charlene Foong Shu Fen, Alicia Tay, Josephine Lum, Francesca Zolezzi, Anis Larbi, Michael Poidinger, Jerry K.Y. Chan, Qingfeng Chen, Laurent Renia, Muzlifah Haniffa, Philippe Benaroch, Andreas Schlitzer, Joachim L. Schultze, Evan W. Newell, and Florent Ginhoux*

#These authors contributed equally to this work.
*Correspondence to: Florent_Ginhoux(

**This PDF file includes:**

Supplementary Materials and Methods
Figs. S1 to S22
Captions for Tables S1 to S6

**Other supplementary material for this manuscript includes the following:**

Tables S1 to S6 (Excel format)

## Supplementary Materials and Methods

Validation of down sampling threshold for normalization of MARS-seq single cell transcriptome data

High variance in terms of quality of single-cell transcriptomes is expected in a single-cell RNA sequencing experiment due to the low quantity of RNA input material. This caveat necessitates stringent quality control in order to avoid a bias introduced by low quality single-cell transcriptomes. In single-cell transcriptomics it is, therefore, common practice to remove low quality transcriptomes to ensure an unbiased and biologically meaningful analysis (*71*, *72*). Different strategies have been used to filter out low quality cells, including an empirically determined cutoff for cell filtering (*71*), a down sampling strategy to normalize and filter low quality cells (*15*), and various filtering cutoffs from 600 UMIs/cell or 400 UMIs/cells (*15*), <500 molecule counts per cell (*73*) and<200 UMIs/cell (*74*). To the best of our knowledge, a mathematically determined cut-off was not reported in any of these studies. As these previous studies were performed on murine cells, and quality filters in single-cell data have to be established within the respective dataset, we had to adapt the filtering strategy to human cells. To determine the quality threshold for our dataset, several diagnostics were used to estimate the optimal cutoff for down sampling of molecule counts. Firstly, we visualized the cumulative distribution of molecule counts, where cells on the x-axis were ordered by decreasing UMI count (**fig. S1c**). Here, we noted that in a certain region there was a period of strong decline in the number of molecule counts per cell. This region corresponded to a range of molecule counts between 400 and 1200 UMIs per cell. The next metric used to judge an objective threshold (**fig. S1d**) was the molecule count distribution of all cells. We found that many of the cell barcodes had <650 molecule counts: these cell barcodes most likely represented the background signal of our MARS-seq data set. The number of cell barcodes with a certain number of molecules decreases with increasing molecule count per cell; through this visualization, we identified natural breakpoints in the distribution that could be used as an objective threshold for filtering and normalization, as these breakpoints mark a change in the data structure and quality, and indicate the transition from background to signal, or from low-quality transcriptomes to high-quality transcriptomes. Here, three notable points were identified (**fig. S1d**), which corresponded to molecule counts of 650 (blue), 1,050 (red) and 1,700 (green) per cell. To objectively determine which of these points represented a shift in data quality from low to high quality transcriptomes, we asked where in the graph a turning point could be identified (**fig. S1d**). In the density plot (**fig. S1d,** top panel), the three lines (blue, red, green) are the breakpoints where the slope of the density function (1$^{st}$ derivative of density, **fig. S1d**, middle panel) has a sudden change. On the blue line, the downward slope (1st derivative) changes from being very steep to less steep, so that the 2$^{nd}$ derivative is the highest at this point. Similarly, on the red line, the downward slope changes from less steep to more steep, so the 2$^{nd}$ derivative is the lowest. Based on these observations, the three turning points were identified by the 2$^{nd}$ derivative (**fig. S1d**, bottom panel). When a cutoff of 650 was applied, the number of molecule counts per cell was too low and the three DC populations – plasmacytoid DC (pDC) and conventional DC (cDC) subsets cDC1 and cDC2, could not be distinguished by principal component

analysis (PCA; **fig. S1e**). When a cutoff of 1,700 was applied, the number of cells retained was too low. Therefore, the 1,050 cutoff was an optimal tradeoff between the number of cells analyzed (cells retained after filtering by down sampling normalization) and the number of molecule counts in a cell (gene expression information that remains after discarding molecule counts by down sampling).

To ensure data reproducibility, stability and independence of the chosen molecule cutoff, we simulated our initial analyses using cutoffs of 650, 1,050, 1,700 and 2,350 molecule counts (**fig. S1e**). All four chosen simulation values exhibited the same general data topology if the data were dimensionally-reduced using PCA, thus proving that the biological data structure was robust and independent of filtering thresholds. In addition, we correlated the influence of the filtering threshold on the gene loadings within the first two principal components. Principal component 1 (PC1) of the dataset down-sampled to 1,050 molecule counts was highly correlated with PC1 of the datasets down-sampled to either 650 or 1,700 molecule counts (Pearson = 0.996 and 0.999, respectively). The same was true for PC2 (Pearson = 0.960 and 0.925, respectively). These results indicated that the chosen filtering cutoff of 1,050 was representative and objectively-derived.

The MARS-seq data obtained in this study were generated by two independent experiments (run1 and run2), which were combined for further data analysis. After normalization, the correlation between the average molecule count of all genes in run1 vs run2 was assessed (**fig. S1f**, which shows the high correlation between the average molecular counts in both runs (r=0.994)). When assessing for a batch effect, it is important to ensure that runs do not determine the clustering itself. We therefore plotted the t-distributed stochastic neighbor embedding (tSNE) values (**fig. S1g**) (cells of run1 and run2 in equal proportions) together with their density estimates. This analysis showed that the general distribution and, therefore, the clustering was not governed by the run, which is in line with the observation that our clustering identified biologically reasonable groups that clearly corresponded to the three DC populations (pDC, cDC1 and cDC2) (**Fig. 1A**). Consequently, the observed clusters were not explained by the variance between the runs, but by biology.

We next compared the frequencies of cell types, as determined by the clustering, within the two runs (**fig. S1h**). This showed that the ratio between the cells in different clusters was comparable between the two runs. Of note, the ratio does not need to be identical in both runs (*73*). In addition, this analysis showed that no cluster dominated a single run. Due to the fact that we are taking relatively small samples from a large total population, the frequencies of cell types are expected to show natural variation between runs, which could explain slight shifts in cellular frequencies.

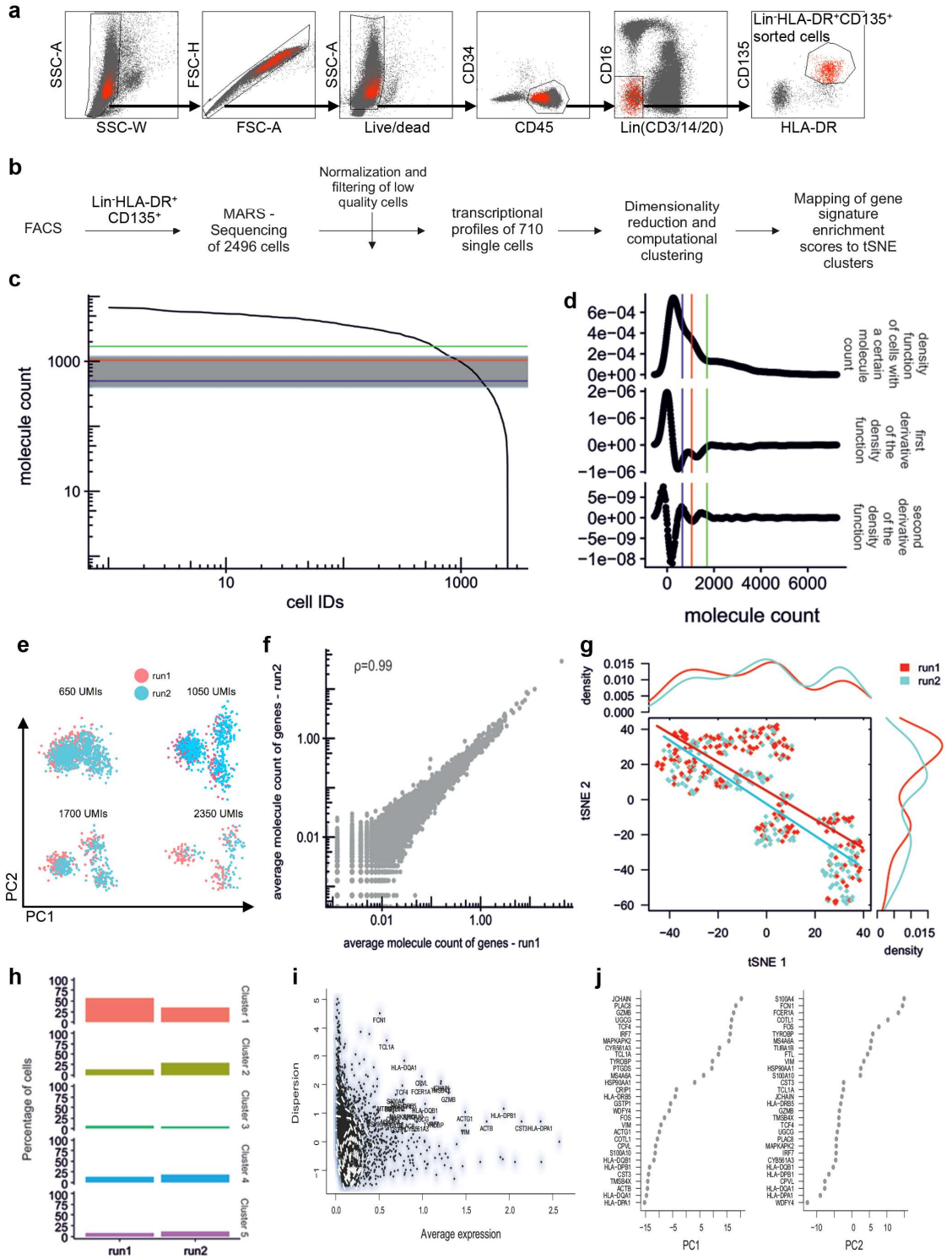



**Fig. S1**. **a**. Gating strategy for FACS of single cells from total Lin$^{-}$HLA-DR$^{+}$CD135$^{+}$ cells. **b**. Workflow of the MARS-seq single cell data analysis. **c**. Association between molecule counts and cells. Cell IDs were sorted from highest to lowest number of unique

molecular identifier (UMI) or molecule counts. The data are presented on a log10 axis. The three colored lines correspond to molecule counts of 650 (blue), 1,050 (red) and 1,700 (green) per cell. The grey area indicates the range of molecule counts from 400 to 1,200 UMIs per cell. Cells with <1,050 molecules were removed from the analysis (n=1,786 cells). A total of 710 high-quality cells were used for further downstream analyses. **d.** Density plot (top panel) representing the distribution of cells with a certain number of molecules, and the first (middle panel) and second derivative (bottom panel) of the density function. The three colored lines correspond to molecule counts of 650 (blue), 1,050 (red) and 1,700 (green) per cell. **e.** Principal component analysis (PCA) after simulation at different normalization thresholds. Points were colored according to the different runs. **f.** Correlation plot of average expression of genes in run2 (y-axis) versus average expression of genes in run1 (x-axis). The data are presented on a log10 axis. The Pearson correlation coefficient was 0.99. **g.** t-distributed stochastic neighbor embedding (tSNE) analysis of the 710 single cells, colored by run association, showed an even distribution of the cells within the tSNE plot. Lines represent a linear fit of the points. The distributions of the points along the tSNE component 1 and component 2 were represented as density plots on the top or right panel, respectively. **h.** Frequency of cells in the five determined clusters for run1 and run2. **i.** Mean-variability plot showed average expression and dispersion for each gene. This analysis was used to determine highly variable gene expression (labeled by gene symbol). The 36 highly variable genes were used to perform a dimensionality reduction of the single-cell data by PCA. **j.** The highest gene loadings in the first and second principal component (PC1 and PC2) from the PCA of 710 high quality cells are shown.

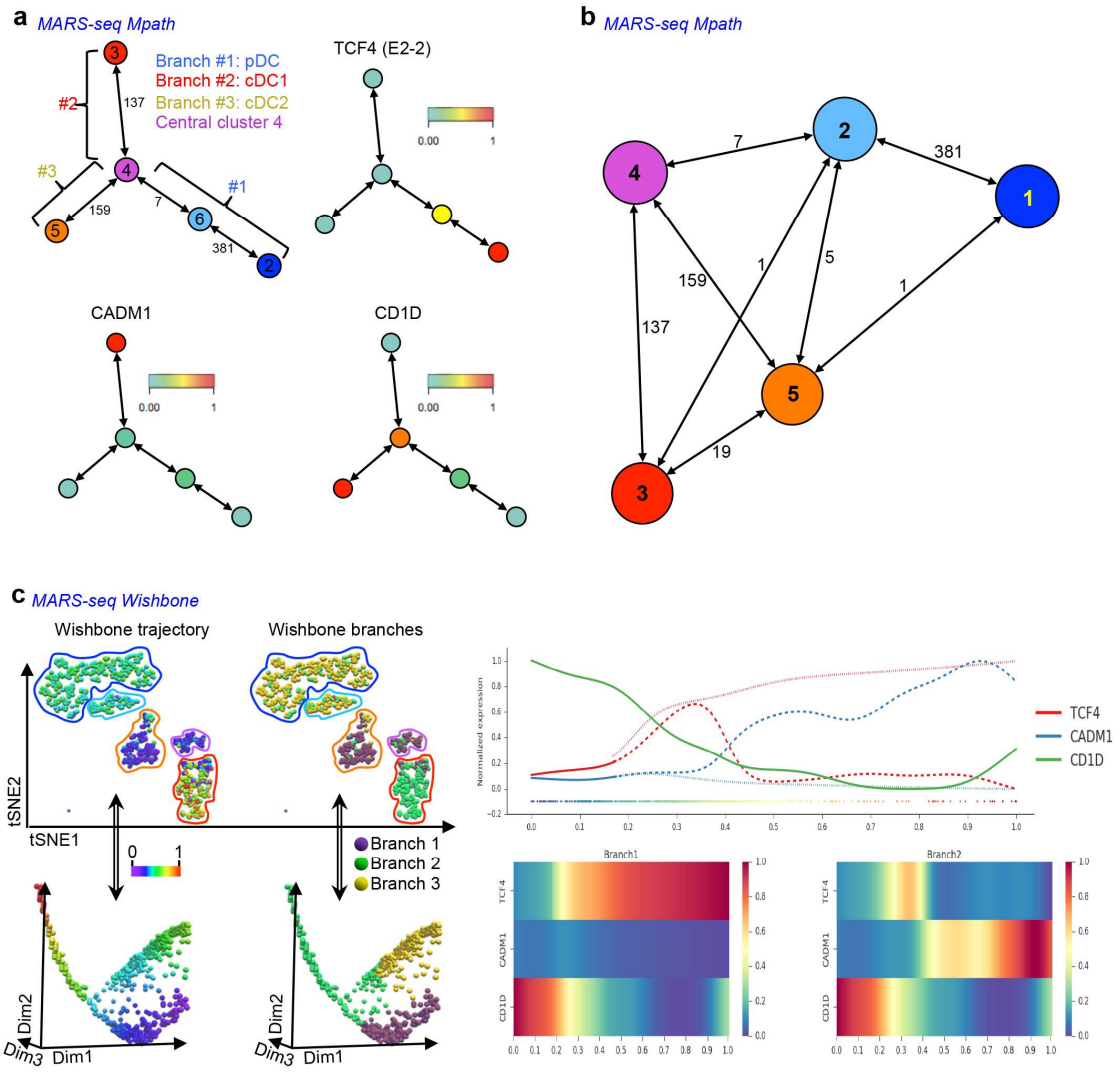



**Fig. S2. a.** Relative expression of signature genes of pDC (TCF4), cDC1 (CADM1) and cDC2 (CD1D) in Mpath clusters defined in Fig. 1C. **b.** Weighted neighborhood network of the Mpath analysis shown in Fig. 1C. **c.** Analysis of MARS-seq data using the Wishbone algorithm. In the 2D-t-distributed stochastic neighbor embedding (tSNE) plot (upper panels) and in the 3D-Diffusion Map (lower panels) (See Fig. 1, A and F, respectively), cells were colored according to the values of the Wishbone trajectory (left panels) or the values of the Wishbone branches (right panels). Line chart (top right panel) of expression of signature genes along Wishbone trajectory. X-axis represents pseudo-

time of Wishbone trajectory. Solid line represents backbone trajectory, dotted lines represent separate trajectories along the two branches. Heat maps (bottom right panels) of expression of signature genes along Wishbone trajectory on the two branches.

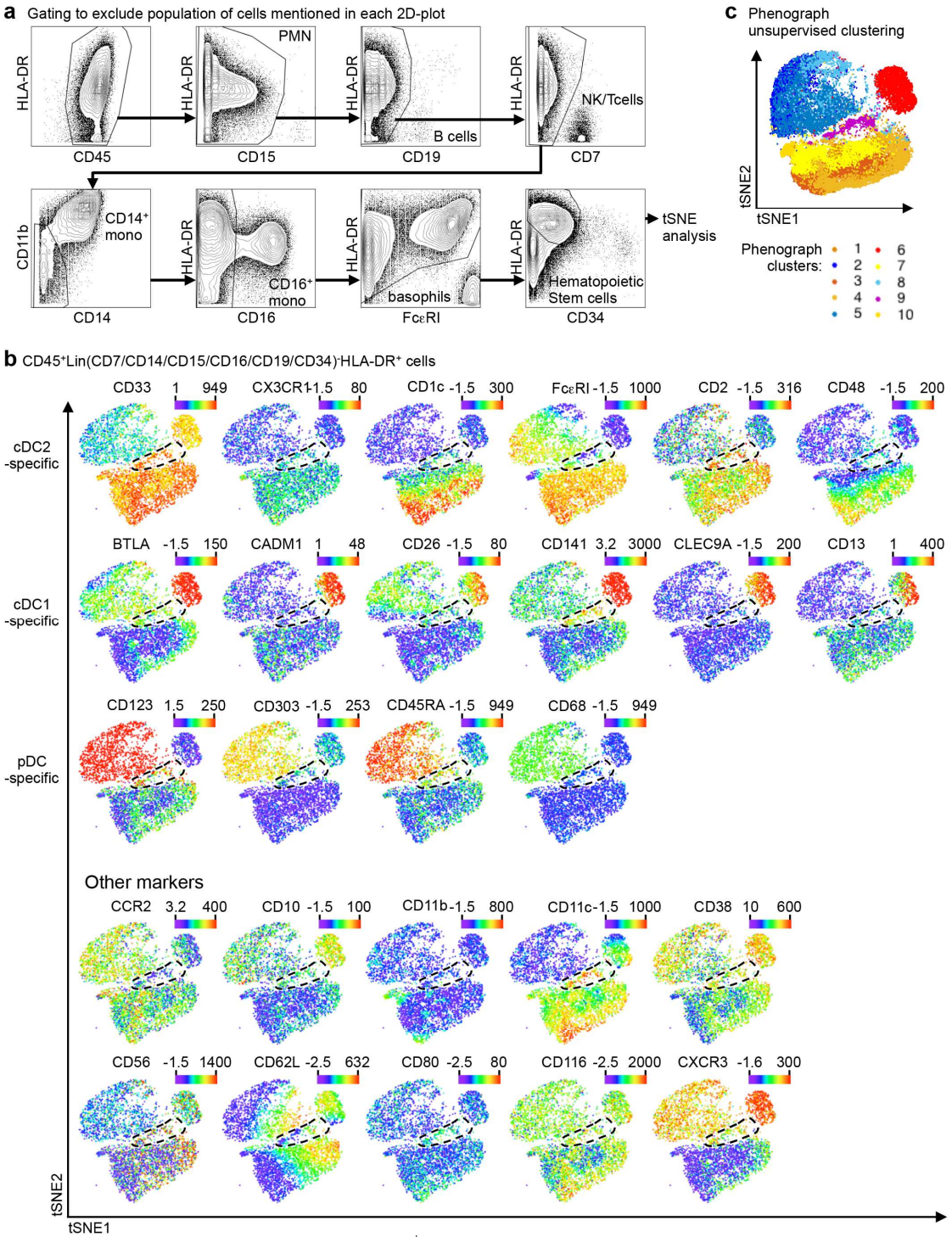



**Fig. S3**. **a.** Gating strategy of CD45$^{+}$Lin(CD7/CD14/CD15/CD16/CD19/CD34)$^{-}$HLA-DR$^{+}$ blood mononuclear cells from CyTOF analysis for downstream t-distributed stochastic neighbor embedding (tSNE) as shown in Fig. 1, E to G. The name of the

excluded population(s) is indicated in each corresponding 2D-plot. **b.** tSNE plots of the CyTOF data from Fig. 1, H to J showing the expression level of cDC2-, cDC1- and pDC-specific markers. **c.** Unsupervised phenograph clustering identified 10 clusters that were overlaid onto the tSNE1/2 plot of the CyTOF data from Fig. 1, H and I.

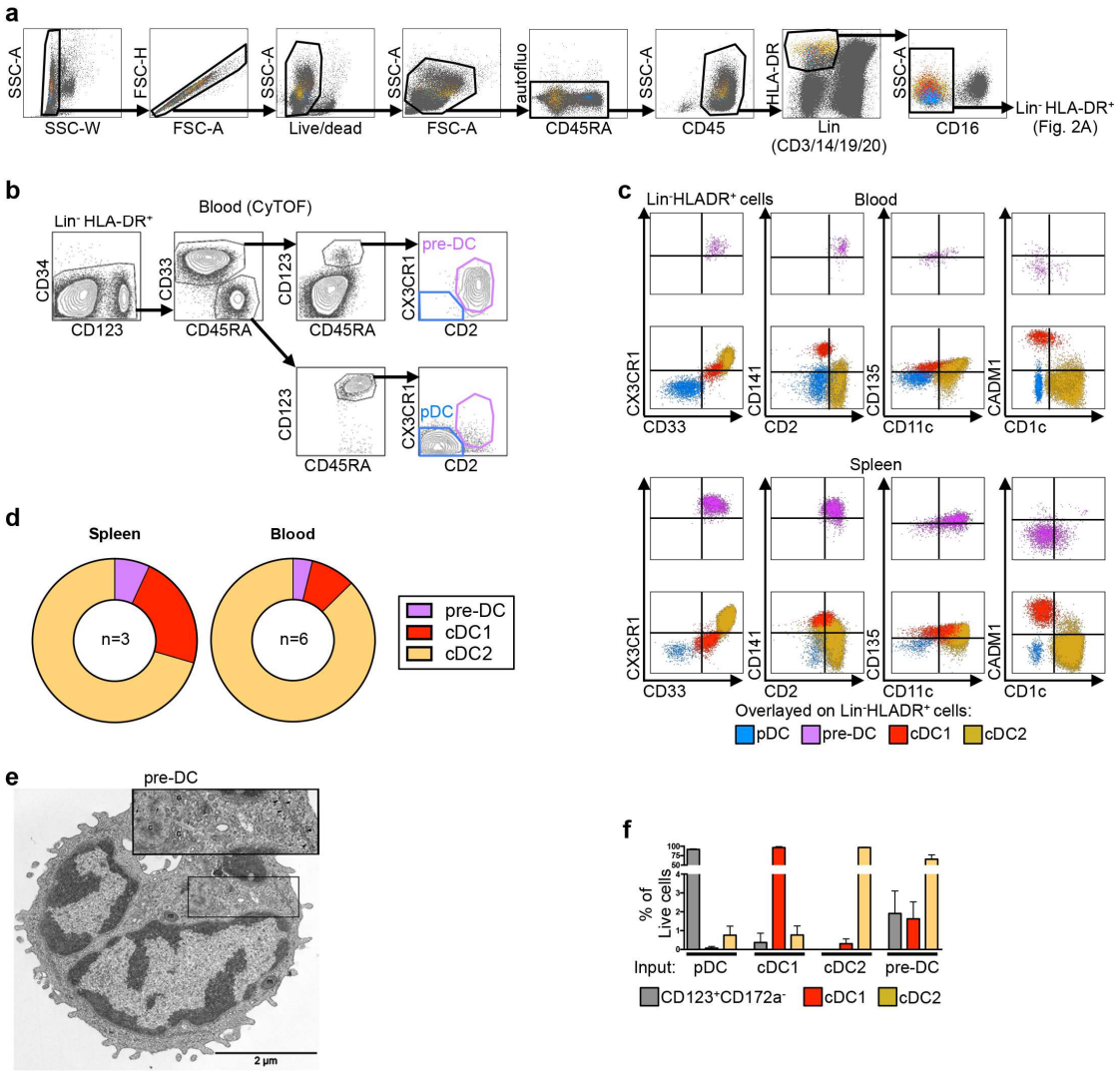



**Fig. S4**. **a.** Gating of flow cytometry data to identify the Lin$^-$HLA-DR$^+$ cell population displayed in Fig. 2A (blood data displayed). **b.** Classical contour plots of CyTOF data from Fig. 1 showing the same gating strategy as applied in the flow cytometry analyses shown in Fig. 2A. **c.** Flow cytometry data showing the relative expression of CD33, CX3CR1, CD2, CD141, CD11c, CD135, CD1c and CADM1 by pre-DC, pDC, cDC1 and cDC2 defined in Fig. 2A in the blood (upper panels) and spleen (lower panels). **d.** Ring graphical representation of the proportion of pre-DC, cDC1 and cDC2 among total Lin$^-$

$CD34^{-}HLA\text{-}DR^{+}CD33^{+}$ cDC defined in Fig. 2A in the spleen (left) and blood (right). **e.** Representative electron micrographs showing morphological characteristics of a pre-DC. **f.** Histograms of the mean relative numbers of $CD123^{+}CD172\alpha^{-}$ cells, $Clec9A^{+}CADM1^{+}$ cDC1 or $CD172\alpha^{+}CD1c^{+}$ cDC2 from the *in vitro* differentiation assays as described in Fig. 2F (n=4). Error bars represent mean ± SEM.

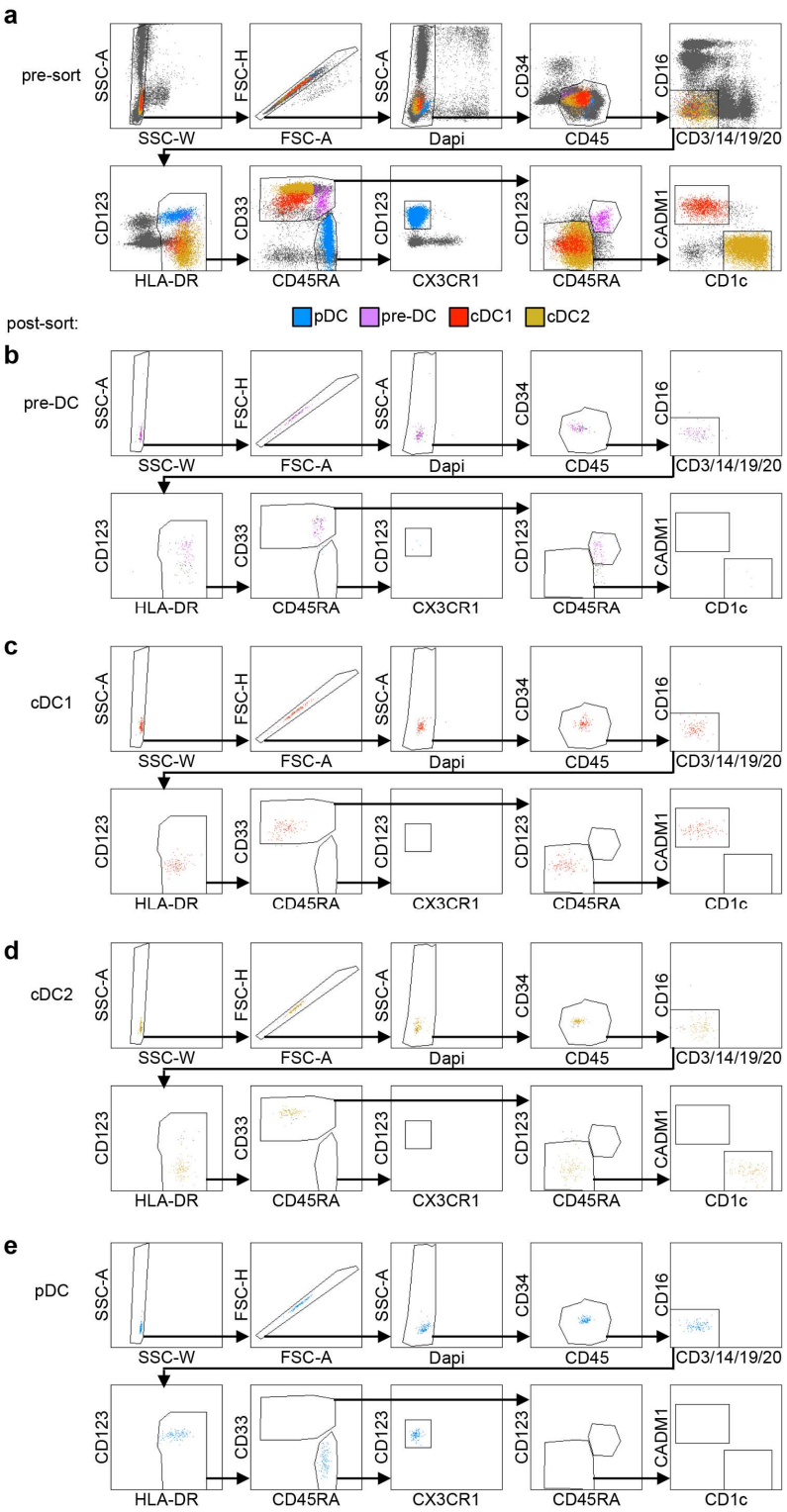



**Fig. S5**. Gating strategy for the fluorescence-activated cell sorting of DC subsets and pre-DC used in the *in vitro* differentiation assays (Fig. 2F). **a.** Pre-sorted data and **b-d.** post-sorted re-analysis of **b.** pre-DC, **c.** cDC1, **d.** cDC2, and **e.** pDC.

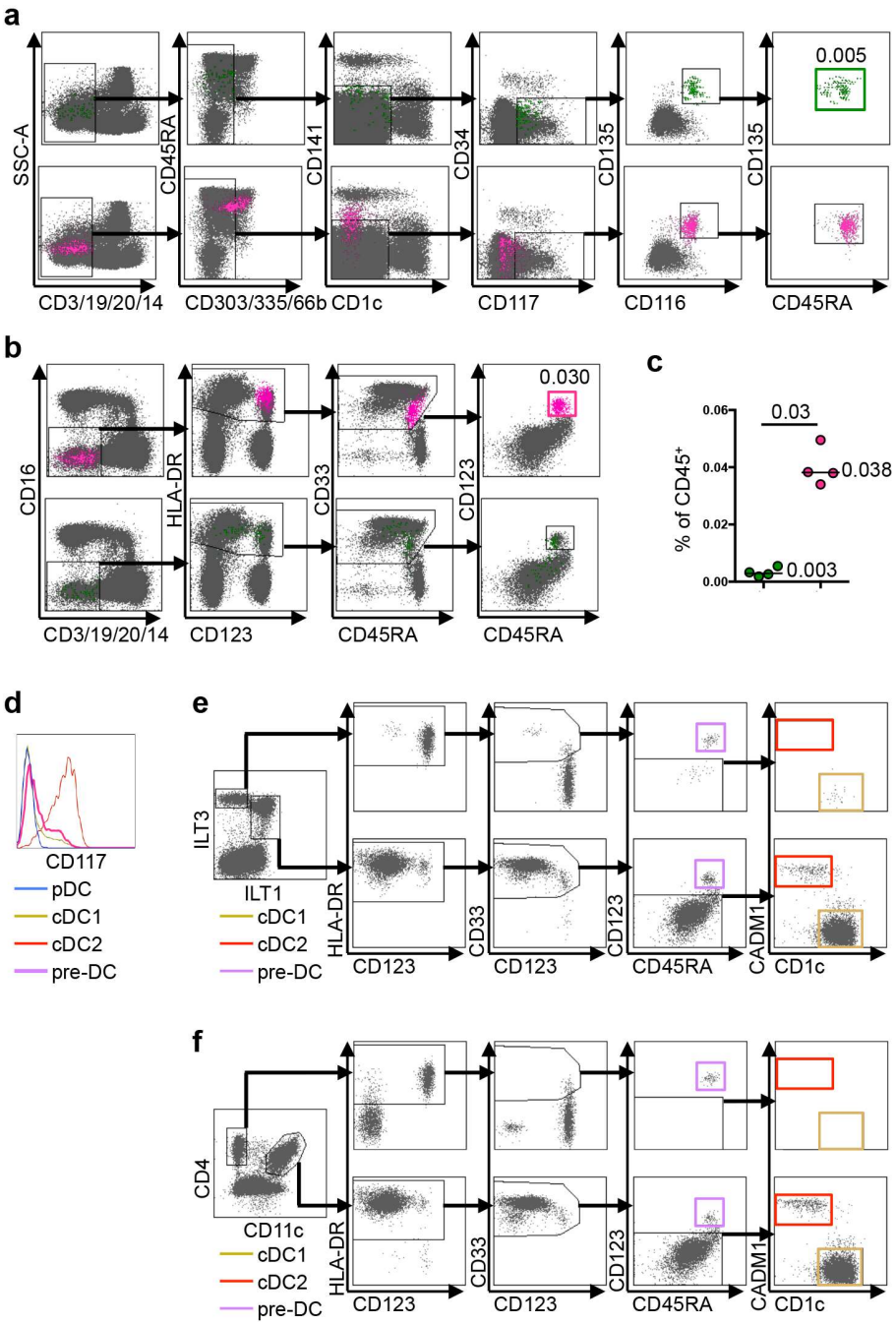



**Fig. S6**. **a-c.** Comparison of **a.** the gating strategy from Breton *et al*. (*9*) pre-DC are shown in green) and **b.** the gating strategy used in Fig. 2A and fig. S4a (pre-DC displayed in purple) to define pre-DC. The relative numbers of pre-DC defined using the two gating strategies among live $CD45^+$ peripheral blood mononuclear cells are indicated in the dot

plots. **c.** Graphical representation of the median relative numbers of pre-DC defined using the two gating strategies among live $CD45^+$ blood mononuclear cells (n=4). The median percentages of $CD45^+$ values are indicated. P-values were calculated using the Mann-Whitney test. **d.** Histogram showing the expression of CD117 by DC subsets and pre-DC determined by flow cytometry. **e-f.** Identification of pre-DC (purple gate), cDC1 (red gate) and cDC2 (beige gate) among $Lin^-HLA\text{-}DR^+$ **e.** $ILT3^+ILT1^-$ cells (*10*) or $ILT3^+ILT1^+$ (cDC), and **f.** $CD4^+CD11c^-$ cells (*11*) or $CD4^{int}CD11c^+$ cDC. .

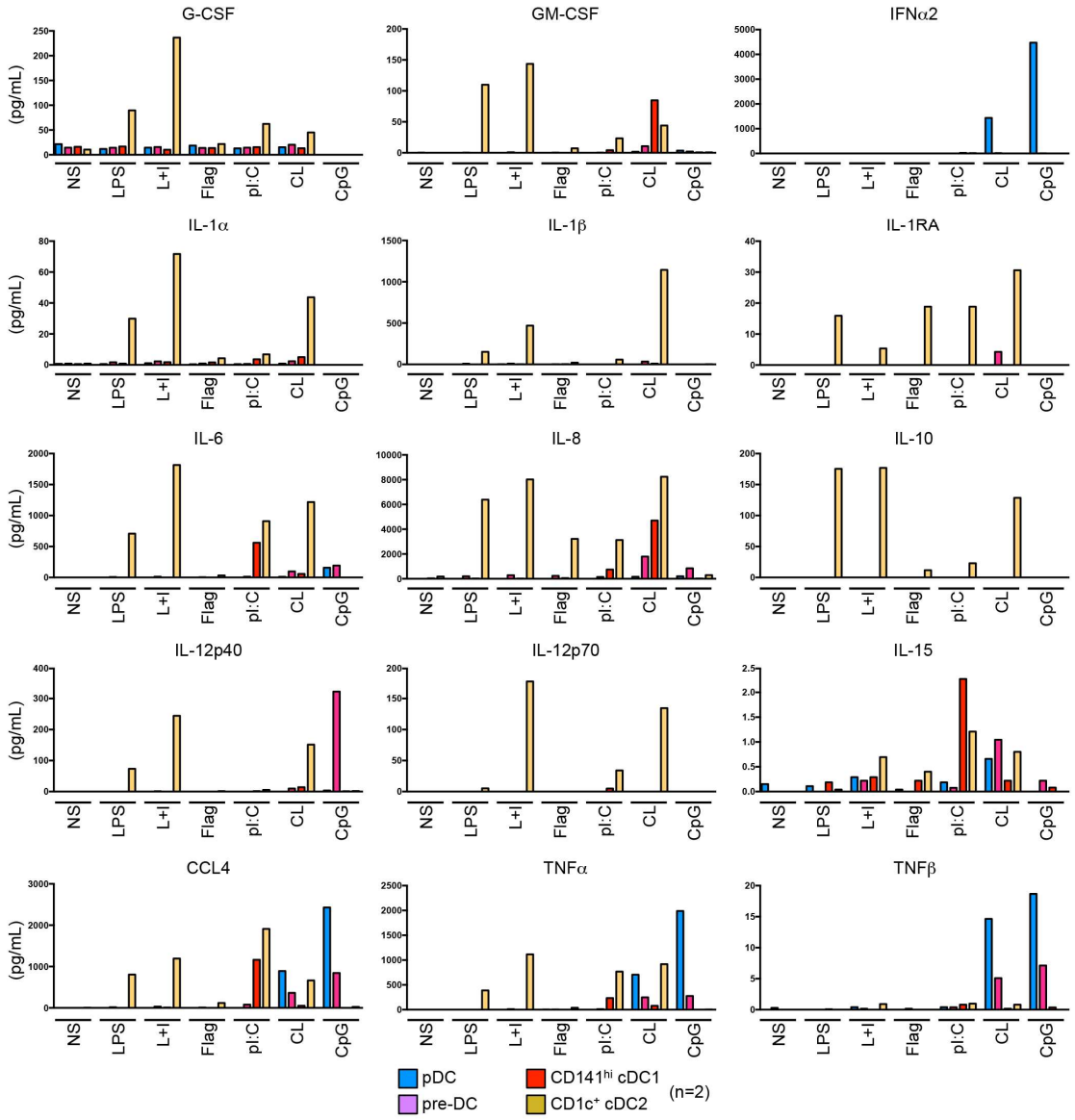



**Fig. S7**. pDC, pre-DC, cDC1 and cDC2 isolated by fluorescence-activated cell sorting were stimulated *in vitro* with LPS, LPS+IFNγ (L+I), Flagellin (Flag), polyI:C (pI:C), CL097 (CL) or CpG ODN2216 (CpG), and the soluble mediators (as indicated above each histogram) in the culture supernatants were quantified by Luminex Multiplex Assay (n=2).

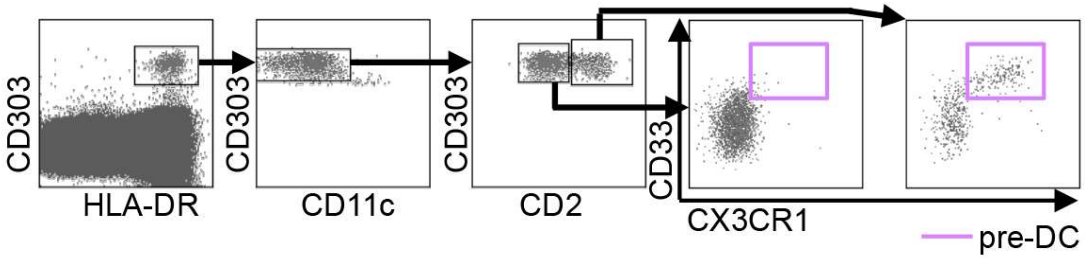



**Fig. S8**. Identification of $CD33^{+}CX3CR1^{+}$ pre-DC among $Lin^{-}HLA\text{-}DR^{+}CD303^{+}CD2^{+}$ cells (*34*).

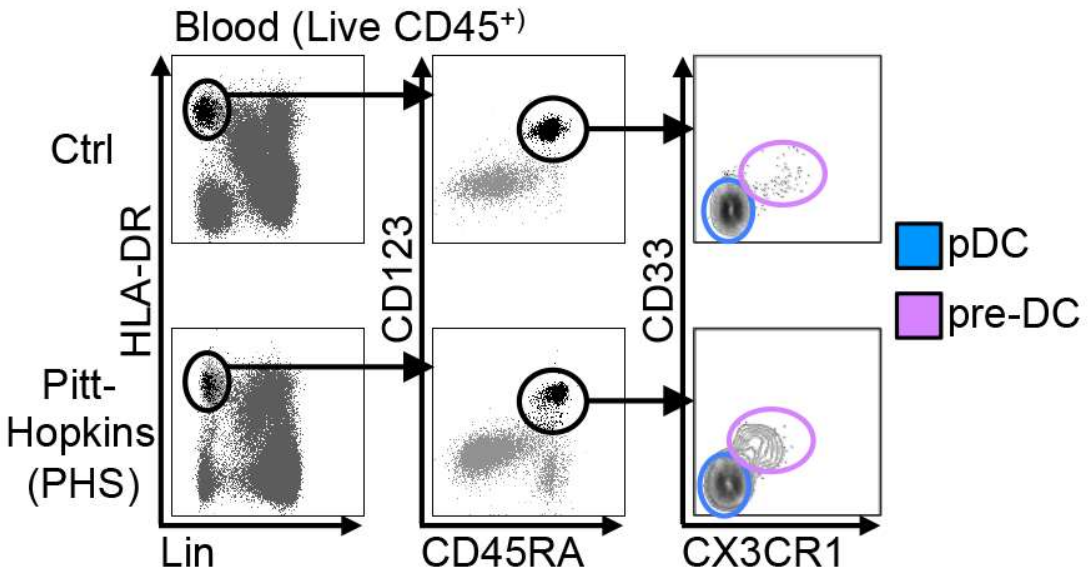



**Fig. S9**. Gating strategy for the fluorescence-activated cell sorting analysis of peripheral blood mononuclear cells from control subjects (Ctrl, n=11) and patients with Pitt-Hopkins Syndrome (PHS; n=4). pDC (circled in blue) and pre-DC (circled in purple) were defined among Lin-HLA-DR$^{+}$CD45RA$^{+}$CD123$^{+}$ cells.

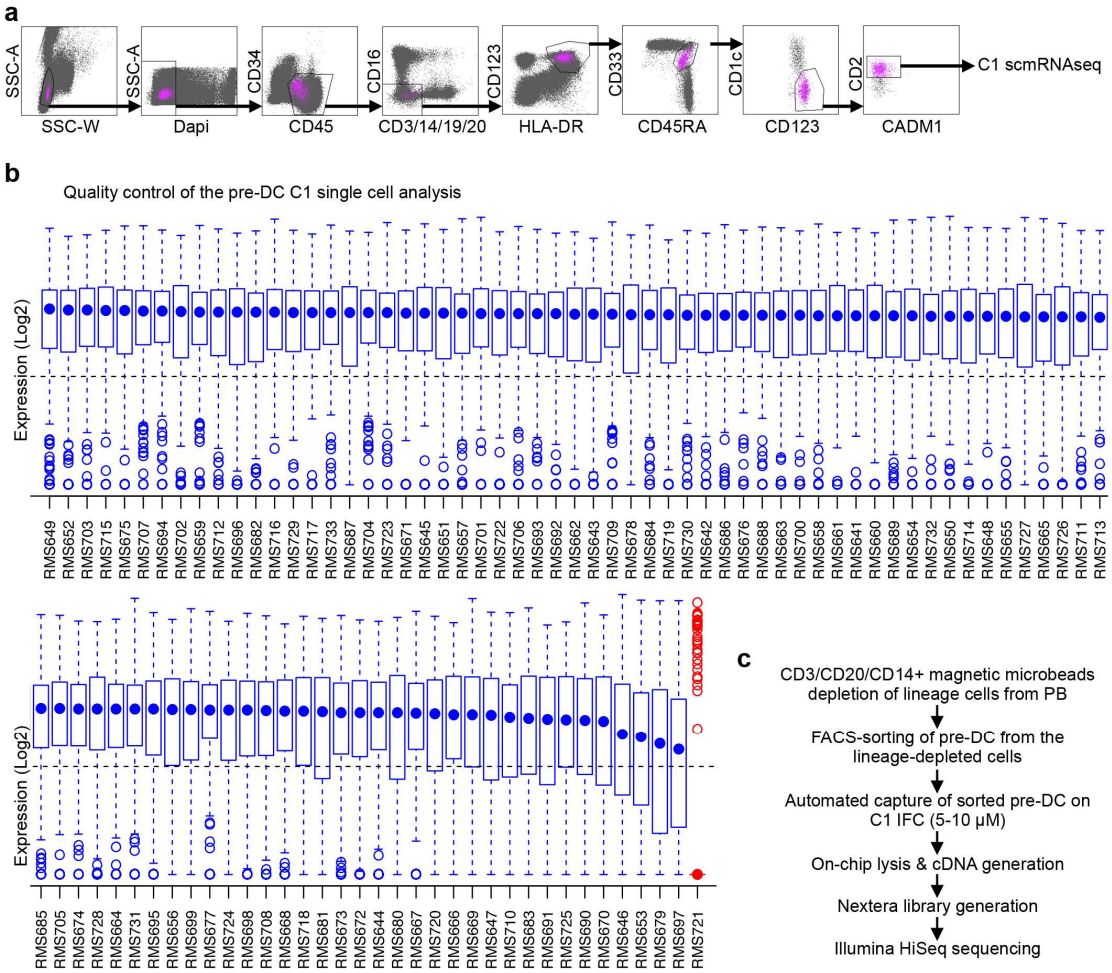



**Fig. S10**. **a.** Gating strategy for FACS of Lin$^-$HLA-DR$^+$CD33$^+$CD45RA$^+$CD1c$^{lo/-}$CD2$^+$CADM1$^{lo/-}$CD123$^+$ pre-DC analyzed by C1 single cell mRNA sequencing (scmRNAseq). **b.** Quality control (removing low-quality cells and minimally-expressed genes below the limits of accurate detection; low-quality cells that were identified using SINGuLAR toolbox; minimally-expressed genes with transcripts per million (TPM) values ≥1 in <95% of the cells) and **c.** work flow of the C1 scmRNAseq analyses shown in Fig. 3A-B. Error bars represent the maximum, third quartile, median, first quartile and minimum.

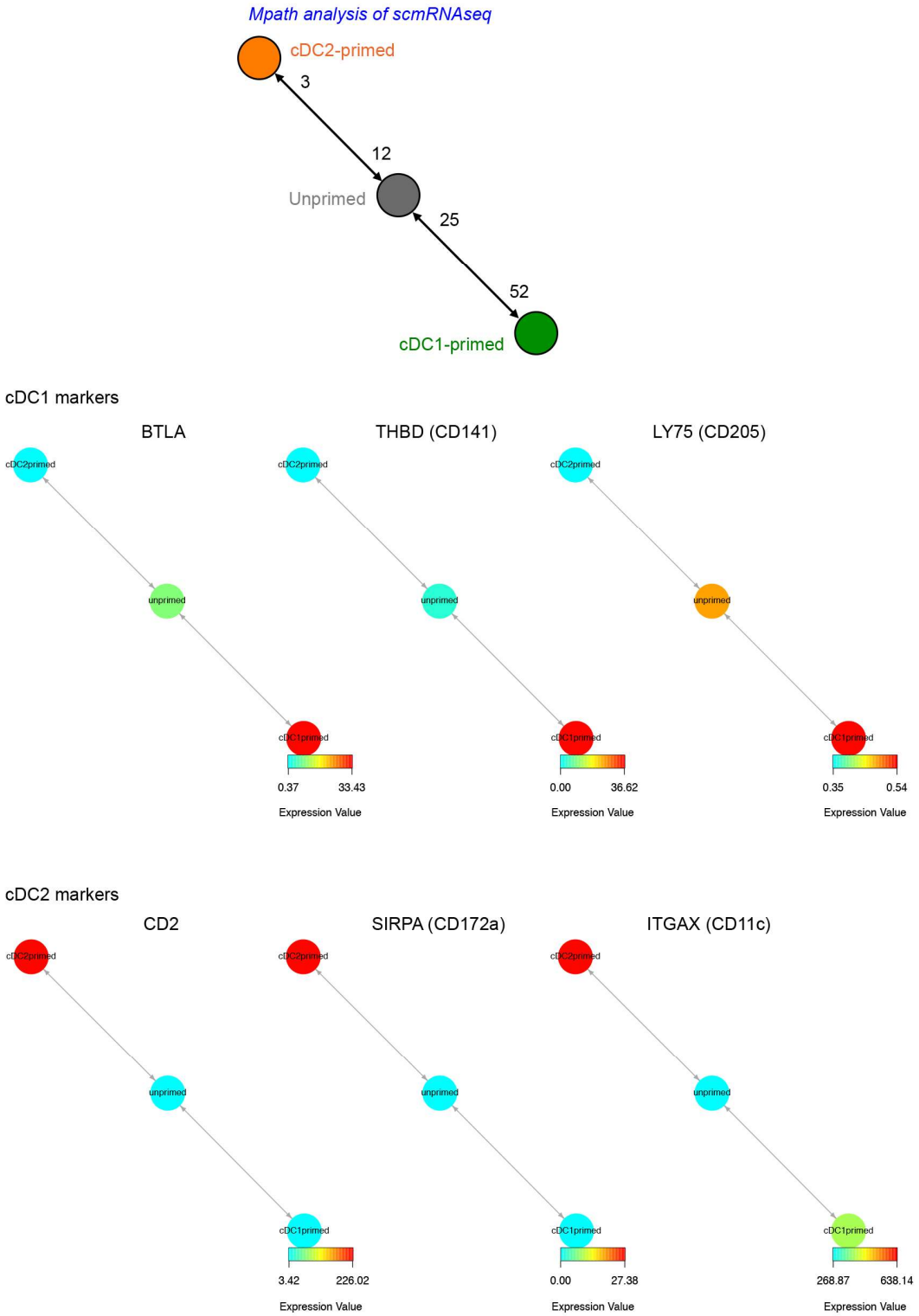



**Fig. S11**. Relative expression levels of signature genes of cDC1 (BTLA, THBD and, LY75) and cDC2 (CD2, SIRPA and ITGAX) in Mpath clusters defined in Fig. 3B.

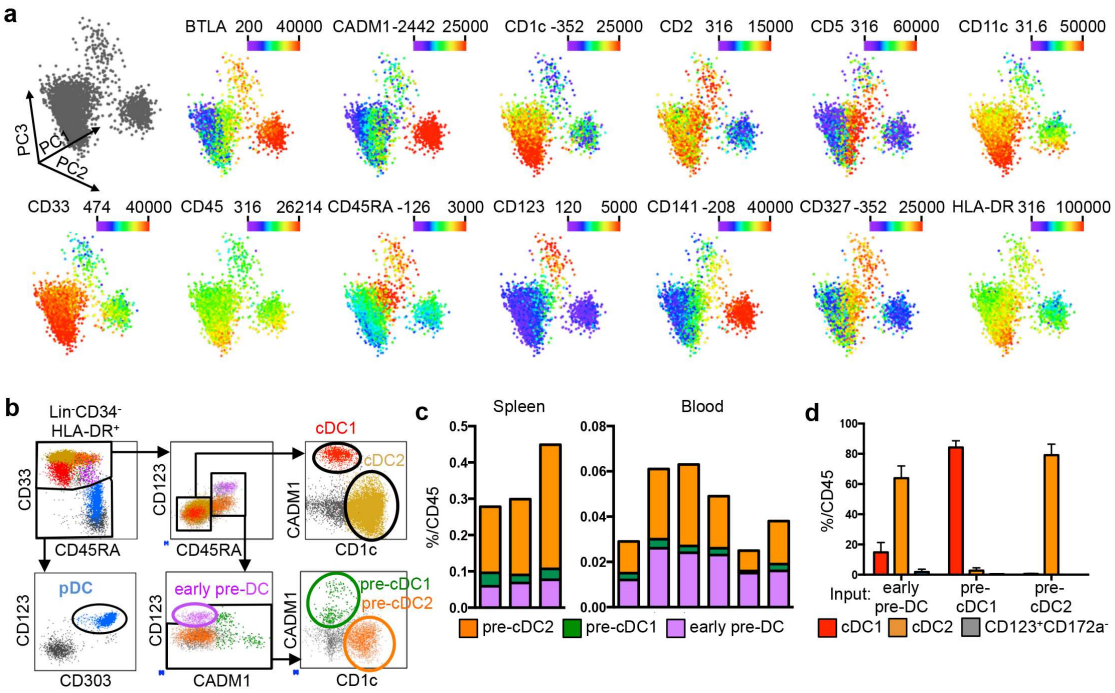



**Fig. S12**. **a.** Expression level of markers in the 3D-Principal Component Analysis (PCA) plots from Fig. 3, C and D are shown. **b.** Sequential gating strategy of flow cytometry data starting from Live $CD45^{+}Lin(CD3/14/16/19/20)^{-}CD34^{-}HLA\text{-}DR^{+}$ peripheral blood mononuclear cells defining $CD33^{-}CD123^{+}CD303^{+}$ pDC, $CD33^{+}CD45RA^{-}$ differentiated cDC ($CADM1^{+}$ cDC1, $CD1c^{+}$ cDC2), and $CD33^{+}CD45RA^{+}$ cells (comprising $CD123^{+}CD45RA^{+}$ pre-DC and $CD123^{lo}CD45RA^{+}$ intermediate cells). **c.** Proportion of $CD45^{+}$ mononuclear cells in spleen (n=3) (left) and peripheral blood (n=6) (right) of the above-mentioned pre-DC subsets. **d.** Histograms of the mean proportion of $CD303^{+}CD172\alpha^{-}$ cells, $Clec9A^{+}CADM1^{+}$ cDC1 or $CD1c^{+}CD11c^{+}$ cDC2 obtained in the *in vitro* differentiation assays as described in Fig. 3H (n=3). Error bars represent mean ± SEM.

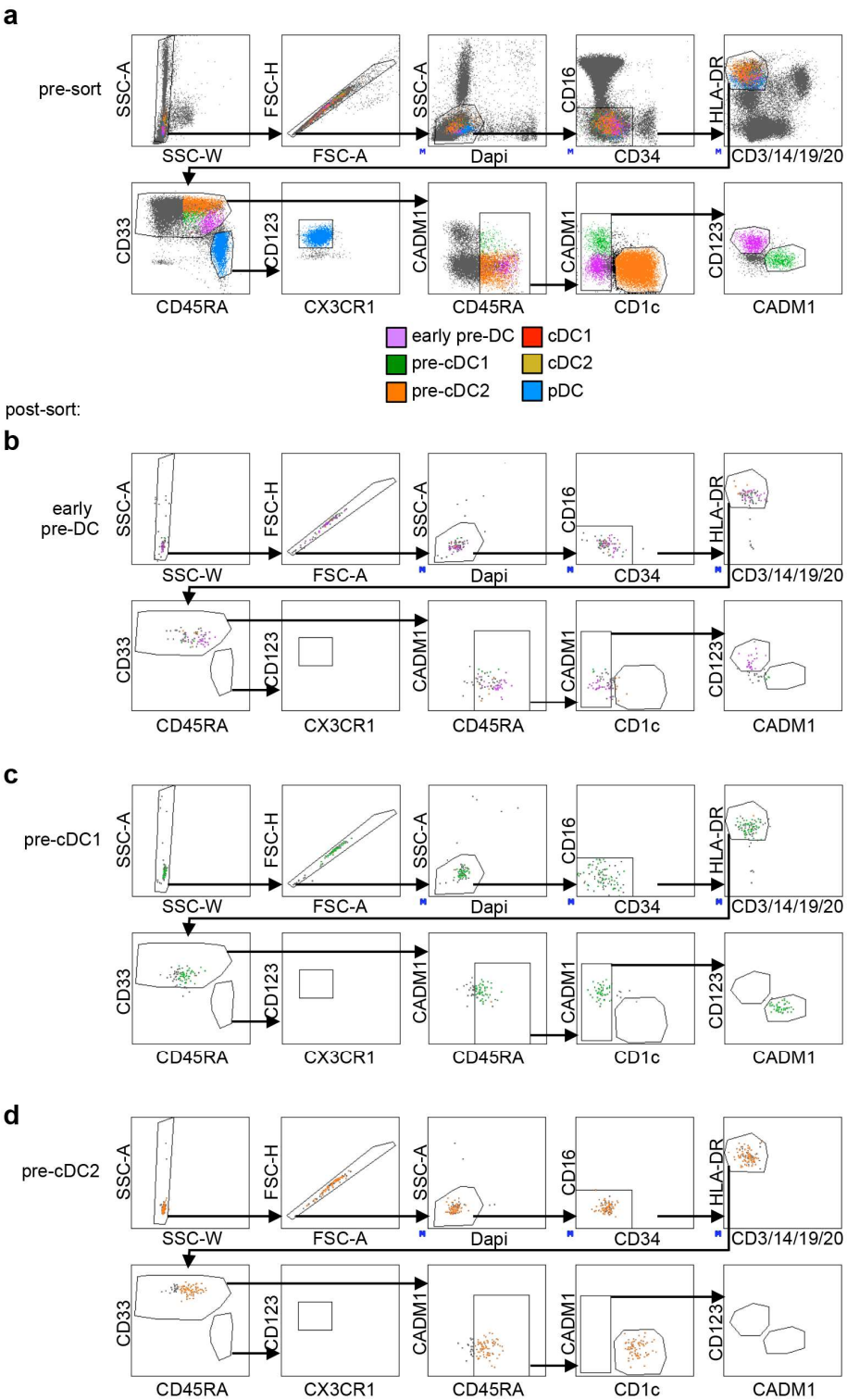



**Fig. S13**. Gating strategy for sorting of pre-DC subsets used in the *in vitro* differentiation assays (Fig. 3G). **a.** Pre-sorted data and **b-d.** post-sorted re-analysis of **b.** early pre-DC, **c.** pre-cDC1, and **d.** pre-cDC2 are shown.

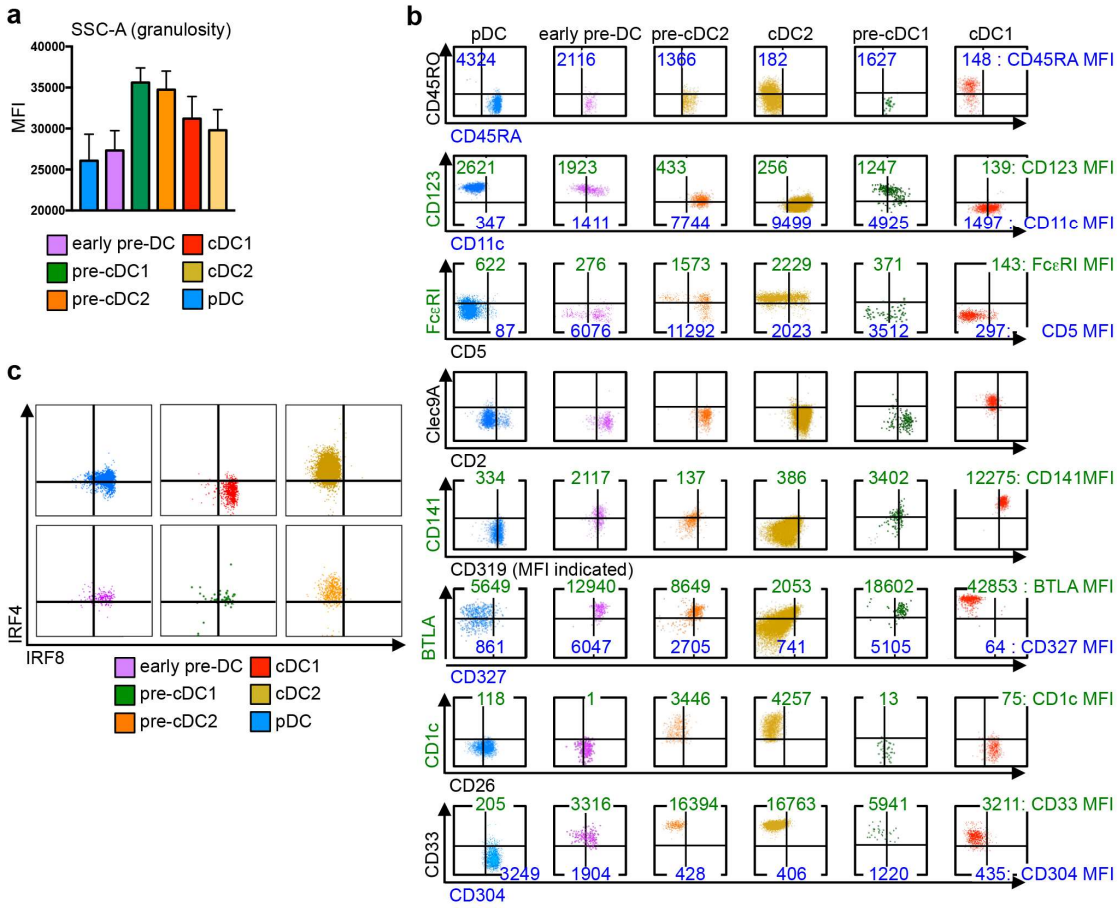



**Fig. S14**. **a.** Expression level in terms of mean fluorescence intensity (MFI) of the side scatter area (SSC-A) indicating cellular granularity of blood pre-DC and DC subsets from five individual human donors (n=5) Error bars represent mean ± SEM.. **b-c.** Flow cytometry data showing the relative expression of **b.** CD45RA, CD169, CD11c, CD123, CD33, FcεRI, CD2, Clec9A, CD319, CD141, BTLA, CD327, CD26, CD1c, CD304 or of **c.** IRF4 and IRF8 by pDC, early pre-DC, pre-cDC2, cDC2, pre-cDC1 and cDC1 defined in Fig. 3G and in fig. S12b.

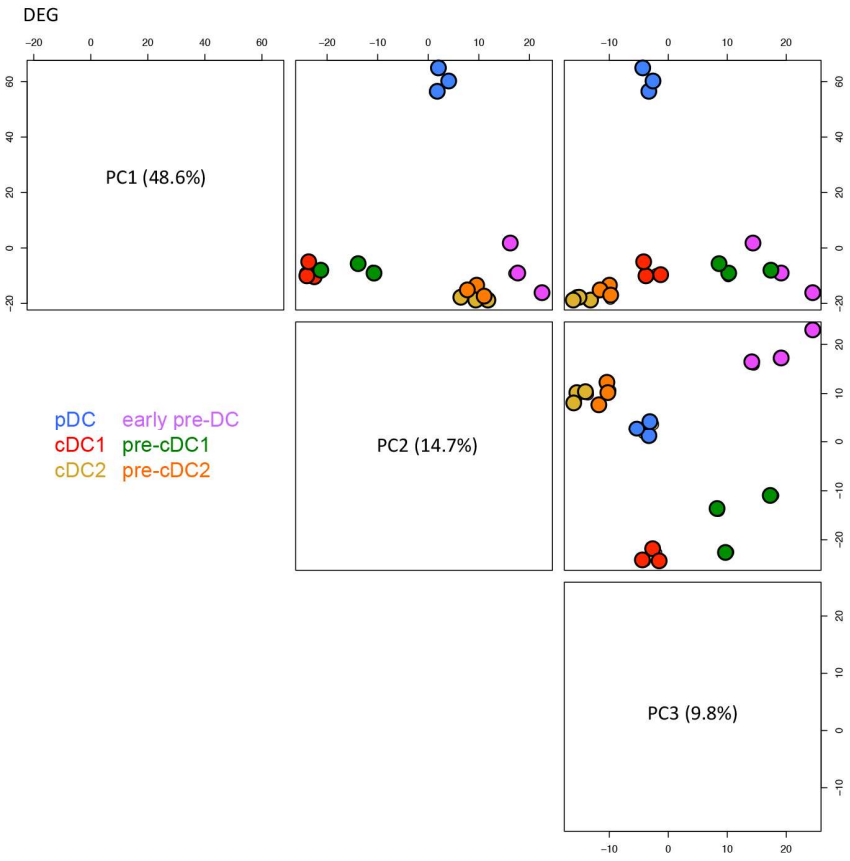



**Fig. S15**. 2D-plots showing combinations of Principal Component Analysis components 1, 2 or 3 (PC1-3) using differentially-expressed genes from the microarray analysis of Fig. 4.

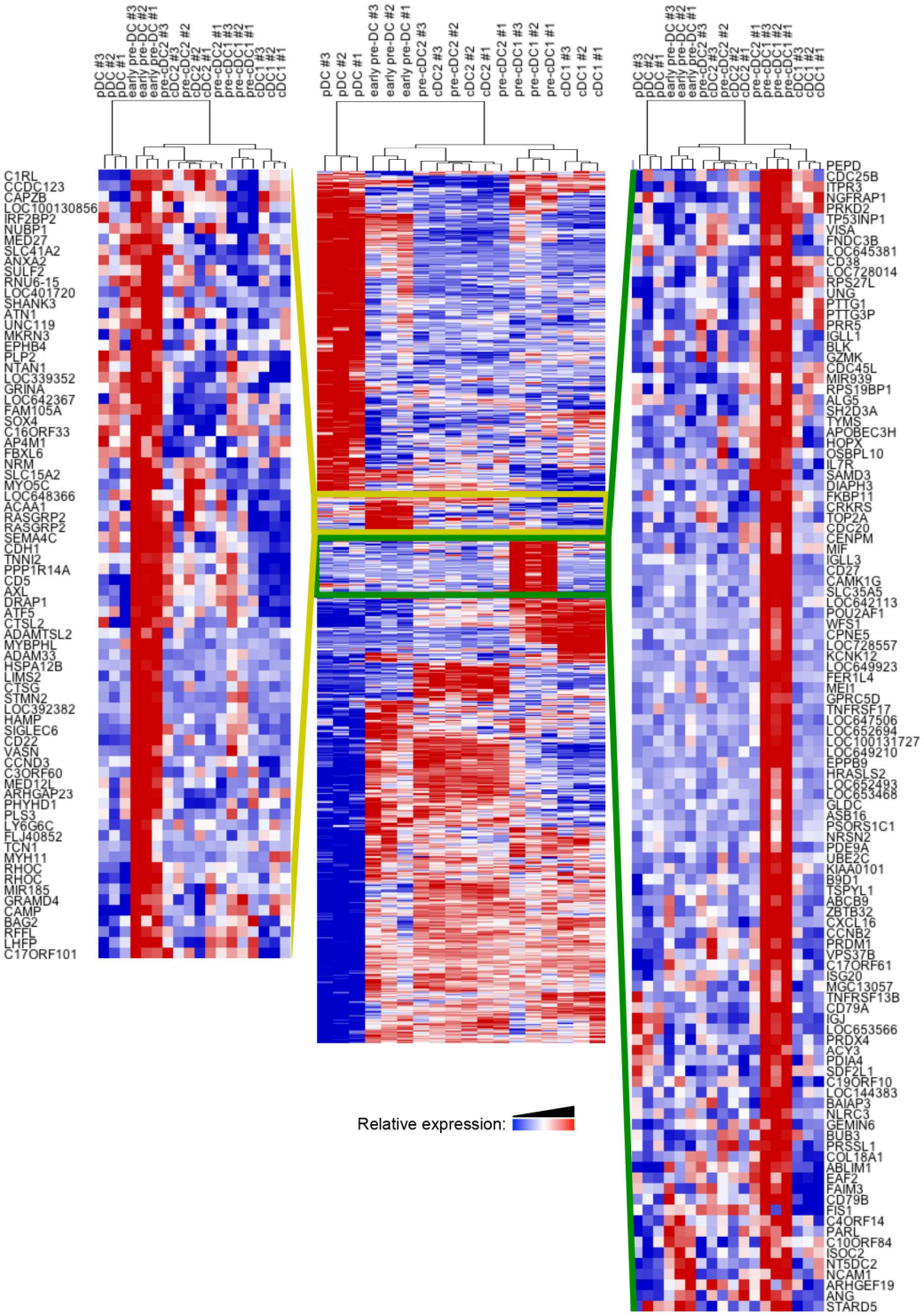



**Fig. S16**. Heat maps of relative expression levels of all differentially-expressed genes, with magnifications of the specific genes in early pre-DC (boxed yellow region) and pre-cDC1 (boxed green region) from the microarray analysis of Fig. 4.

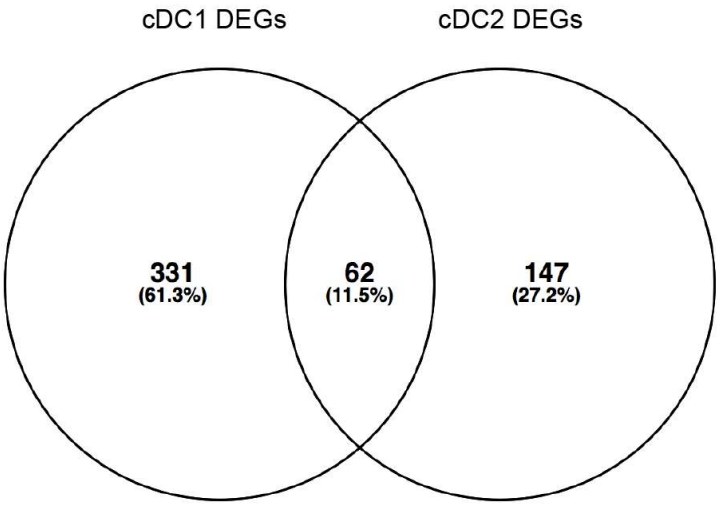



**Fig. S17.** Venn diagram showing genes common between the lists of cDC1 DEGs (the union of DEGs from comparing pre-cDC1 vs early pre-DC and cDC1 vs pre-cDC1) and cDC2 DEGs (the union of DEGs from comparing pre-cDC2 vs early pre-DC and cDC2 vs pre-cDC2). These 62 genes were then plotted in Fig. 4E with the log2 fold-change values (versus early pre-DC).

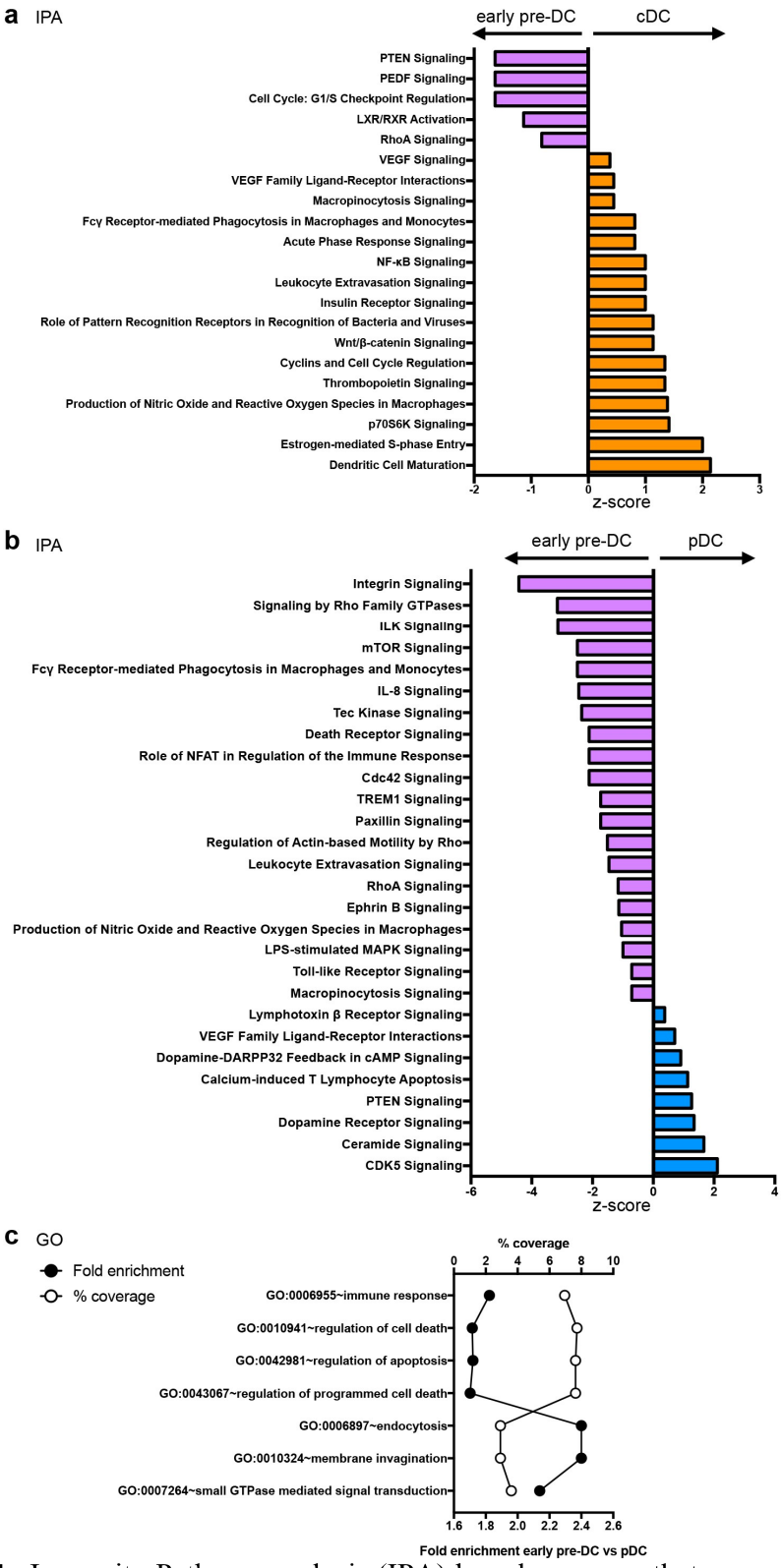



**Fig. S18**. **a-b.** Ingenuity Pathway analysis (IPA) based on genes that were differentially-expressed between **a.** cDC and early pre-DC or **b.** pDC and early pre-DC. Only the DC

biology-related pathways were shown, and all displayed pathways were significantly enriched ($P<0.05$, right-tailed Fischer's Exact Test). The heights of the bars correspond to the activation z-scores of the pathways. Enriched pathways predicted to be more activated in early pre-DC pathways are shown in pink and enriched pathways predicted to be more activated in cDC or pDC are shown in orange and blue, respectively. IPA predicts pathway activation/inhibition based on the correlation between what is known about the pathways in the literature (the Ingenuity Knowledge Base) and the directional expression observed in the user's data. Please refer to IPA Upstream Regulator Analysis Whitepaper and IPA Downstream Effectors Analysis Whitepaper for full description of the activation z-score calculation. **c.** Gene Ontology (GO) enrichment analysis of differentially-expressed genes (DEGs) in early pre-DC and pDC indicating biological processes that were significantly enriched (Benjamini-Hochberg adjusted p value $< 0.05$) with genes expressed more abundantly in early pre-DC as compared to pDC. Note that no biological process was significantly enriched with genes expressed more abundantly in pDC as compared to early pre-DC.

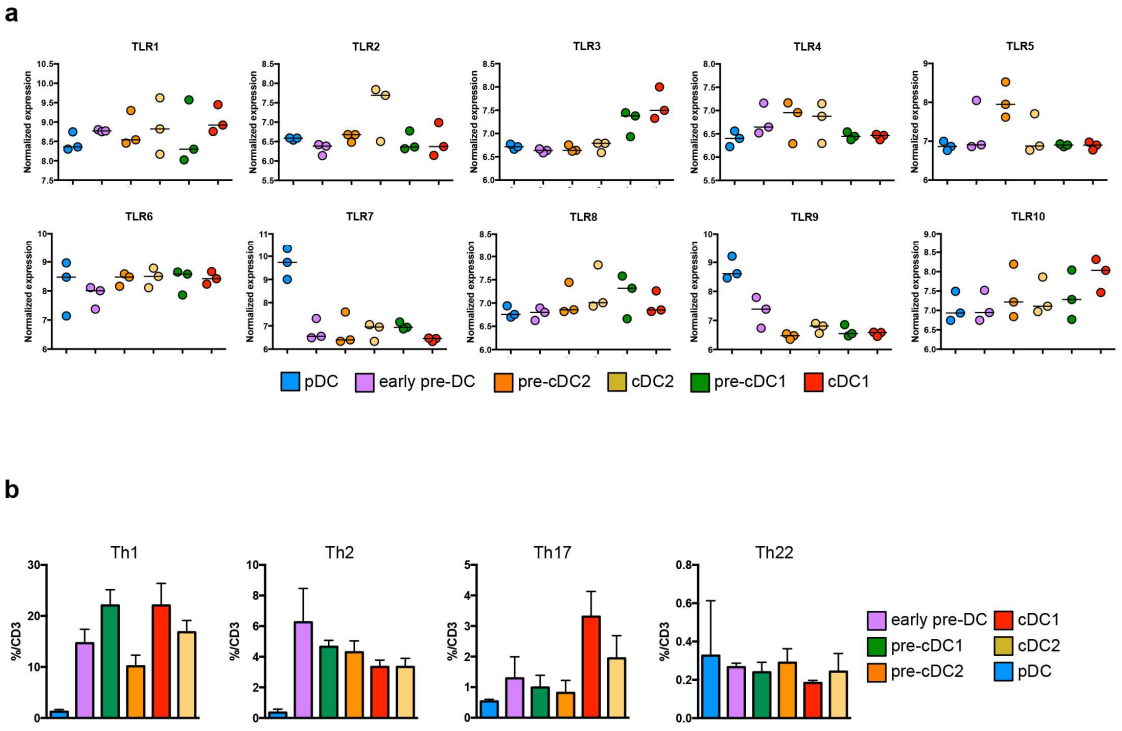



**Fig. S19**. **a.** Normalized abundance of all TLR mRNA in DC and pre-DC subsets obtained from the microarray analysis of Fig. 4. **b.** Polarization of naïve $CD4^+$ T cells into $IFN\gamma^+IL\text{-}17A^-$ Th1 cells, $IL\text{-}4^+$ Th2 cells, $IL17A^+IFN\gamma^-$ Th17 cells and $IL\text{-}22^+IFN\gamma^-IL\text{-}17A^-$ Th22 cells after 6 days of culture in a mixed lymphocyte reaction with allogenic pre-DC and DC subsets (n=3). Error bars represent mean ± SEM.

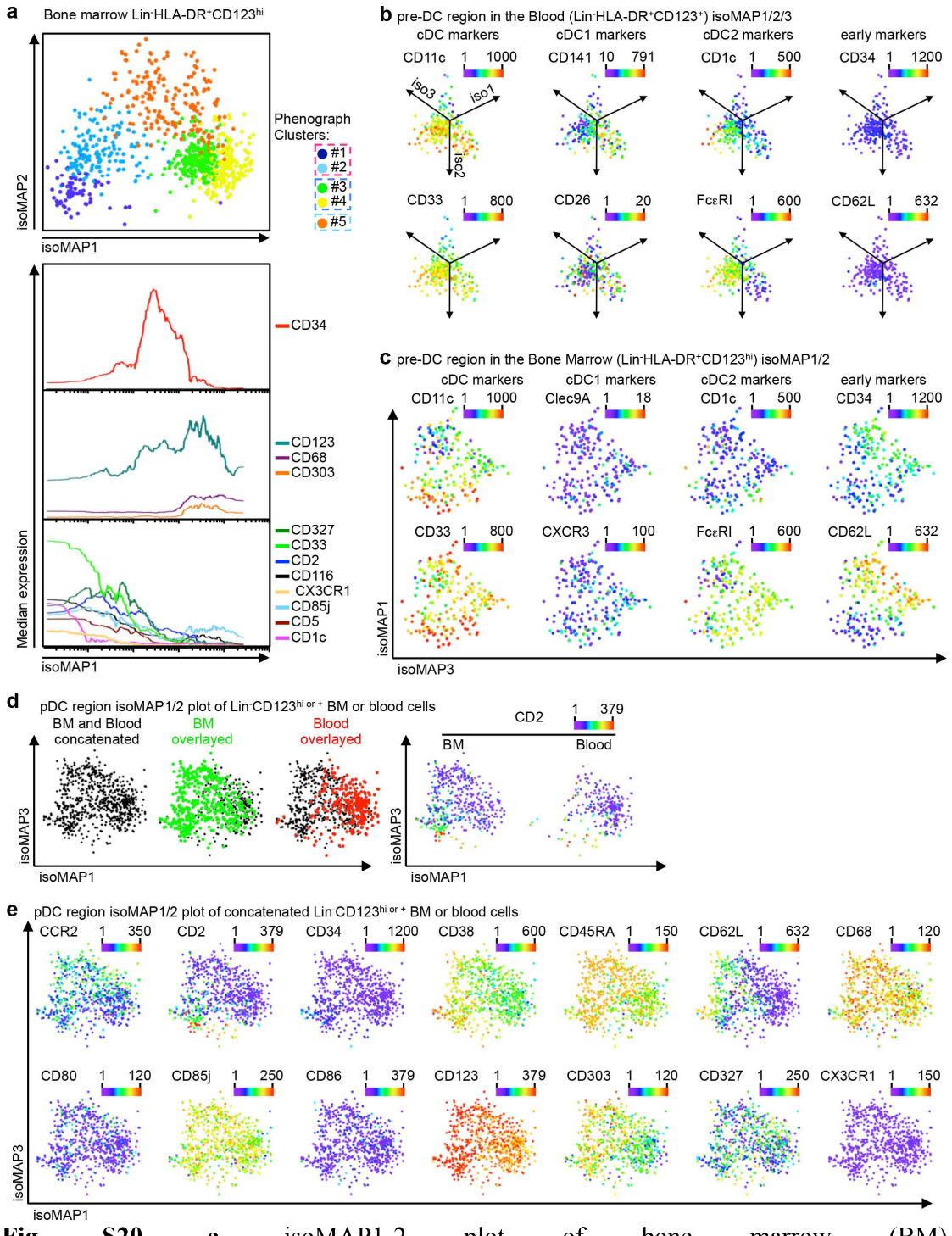



**Fig. S20. a.** isoMAP1-2 plot of bone marrow (BM) Lin(CD3/CD7/CD14/CD15/CD19/CD34)$^{-}$CD123$^{hi}$ cells (upper panel) and graphics of the binned median expression of defining markers along the phenotypic progression of cells

defined by the isoMAP1 dimension (lower panels) are shown. **b.** Expression level of selected markers in the isoMAP1-2-3 3D-plots (Fig. 6C, lower left panel) corresponding to cells within the pre-DC phenograph clusters (#1 and #2) of the blood $Lin^-CD123^+$ cells isoMAP analysis. **c.** Expression level of selected markers in the isoMAP1-2 plots (Fig. 6C, upper left panel) corresponding to cells within the pre-DC phenograph clusters (#3 and #4) of the BM $Lin^-CD123^{hi}$ cells isoMAP analysis. **d.** pDC defined in BM $Lin^-CD123^{hi}$ (green: phenograph clusters #3 and #4) or blood $Lin^-CD123^+$ (red: phenograph cluster #7) cells of Fig. 6A and 6B, respectively, were exported and analyzed using the isoMAP method and subdivided into clusters using the phenograph algorithm. BM and blood concatenated (black) and overlaid BM (green) and blood (red) isoMAP1/3 plots are shown (left panels). Expression level of CD2 in BM (left) and blood (right) pDC in the isoMAP1/3 plot. **e.** Expression level of selected markers in the BM and blood concatenated isoMAP1/3 plot of Fig. 6C (right panels).

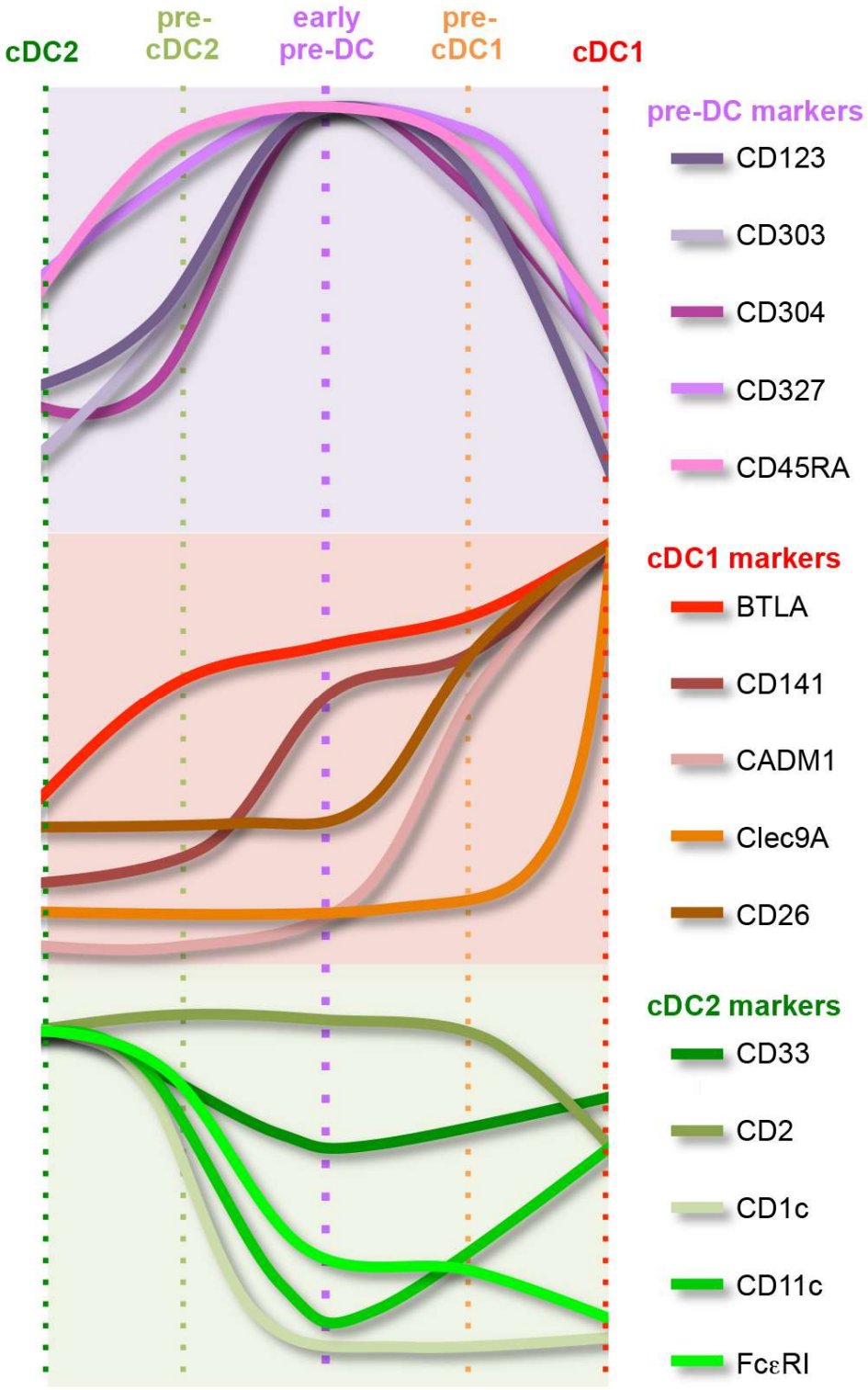



**Fig. S21.** Schematic representation of the expression of major pre-DC, cDC1 and cDC2 markers as pre-DC differentiate towards cDC.

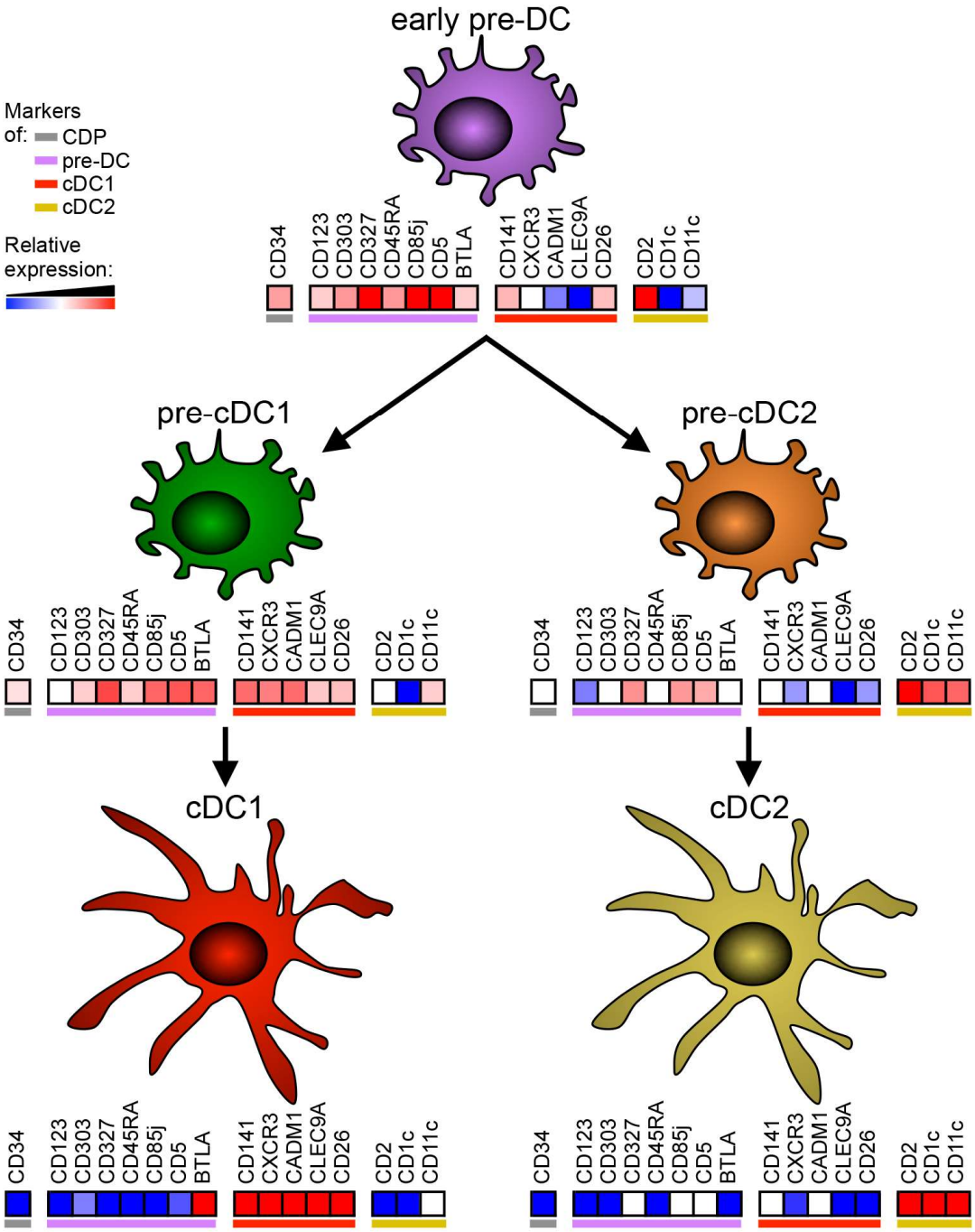



**Fig. S22.** Schematic representation of the expression of major pre-DC, cDC1 and cDC2 markers as pre-DC differentiate towards cDC.

**Supplementary Tables S1-S16 are provided online as a separate file:**

**Table S1.** Number of detected genes per cell in the total DC MARS-seq experiment

**Table S2.** DC subsets signature genes derived from Gene Expression Omnibus data series GSE35457 and used for MARS-seq and C1 data analyses

**Table S3.** List of anti-human antibodies used for mass cytometry (CyTOF)

**Table S4.** Number of expressed genes detected per cell in the pre-DC C1 scmRNAseq experiment

**Table S5.** Lists of genes identified from the microarray DEG analysis comparisons along the lineage progression from early pre-DC to mature cDC, for cDC1 and cDC2 respectively, and the list of the 62 common genes

**Table S6.** List of anti-human antibodies used for flow cytometry

Publication IV

**Günther, P.*,** Cirovic, B.*, Baßler, K.*, Händler, K., Becker, M., Dutertre, C.A.A., Bigley, V., Newell, E., Collin, M., Ginhoux, F., Schlitzer, A. and Schultze, J.L. (2019) ‘A Rule-Based Data-Informed Cellular Consensus Map of the Human Mononuclear Phagocyte Cell Space’. *BioRxiv*, p. 658179. DOI: 10.1101/658179.

* Shared first authorship

My contribution to this publication includes:

Conceptualization, study design and planning of experiments; Methodology and Software; Investigation; Data analysis, visualization and Interpretation of all scRNA-seq data (excluding cord blood dataset); Writing

# A rule-based data-informed cellular consensus map of the human mononuclear phagocyte cell space

Patrick Günther[1,†], Branko Cirovic[2,†], Kevin Baßler[1,†], Kristian Händler[3], Matthias Becker[3], Charles Antoine Dutertre[4], Venetia Bigley[5], Evan Newell[4], Matthew Collin[5], Florent Ginhoux[4,6], Andreas Schlitzer[2,*], Joachim L. Schultze[1,3,*]

**Affiliations:**

[1]Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany.

[2]Myeloid Cell Biology, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany.

[3]Platform for Single Cell Genomics and Epigenomics, German Center for Neurodegenerative Diseases and the University of Bonn, 53175 Bonn, Germany.

[4]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), BIOPOLIS, Singapore, Singapore.

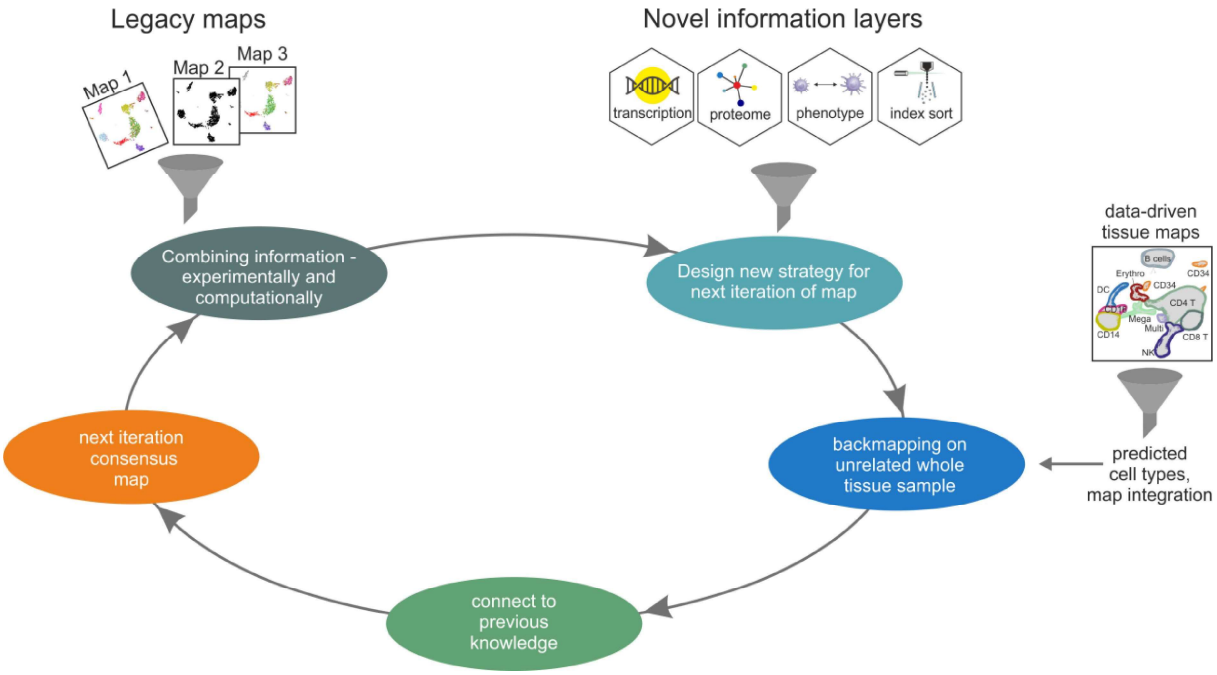
[5]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

[6]Shanghai Institute of Immunology, Shanghai JiaoTong University School of Medicine, 280 South Chongqing Road, Shanghai, China.

[†]shared first authors

[*]shared last authors, corresponding authors

## Graphical Abstract



**Highlights**

- Defining a consensus of the human myeloid cell compartment in peripheral blood
- 3 monocytes subsets, pDC, cDC1, DC2, DC3 and precursor DC make up the compartment
- Distinguish myeloid cell compartment from other cell spaces, e.g. the NK cell space
- Providing a generalizable method for building consensus maps for the life sciences

**Abstract**

Single-cell genomic techniques are opening new avenues to understand the basic units of life. Large international efforts, such as those to derive a Human Cell Atlas, are driving progress in this area; here, cellular map generation is key. To expedite the inevitable iterations of these underlying maps, we have developed a rule-based data-informed approach to build next generation cellular consensus maps. Using the human dendritic-cell and monocyte compartment in peripheral blood as an example, we performed computational integration of previous, partially overlapping maps using an approach we termed 'backmapping', combined with multi-color flow-cytometry and index sorting-based single-cell RNA-sequencing. Our general strategy can be applied to any atlas generation for humans and other species.

## Introduction

Since Robert Hooke's first observations of cells as the basic unit of life, generations of life scientists have been driven to understand, map and characterize individual cells (Cavaillon, 2011). For many decades, morphological parameters were the major driving force to establish new cell identities (Hussein et al., 2015). In immunology, technologies such as flow cytometry have been developed that permit quantitative enumeration of single cells based on measuring combinations of predominantly cell-surface proteins (Hulett et al., 1969; Perfetto et al., 2004). These technologies, however, have some undisputable limitations, most notably, their reliance on a predefined subset of biomolecules. Conversely, single-cell-omics, particularly single-cell transcriptomics, allow for cells to be assessed, in principle, without predefined markers. Here, the complete spectrum of transcriptomic parameters is investigated and used as a defining unit of cell identity (Islam et al., 2014; Macosko et al., 2015; Tang et al., 2009). Such single-cell technologies allow for a fully data-driven analysis to establish cell maps of an organism, such as those proposed by the Human Cell Atlas consortium (Rozenblatt-Rosen et al., 2017). We have learnt from other disciplines that maps require iterations over time, often due to new data generated as a result of technological advances. These iterations improve the precision, accuracy and available content per data point (Edney, 2019; Monmonier, 2015; Ridpath, 2007).

Reliable consensus maps are a prerequisite to reconcile conflicting data that might have been generated based on different data generating approaches (Edney, 2019; Monmonier, 2015). Here we generalize the approach of building geographic or astronomic consensus maps to human cellular consensus maps. We exemplify our approach by integrating two recently introduced single-cell transcriptomics-based cellular maps of the human blood mononuclear myeloid cell compartment (See et al., 2017; Villani et al., 2017) with novel single-cell transcriptomics and flow cytometry data. The human blood mononuclear myeloid cell compartment has been recognized to harbor a complex mixture of cells of diverse origins exemplified by the ongoing efforts to map

this cellular compartment with increasing resolution (Dutertre et al., 2019). The two mapping efforts present with discrepancies and commonalties in terms of cell type identification, naming and breadth of sampling. In order to establish a consensus map of the human mononuclear myeloid cell compartment we allow for the integration of prior knowledge in that we define *a priori* criteria for the cellular compartment under study in order to increase resolution and to allow building of a consensus map. Overall, our approach generates rule-based data-informed cellular consensus maps that resolve discrepancies between the two recently generated maps, and clarifies cellular identities of human dendritic-cell (DC) and monocyte subsets resulting in a novel, integrated consensus map of the human blood myeloid compartment.

## Results

### *Integrated phenotypic characterization of the myeloid cell compartment in human peripheral blood*

We aimed to build a consensus map of healthy human blood myeloid cells that integrates legacy dataset knowledge into a revised consensus map. To do so, we generated a novel single-cell-omics dataset of the blood $CD45^+Lin^-HLA\text{-}DR^+$ cell space using a 17 parameter index sorting panel incorporating important markers from two recently published single-cell–omics datasets, here termed map 1 (Villani et al., 2017) and 2 (See et al., 2017) and an established panel of myeloid cell markers including CD14, CD16, HLA-DR, CD1c and CADM1 (Dutertre et al., 2014; Guilliams et al., 2016; Haniffa et al., 2012), to link the data to the body of knowledge already present within the literature (**Figure 1A, S1A-C, Table S1**). This strategy allowed us to directly include several cell populations defined by either map 1 or 2 into our single-cell transcriptomics dataset, compare these populations within an unbiased myeloid cell space dataset, and assess differences and commonalities between the two maps.

To understand the organization of the blood-derived myeloid cell compartment, we performed dimensionality reduction using the uniform manifold approximation and projection (UMAP) algorithm (Becht et al., 2018) on the complete flow cytometry space of live $CD45^+$ $Lin^-$ cells (**Figure 1B**). UMAP revealed a complex topology of the flow cytometry data, segregating a large cluster on the right and multiple small entities on the left of the topology. A fraction of the $Lin^-$ cells (**Figure 1C**, cluster two**)** was not part of the monocyte or DC cell space according to CD16, CD14 and HLA-DR expression (**Figure 1D**). These cells most likely represent basophils due to their lack of HLA-DR expression but high CD123 expression (**Figure 1D, Figure S1**). To fully understand the population structure of the presented FACS-based UMAP, we performed Phenograph clustering of the live $CD45^+$ $Lin^-$ blood-derived flow cytometry UMAP space and detected 27 clusters (**Figure 1C**). To link these novel data to the two existing maps for the blood myeloid cell compartment (Guilliams et al., 2014; See et al., 2017; Villani et al., 2017) we reapplied the gating

strategies of either map 1 or map 2 and overlaid these onto our novel flow cytometry-derived UMAP topology (**Figure 1E, Figure S1A-C**). This analysis revealed several commonalities and discrepancies between maps 1 and 2 in the combined novel flow cytometry panel used in this study. On the upper-most level, map 1 was less stringent in excluding HLA-DR$^-$ cells within the myeloid cell space (cells labeled light grey, **Figure 1D, S1B**), a feature rigorously adhered to in map 2 (cells labeled dark grey, **Figure 1D, S1C**). Furthermore, Axl$^+$Siglec6$^+$ DCs (AS-DC; DC 5, **Table S1**) in map 1 occupied the same topological space as pre-DCs in map 2, indicating potential cellular overlap. Finally, map 1 mono 2/4, resembling non-classical monocytes (ncMono) (**Table S1**), occupied two different locations on the UMAP topology: one of them being within the HLA-DR$^-$ compartment of the topology and the other being within the space assigned to monocytes by a classical investigator-derived flow cytometry gating (**Figure 1D, 1E, S1**). These data suggest that there is a commonality in the identity of map 1 Axl$^+$Siglec6$^+$ DCs (AS-DC; DC 5) and map 2 pre-DCs whereas mono 2/4 may represent a heterogeneous mixture of various cell types – apparently not all of them related to the myeloid cell lineage.

***Novel integrated single cell-omics data identifies commonalties and discrepancies between two recent myeloid cell maps***

To investigate the cell population structure at the transcriptomic level we performed single-cell RNA-sequencing (scRNA-seq) of 2,509 blood-derived single cells following index sorting to encompass all major populations identified in either map 1 or 2 after lineage exclusion and generated a UMAP dimensionality reduction-based transcriptome map (**Figure 1B, S2A-F**). *De novo* clustering of the scRNA-seq data revealed 11 transcriptionally different clusters (**Figure 2A, 2B, Data Table S1**). We projected the cluster identities onto the flow cytometry-derived UMAP topology, which allowed us to validate our index sorting strategy and link identities across the flow cytometry and scRNA-seq data (**Figure 2A**).

Gene level inspection of these clusters revealed that cluster one had a natural killer (NK) cell signature, as indicated by PRF1, GNLY and NKG7 gene expression. Cluster two was identified by high IGH family gene expression, thus implying contaminating B cells with a strong plasmablast signature (Shi et al., 2015) (**Figure 2B**). Cluster three was represented by a minor fraction of the cells within our dataset, with a profile of microRNA-related transcripts. Cluster four expressed SPINK2, GAS5, SATB1and STMN1 genes, and thus corresponded to blood-derived $CD34^+$ hematopoietic stem cells (Satoh et al., 2013; Will et al., 2013). Cluster five expressed the plasmacytoid DC (pDC)-related IRF7, TCF4 and GZMB transcripts (See et al., 2017; Villani et al., 2017), whereas cluster six expressed a conventional dendritic cell 1 (cDC1) gene-set, with high expression of CLEC9A, IDO1 and CD74 (van der Aa et al., 2015; Zhang et al., 2012). Interestingly, cluster seven expressed genes either affiliated to pre-DCs (See et al., 2017) or AS-DCs (DC5) (Villani et al., 2017), such as SIGLEC6, AXL, PLAC8 or LILRA4, thus associating them to the human pre-DC continuum. As expected from our sorting strategy, we also detected several clusters belonging to the monocyte lineage. Cluster eight represented $CD16^+$ ncMono cells based on high FCGR3A (CD16) with SERPINA1 and DUSP6 expression. Conversely, cluster nine expressed S100A8, S100A9 and S100A12 together with VCAN and FCN1, identifying them as classical $CD14^+$ monocytes (Mono1/cMono). Clusters 10 and 11 represented two cDC2 identities (DC2, DC3): both clusters expressed high levels of the cDC2-related CD1C, CD1E and several HLA-DR transcripts. Interestingly, and as shown in map 1 (Villani et al., 2017), cluster 11 co-expressed certain monocyte-affiliated gene products, such as S100A8, S100A89 and FCN1 (**Figure 2B**), as also shown in Dutertre et al. (Dutertre et al., 2019).

To develop our consensus map, we utilized the index sorting data of the populations identified in map 1 and 2 and mapped them onto our single-cell transcriptomic dataset (**Figure 2C**). Overlaying this index-sorting data onto the scRNA data-derived UMAP topology reiterated several commonalities between maps 1 and 2, including DC1/cDC1 (purple), DC6/pDC (pink), Mono1/cMono (ochre) and $CD14^+CD16^+$ intermediate monocytes (Mono3/intMono, dark red).

Importantly, the detected discrepancies between maps 1 and 2 were also apparent on the transcriptomic level. Firstly, we noticed that map 1 double negative DCs (DN-DC) populated the same position within the UMAP topology as cDC2s in map 2. Furthermore, mapping the index sorting data of the map 1 ncMono population (transcriptionally defined in map 1 as Mono2 and 4, **Table S1**) revealed two separate cell clusters within the transcriptomic UMAP topology, indicating considerable cell-type heterogeneity within this population as defined by map 1. Interestingly one of the Mono2/4 clusters overlapped with the ncMono (magenta) cluster, whereas another cluster was mapped as HLA-DR$^-$ within the flow cytometry gating strategy used in map 2 (**Figure S1**). Moreover, we noticed that map 1 AS-DCs (red) and map 2 pre-DCs (red) occupied the same topological space, indicating considerable transcriptomic similarity despite different markers were used for their flow cytometric identification (**Figure 2C**). Taken together, the combined phenotypic and transcriptomic analysis presented here strongly argues for the need to further assess cellular identities within the myeloid cell compartment.

***Axl$^+$Siglec6$^+$ DCs phenotypically and transcriptionally overlap with human pre-DC***

To clarify the relationships and cellular identities of the different DC subsets and their progenitors in maps 1 and 2, we mapped individual protein and transcript information (**Figure S3**) and the transcriptomic signatures of DC subsets and their progenitors derived from map 2 (pDC, cDC1, cDC2, pre-DC) onto our scRNA-seq myeloid-cell-space data set (**Figure 3A**). By overlaying index sorting information and the initial unbiased clustering data, we revealed that specific map 2 pDC, cDC1 or pre-DC signatures were enriched in dense discrete cell clusters within the UMAP topology of the myeloid-cell-space scRNA-seq data, whereas the cDC2 signature was more broadly enriched within both the clusters associated with cDC2 and monocytes (**Figure 2A**, **3A**) suggesting a close relationship between these two cell types which is studied in further detail by Dutertre et al..

To integrate the identified DC subsets in map 1 and map 2 with each other, we computed a UMAP topology from the original map 1 single-cell transcriptome data comprising the DC cell space and overlaid the signatures of the map 2 DC subsets (pDC, cDC1, cDC2, pre-DC) (**Figure 3B**). This analysis showed that within the original map 1 transcriptomic data, map 2 pDC signatures mapped to the same topological space as DC6, thus identifying DC6 as *bona fide* pDCs. Similarly map 2 cDC1 transcriptomic signatures were enriched within map 1 DC1, whereas map 2 cDC2 signatures enriched in map 1 DC2, DC3 and DC4. Furthermore, mapping map 2 pre-DC signatures revealed the highest enrichment of this signature in map 1 DC5 (AS-DC), indicating the highest level of similarity between map 1 DC5 and map 2 pre-DC.

To validate these correlations between the DC types defined in maps 1 and 2, we investigated the enrichment of map 1-defined DC1-6 signatures within our new scRNA-seq consensus data (**Figure 3C**). Visualizing the scaled signature enrichment scores for DC1 showed correspondence between maps 1 and 2 cDC1 locations and between map 1 DC2, DC3 and map 2 cDC2 locations, respectively. Similarly, map 1 DC6 and map 2 pDC localized to the same topological space within our new scRNA sequencing data. The highest enrichment of map 2 pre-DC signatures (**Figure 3A**) and map 1 DC5 signatures (**Figure 3C**) was seen in cluster seven of our new scRNA-seq consensus data (**Figure 2A**), again indicating substantial transcriptomic overlap between map 1 DC5 and map 2 pre-DC.

We then investigated the potential differences in cell-type identity between map 1 AS-DCs (DC5) and map 2 pre-DCs (**Figure 2C**). We separately projected cells identified as pre-DCs by unbiased clustering of the flow cytometric data (cluster 26 in **Figure 1C**), map 1 DC5, map 2 pre-DC gated cells and cluster seven from our new scRNA-seq consensus data, which displayed precursor gene expression patterns, onto the novel combined flow cytometric-based UMAP topology (**Figure 3D**). This approach showed that FACS cluster 26 represented the intersection of map 1 DC5, map 2 pre-DCs and scRNA-seq cluster 7 and best reflected these progenitor cells at the protein level in an unbiased fashion. Certain differences between map 1 AS-DCs and map 2 pre-

DCs, however, became visible. Specifically, map 1 DC5 located only to a very discrete part of the topology and reached into a contaminating cDC2 space. FACS cluster 26 and map 2 pre-DCs occupied almost identical topological locations within the UMAP space, further illustrating the difficulties in discriminating pre-DCs and pDCs (**Figure 3D, 1C-D, 2A**). Interestingly, both map 1 AS-DCs and map 2 pre-DCs were best defined by FACS cluster 26, indicating that these cells represent the same cellular identity at both the surface marker and transcriptomic level. This finding was further supported when enriching transcriptomic signatures of map 2 pre-DCs across the spectrum of identified DC subtypes in map 1, resulting in a high enrichment of map 2-derived pre-DC signature genes within map 1 AS-DCs (**Figure 3E**). This enrichment was further reinforced by comparing hallmark genes within the cell populations defined in the legacy maps 1 and 2 (**Figure 3F**). In conclusion, these analyses demonstrate that map 1 DC5 and map 2 pre-DCs represent, to a large extent, the same pre-DC identities and therefore, might be best named according to already published guidelines (Guilliams et al., 2014; Schlitzer and Ginhoux, 2014) as pre-DCs.

***DN-DCs/DC4 resemble CD16$^{+}$ non-classical monocytes***

We were unable to locate the novel map 1 DC4 (DN-DC) subtype within a distinct cluster in our new scRNA-seq consensus data (**Figure 2C**). According to map 1 DC4 derived from a DN-DC subtype, being negative for the classical cDC subset markers CD1c, CD141 and CADM1 and pDC marker CD123 but positive for CD11c (Villani et al., 2017). To understand the role and placement of DC4 within the entire monocyte and DC space of both maps, we recapitulated the gating strategy originally used to delineate DC4 by map 1 (**Figure 4A**). Using the additional information from the newly included surface markers, such as CD16, we revealed that the large majority of DC4/DN-DCs (96.6%) were CD16$^{+}$ mononuclear cells (**Figure 4A**). We subsequently mapped the CD16$^{-}$ and CD16$^{+}$ fraction of the DC4/DN-DC compartment of map 1 onto our integrated flow cytometry-derived and scRNA-seq-derived UMAP topologies (**Figure 4B**).

Mapping onto the flow cytometry and scRNA-seq-derived UMAP topologies revealed that the CD16$^{+}$ DC4/DN-DC compartment was associated with the location traditionally occupied by ncMono and the CD16$^{-}$ fraction mapped into the topological region of the UMAP associated with pre-DCs and cDCs on both the phenotypic and transcriptomic level. To address the ambiguous DN-DC identity, we cross-referenced map 1 DN-DCs towards map 2 ncMono and the flow cytometry-based Phenograph cluster 15 (**Figure 4C**). Here we detected map 1 CD16$^{+}$ DN-DCs almost exclusively within the map 2 ncMono cluster and primarily contained within Phenograph cluster 15 (**Figure 1C**) derived from the combined flow cytometry panel, expressing ncMono-associated surface markers.

To extrapolate these surface phenotypic findings to the transcriptome level and understand the transcriptomic identity of map 1 DN-DCs, we correlated all transcriptomes of map 1 DC subsets with the map 1 monocyte subsets (mono 1-4) (**Figure 4D**). Pearson correlation revealed the highest level of correlation between DC4 and the map 1 mono 2 subset, with intermediate correlation with the mono 1, 3 and 4 subsets, and poor correlation with any map 1-identified DC subset (**Figure 4D**). Furthermore, enrichment of a mono 2-specific gene signature across all map 1-identified mononuclear cell identities showed enrichment in all monocyte-associated cell entities and DC4, further supporting that DC4 might be ncMono (**Figure 4E, S4A-C**). Additionally, we used map 1-derived DC4 signature genes and mapped them onto our scRNA-seq consensus data of the blood myeloid cell space (**Figure 4F**). This analysis showed a strong enrichment of DC4 signature genes within the cluster identified by unbiased cluster detection as having ncMono identity.

To reconcile DC4 with the existing spectrum of monocyte and DC subsets, we examined DC4 expression of SLAN — a marker for inflammation-associated ncMono (Hansel et al., 2011) using a new marker panel (**Figure 4G**) and UMAP-based visualization (**Figure 4H**). DC4 showed the expected SLAN expression levels for ncMono. To validate this finding and to exclude that DC4 are another subset within peripheral blood mononuclear cells (PMBCs) that we might not have

accounted for, we performed dimensionality reduction of the flow cytometric analysis in **Figure 4G** (**Figure 4H**) and mapped both CD16$^-$ and CD16$^+$ DN-DCs onto the UMAP topology (**Figure 4I**). Again, we found that DC4 associated with two different positions within this UMAP topology. Putting these two separate clusters within the DC4/DN-DC in the context of a conventional gating strategy of PBMC-derived mononuclear cells revealed a co-association between (i) classically defined ncMono and DC4/DN-DC that are CD16$^+$ and constitute the already known ncMono fraction (Schakel et al., 1999), and (ii) a CD16$^-$ pre-DC contamination associating with the areas within the UMAP defined as cDC2 and pre-DC by traditional investigator-informed gating (**Figure 4I**). Taken together, our new consensus map clarifies that map 1 DC4 is comprised of CD16$^+$ ncMono and pre-DCs, rather than a phenotypically defined novel cell type within the human mononuclear myeloid cell compartment (Calzetti et al., 2018).

***Backmapping identifies mono 4 as bona fide CD56$^{dim}$ NK cells***

Next, we wanted to use our new consensus map to define the monocyte population structure. In particular, we aimed to consolidate the newly defined map 1 subtype structure with the four monocyte subtypes (mono 1-4) (Villani et al., 2017) in light of the traditional view of only three phenotypically different monocyte subsets based on CD14 and CD16 expression (Ziegler-Heitbrock et al., 2010). As a first step, we recapitulated the map 1 flow cytometry sorting strategy and overlaid the cellular contents of this gate onto our novel flow cytometry-derived UMAP topology (**Figure 5A**). Within the CD16$^+$ compartment of map 1, two different monocyte populations (mono 2 and 4) were defined by phenotypical and transcriptional differences. Mapping the CD16$^+$CD14$^-$ compartment of map 1 onto the new UMAP topology indeed showed that it is composed of two transcriptionally different cellular entities, one mapping into the HLA-DR$^-$ space of the flow cytometry-derived UMAP topology and one mapping to the location occupied by ncMono in an investigator-driven gating approach and named mono 2 in map 1 (**Figure 1E, 2A, S1**).

To understand the identity of the cells mapping to the HLA-DR$^{-}$ cell space, we utilized the transcriptomic marker genes derived from map 1 mono 4, as mono 2 mapped to the ncMono space. We then interrogated a publicly available database of population-based proteome fingerprints (Rieckmann et al., 2017) from a variety of blood-borne immune cells for the mono 4 signature (**Figure 5B**). Here, we found high expression of mono 4-related proteins in NK cell subsets, including a CD56$^{dim/neg}$ subset (NK$^{dim}$). To validate these findings, we generated a transcriptomic NK-cell signature based on previous knowledge (Costanzo et al., 2018; Liberzon et al., 2011; Rieckmann et al., 2017; Subramanian et al., 2005) and calculated the signature enrichment scores across all monocyte subsets defined in map 1 (**Figure 5C-D**). This calculation revealed that the mono 4 subset was significantly enriched in NK-cell-specific transcripts. Subsequently, we integrated the original monocyte map 1 single-cell transcriptome data (mono 1-4) into an external dataset of 33,148 PBMCs (short: 33k-PBMC dataset, https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc33k) and performed dimensionality reduction of the corresponding monocyte and NK-cell-related cellular spaces using UMAP (**Figure 5E-F**). We termed this approach 'backmapping', where we utilized an unrelated single-cell data set derived from the same tissue origin. In a next step, we annotated the integrated map 1 specific monocyte subsets within the combined UMAP topology according to the terminology used in map 1, to understand where these cells would associate in the context of an unbiased assessment of the complete mononuclear PBMC fraction. This analysis showed that the mono 1-3 subsets mapped to the topological UMAP space initially assigned to monocytes, whereas mono 4 mapped to the topological UMAP space of NK cells, further supporting the hypothesis that mono 4 are NK cells. Overlaying of the NK cell signature onto the original map 1 also revealed strong enrichment in the mono 4 cluster (**Figure 5G**).

We then modified our combined flow cytometry panel to specifically verify NK cell contamination within the map 1-defined flow cytometry CD16$^{+}$ monocyte cell space (**Figure 5H, S5A**). Specifically, we removed CD56 from the lineage to track the expression of this NK-cell marker

separately and added the granulocyte marker CD66b, the lymphoid marker CD7 and the NK cell markers NKp46, CD160 and CD107 (**Figure S5A**). We then examined CD16 and CD56 expression within the $CD14^-$ compartment of the PBMC $CD45^+Lin^-$ fraction and generated a reference UMAP topology of the $CD45^+Lin^-$ cell space (**Figure 5H**, **S5A-B**). This analysis identified seven cell populations based on CD16 and CD56 expression levels (**Figure S5B-C**). Two populations (turquoise and pink) displayed high CD16 and SSC and no (light blue) to mid (pink) CD56 expression, with granulocytic forward and sideward scatter characteristics identifying them as granulocyte contaminants (**Figure S5C**).

Next, we focused our analysis on the $CD16^+$ compartment of this cell space (green and purple gates). Two populations were identified as $CD16^+$, in which one $CD16^+CD56^+$ population (purple) matched the surface phenotype of classical $CD56^+$ NK-cells that are normally dismissed by including CD56 in the lineage panel of map 1. To determine the identity and heterogeneity of the remaining $CD16^+CD56^-$ cell compartment (green, orange, yellow, grey gates, **Figure S5B**), we mapped this compartment back to a UMAP topology of either the $Lin^-CD16^+$, $Lin^-CD56^-CD16^+$ or the $Lin^-CD56^-CD16^+HLA\text{-}DR^+$ cell space, to represent a stepwise cleanup of non-monocytic $CD16^+$ cells (**Figure 5H**). This analysis showed that if the totality of the $Lin^-CD16^+$ compartment is mapped back onto the $Lin^-$ UMAP topology (**Figure 5H**, pink overlay, most left panel), NK cells ($CD56^+$), monocytes ($CD56^-CD16^{+/-}$) and granulocyte fractions ($CD16^{high}$) are included in this cellular compartment. When excluding CD56 in the UMAP topology, classical $CD56^+$ NK cells are excluded; however, within the $CD16^+$ gate a $CD56^-$ population became apparent that mapped to the UMAP space previously associated with classical NK cells (**Figure 5H**, pink overlay, mid panel). Another CD16+$CD56^-$ population mapped to the topological UMAP space occupied by ncMono, as defined by their high expression of HLA-DR and CD11c and no expression of classical NK-cell markers, including CD56, CD7, CD160 and Nkp46 (**Figure S5C**). We next excluded HLA-$DR^-$ cells and mapped the remaining $CD16^+$ cells onto the UMAP topology. This step revealed that including a positive HLA-DR threshold successfully removed mono 4 / NK cells (pink overlay,

**Figure 5H**, right panel). This subclass of NK cells is not easily distinguishable from monocytes, as also evidenced by their very similar morphology (**Figure S5D**). Taken together, we identified the map 1 mono 4 subset as HLA-DR$^-$CD16$^+$CD56$^-$ NK-cells, intruding into the CD16$^+$CD14$^-$Lin$^-$ CD45$^+$ map 1 monocyte sorting gate that was performed without HLA-DR gating stratification according to map 1. Collectively, within our new consensus map, we define the borders between the myeloid compartment and the NK cell compartment and re-stablish a structure of three monocyte subsets in peripheral blood.

***PBMC derived monocyte subsets form a transcriptional continuum during homeostasis***

*De novo* clustering (**Figure 2A, 6A**) did not reveal intMono as a transcriptionally distinct cluster; rather, they were identified as forming part of clusters eight and nine (**Figure 6A, 6B, S6A, S6B**). Pseudo-time analysis of the scRNA-seq data, however, placed intMono in between cMono and ncMono (**Figure 6C**). The visualization of genes changing over the pseudo-time depicts a gradual decrease in expression of cMono marker genes (CD14, etc) and an increase of ncMono marker gene expression (CD) along the trajectory (**Figure 6D**). This was further corroborated by plotting CD14 and CD16 mRNA expression of single cells within the three monocyte subsets (**Figure 6E**). Therefore, these analyses clearly corroborate an existing transcriptional continuum of monocytes within human PBMC and reveal the transcriptional identity of intMono during homeostasis.

***Backmapping integrates legacy datasets and enhances cell type resolution creating novel consensus maps***

To evaluate our findings and put our new consensus map into the framework of data-driven maps of the complete human PMBC compartment, we combined our new scRNA-seq data with three independent PBMC datasets and performed backmapping and cell type prediction (**Figure 7**, **S7**). This approach permitted a detailed annotation of previously undefined cellular identities within the

external PBMC-derived datasets. *De novo* analysis of a combined data set (**Figure S7A**) consisting of a publicly available dataset of approximately 33.000 cells (33k-PBMC) and our new scRNA-seq dataset revealed NK, $CD8^+$, $CD4^+$ and B cells alongside megakaryocytes, $CD16^+$ and $CD14^+$ monocytes, cDC1, DC2, DC3, pDCs and $CD34^+$ progenitor cells (**Figure 7A**). When using the new scRNA-seq-based consensus map as a reference to predict cell annotations in the data-driven PBMC dataset by a nearest neighbor classifier (Kiselev et al., 2018), we obtained similar but not identical results (**Figure 7B**, left panel). We therefore applied the backmapping approach by projecting the cluster identities of our new scRNA-seq dataset onto the UMAP dimensionality reduction of the combined dataset. Backmapping revealed commonalities between all cell types of the myeloid cell compartment including previously unidentified pre-DCs, DC2 and DC3 clusters within the 33-K PBMC dataset (**Figure 7B**) as well as NK cells, $CD34^+$ and plasma cells derived from both datasets (**Figure S7B**). We next applied the above outlined approach including backmapping to a larger data set provided by the Human Cell Atlas (HCA) with approximately 255.000 cells (**Figure 7C, 7D, 7E**). In a first step, we categorized all major subtypes (**Figure 7C**) and predicted cells associated with the mono 4 population of map 1 and found that all cells fell within the NK cell cluster (**Figure S7C, S7D**), further supporting that cells of the mono 4-subset are *bona fide* NK cells. Next, we reduced the datasets to clusters that were part of the myeloid cell compartment (**Figure 7D, S7C**) and again performed the backmapping and prediction approaches. Because of the solely cluster-driven reduction of the dataset, we observed some lymphoid cells in the reduced dataset, which might derive from misclustered cells (**Figure 7D**). Nevertheless, this approach allowed us to identify smaller myeloid cell populations within the larger dataset (**Figure 7E**). Finally, we used a third PBMC dataset based on a targeted scRNA-seq approach, which also allowed us to better define subsets in the unbiased dataset (**Figure S7E, S7F, S7G**) making this overall approach independent of single-cell technology and dataset size.

Collectively, we demonstrate the value of an iterative, rule-based, data-informed approach based on previously existing maps to integrate additional information layers into new consensus maps. Backmapping to whole data-driven tissue maps and providing a connection to previous knowledge are important steps to derive the next iterations of maps that finally serve as the entry point for further iterations.

## Discussion

Consensus maps are an important instrument within an iterative process of producing cellular maps of all organs and tissues in different species, including humans. As within other scientific disciplines, such as geography or astronomy, the maps generated in the life sciences require much iteration to allow for the integration of new content. By combining single-cell transcriptomics with index sorting, and multi-color flow cytometry and applying simple but very effective computational strategies, such as 'backmapping' to cellular maps generated in a purely data-driven fashion, we have generated a new consensus map of the myeloid cell compartment including monocytes, DCs and their precursors (**Figure 7, S7**). Because we propose to include prior knowledge in the respective scientific field into the algorithm for generating such consensus maps, we define the overall strategy as being 'data-informed', combining prior knowledge and data-driven technologies including single-cell omics.

The two previous maps based on single-cell RNA-seq used in our approach as well as a phenotypic analysis of the human blood and tissue myeloid cells were developed to improve our understanding of myeloid cell heterogeneity (Alcantara-Hernandez et al., 2017; See et al., 2017; Villani et al., 2017). Yet there were shortcomings to these maps, which we have overcome in our new consensus map. First, map 2 only identified one cDC2 subset, whereas map 1 and our new consensus map defined two subsets. Furthermore, we established a proximity between DC3 and cMono, which has been further dissected by Dutertre *et al.* (Dutertre et al., 2019), thus already

providing the next iteration of this particular subspace in the myeloid cell map of human peripheral blood. Second, map 1 identified a novel monocyte subset named mono 4. Using backmapping, we reveal that mono 4 are $CD56^{dim}$ NK cells and are not related to monocytes, supporting the current definition of three major monocyte subsets (classical $CD14^{+}$, non-classical $CD16^{+}$, and intermediate double positive monocytes), consisting of two transcriptionally distinct entities and a continuum of intermediate, double-positive monocytes between them (**Figure 6**). This finding is further supported by the changes in expression of the CD14 and CD16 cell-surface markers and results derived from genetic mouse models showing that $Ly6c^{hi}$ monocytes (murine equivalents of classical monocytes) can transition into $Ly6c^{low}$ monocytes (murine equivalents of non-classical monocytes) with only a few cells detectable in the transitory state (Mildner et al., 2017). Third, we could clearly define AS-DCs (DC5) from map 1 as pre-DCs within the consensus map, consistent with their functional definition in map 2 (See et al., 2017). Together with the complete overlap of the three differentiated DC populations between the original maps, these results reassure the validity of single-cell transcriptomic analyses.

We define backmapping as an integral component of the strategy to define novel consensus maps. Here, we use cellular maps derived from tissues – in this case peripheral blood - without prior experimental enrichment of certain cell types. This relatively simple computational approach allows to unequivocally overlay cell subsets from different maps onto a common cell space. As exemplified here for the monocyte / NK cell space, we postulate that potential conflicts for new cell types in other organs can be resolved in a similar fashion.

Collectively, we report on a new consensus map of the myeloid cell compartment in human blood, which was built on two previously introduced maps (See et al., 2017; Villani et al., 2017). The myeloid cell compartment is of particular interest due to its intrinsic heterogeneity, its involvement in many if not all major tissues and organs, and its prime involvement in almost any major disease (Bassler et al., 2019). It is therefore of utmost importance to establish a precise baseline during homeostasis as provided here by our new consensus map, to allow for a better understanding of

any deviations of the myeloid cell compartment during stress, pathophysiological conditions and diseases. Furthermore, the dynamic processes of myelopoiesis during homeostasis but even more so during inflammatory conditions requires precise mapping of cellular identities as a prerequisite to identify targets for precise therapeutic intervention (Dick et al., 2019; Schultze, 2019; Schultze et al., 2019). Furthermore, the necessity to continuously iterate the process of improving the consensus maps is nicely illustrated by the accompanying manuscript by Dutertre *et al.* (Dutertre et al., 2019), further defining the cellular relationship of DC2/3 and monocytes. As many institutions world-wide continue to generate cellular maps, consensus maps will become an increasingly important instrument to reconcile and integrate information. Our approach provides a guide to integrate and value legacy datasets together with newly generated single-cell omics data and build new iterations of consensus maps applicable to any other tissue. These maps can also be adapted to include further technological advancements. With the continuation of technical advances, we anticipate that consensus map building will become a major task within our efforts to create complete cellular atlases for the major species.

**Acknowledgments:**

*Non-author contributions:* We thank Jessica Tamanini for critical review and editing of the manuscript.

**Funding:** This work was supported by the German Research Foundation to JLS (GRK 2168, INST 217/577-1, EXC2151/1), by the HGF grant sparse2big to JLS, the FASTGenomics grant of the German Federal Ministry for Economic Affairs and Energy to JLS and the EU project SYSCID under grant number 733100. A.S. is supported by an Emmy Noether fellowship of the German Research foundation (SCHL2116/1-1, EXC2151/1). J.L.S. and A.S. are members of the Excellence Cluster ImmunoSensation$^2$. F.G is an EMBO YIP awardee and is supported by Singapore Immunology Network (SIgN) and Shanghai Institute of Immunology core funding. VB is supported by Wellcome Trust Intermediate Fellowship (101155/Z/13/Z). MC is supported by CRUK (C30484/A21025) and NIHR Newcastle Biomedical Research Centre at Newcastle upon Tyne Hospitals.

**Author contributions:** Conceptualization, P.G, B.C., A.S. and J.L.S.; Methodology, P.G, B.C., A.S. and J.L.S.; Software, P.G, B.C, K.B., M.B. and K.H., Investigation, P.G, B.C, K.B., K.H., C.A.D. and V.B., Resources, A.S. and J.L.S., Writing – Original Draft, P.G, B.C., A.S. and J.L.S.; Writing – Review & Editing, P.G, B.C, K.B., E.N., M.C., F.G., A.S. and J.L.S.; Visualization, P.G, B.C. and K.B.; Supervision, A.S. and J.L.S.; Project Administration, A.S. and J.L.S.; Funding Acquisition, E.N., M.C., F.G., A.S. and J.L.S..

**Declaration of interests:** The authors declare that there are no competing interests.

**Data and materials availability:** Processed and raw scRNA-seq datasets are available through the Gene Expression Omnibus (GSE126422). In addition, we provide an interactive web tool to visualize the single-cell RNA-Seq data together with the flow cytometry data at https://paguen.shinyapps.io/DC_MONO/.

## Figure Legends

**Figure 1. Generating a new consensus map of the mononuclear myeloid cell compartment in human peripheral blood. (A)** Workflow to generate a new consensus map of the human mononuclear myeloid cell compartment. **(B)** Visualization of ~1.4 mio. live $CD45^{+}Lin(CD3, CD19, CD20, CD56)^{-}$ cells after UMAP dimensionality reduction of the flow cytometry panel introduced in A (left panel), mononuclear myeloid cell compartment (second panel), overlay of index-sorted cells (third panel), UMAP topology of the index-sorted cells based on the single-cell transcriptome data (most right panel, see also **Figure 2**). Grey areas in the third panel represent the $CD45^{+}Lin^{-}$ cell space. **(C)** Phenograph clustering of the flow cytometry data projected onto the FACS-based UMAP topology. **(D)** Color-coded visualization of markers used to define the mononuclear myeloid cell compartment. **(E)** Overlay of the cell gating strategies according to maps 1 (Villani et al., 2017) and 2 (See et al., 2017). See also Figure S1.

**Figure 2. Index-sorted scRNA-seq dataset of the myeloid cell compartment in human blood. (A)** *De novo* clustering of the 2,509 index-sorted cells onto the scRNA-based UMAP topology (left panel) and cluster projection onto the FACS-based UMAP topology (right panel, grey background: complete $CD45^{+}Lin^{-}$ cell space). **(B)** Heatmap of 10 most significant marker genes for each of the 11 clusters identified and visualized in **Figure 2A**. **(C)** Overlay of cell types defined for maps 1 (left panel) and 2 (right panel) onto the scRNA-based UMAP topology of the new consensus map.

**Figure 3. Harmonizing the DC space within the mononuclear myeloid cell compartment.** (**A**) Overlay of signatures derived from map 2 DCs onto the new scRNA-seq-based UMAP topology consensus map. (**B**) UMAP topology based on map 1 single-cell transcriptomes of map 1 DC1-6 cells and overlay of signatures derived from map 2 DCs. (**C**) Overlay of signatures derived from map 1 DC1-6 cells onto the new scRNA-seq-based UMAP topology consensus map. (**D**) UMAP topology based on flow cytometry data derived from ~1.4 mio live $CD45^+Lin(CD3, CD19, CD20, CD56)^-$ cells (see **Figure 1A**) and separate overlays of cluster 26 defined by Phenograph (see **Figure 1C**), map 1 DC5, map 2 pre-DC, and scRNA-seq-based cluster seven representing transcriptomic progenitor DC signatures (see **Figure 2A**). **(E)** Enrichment of map 2 defined pDC, cDC1, cDC2 and pre-DC signatures in the map 1 DC1-6 subsets. **(F)** Heatmap of the average expression values of hallmark genes defined for map 1 DC1-6 subsets in both map 1 DC1-6 as well as map 2 DCs subsets.

**Figure 4. Integrating newly defined DN-DCs into the new consensus space of the myeloid cell compartment.** (**A**) Recapitulation of the map 1 gating strategy to identify a putative DC4 subset within $CD11c^+$ DN-DCs and visualization of CD16 expression. (**B**) $CD16^+$ and $CD16^-$ DN-DC mapping (as in A) onto the flow cytometry derived UMAP topology (left panel, see **Figure 1B, 3A**) and scRNA-seq data (right panel). (**C**) Relationship analysis of $CD16^+$ or $CD16^-$ DC-DN cells (see **Figure** 3A) and their corresponding annotation according to map 2, FACS Phenograph clustering and scRNA-seq data. (**D**) Pearson correlation matrix of all cell types defined in map 1. (**E**) Signature enrichment analysis of map 1 mono 2 signature in all other map 1-defined cell types. (**F**) Enrichment of the DC4 (DN-DC) signature visualized on the scRNA-seq data derived UMAP topology. (**G**) Gating strategy to define SLAN expression on the cell population defined as DN-DC based on CD16 expression. (**H**) Visualization of DN-DC cells in the complete $CD45^+Lin^-HLA\text{-}DR^+$ UMAP topology (grey). (**I**) Mapping of the phenotypic information of cell populations onto the new UMAP topology.

**Figure 5. Relationship of the previously introduced fourth monocyte subset (mono 4) in context of the new consensus map of the myeloid cell compartment.** (**A**) Mapping flow cytometric- and scRNA-seq-defined map 1 mono 2 and 4 onto the flow cytometry based UMAP topology (see **Figure 1B**, **E**, **2A**). **(B)** Heatmap of the protein expression pattern for mono 4 signatures genes derived from mass-spectrometric data of FACS-sorted population as described by Rieckmann et al. (Rieckmann et al., 2017). (**C**) Enrichment of a NK cell signature in map 1 mono 1-4 subsets. (**D**) Heatmap of NK cell hallmark genes within the map 1 defined cell subsets. (**E**) Backmapping by overlaying map 1 mono 1-4 cells onto the 33k-PBMC scRNA-seq dataset. Only monocytes and NK cells are shown. (**F**) Visualization of the percentage of cells that are aligned with either monocytes or NK cells derived from the unrelated 33k-PBMC scRNA-seq dataset. (**G**) UMAP topology of scRNA-seq data derived from the map1 DC and mono subsets (left panel) and overlay of the NK cell signature onto this UMAP topology. (**H**) Top panels: classical gating strategy and stepwise cleanup of $CD45^+$ cells based on lineage (CD3/CD19/CD20) marker expression, then based on CD56 expression followed by HLA-DR expression, left to right. Middle panels: UMAP topology derived from the respective cell populations marked within the corresponding top panels. Mono/DC by HLA-DR expression, green; NK by CD7 expression, violet; and granulocytes by CCR3 or CD66b expression, orange. Bottom panels: Effect of cleanup as shown in top panels on the $CD16^+$ $CD14^-$ cell population.

**Figure 6. Focused analysis of the monocyte compartment (A)** Overlay of the the cluster 8 and 9 defined by de novo clustering of the scRNA-Seq data onto the scRNA-UMAP topology of the new consensus map. **(B)** Bar chart showing the original FACS annotation of cells derived from cluster 8 or 9 following the sorting scheme of map 1 or map 2, respectively. **(C)** Trajectory analysis of the monocyte subset containing cells from cluster 8 and 9. Monocle-based UMAP dimensionality reduction overlaid with cell estimated pseudo-time (left panel) and the FACS

annotations derived from map 1 (center panel) or map 2 (right panel). **(D)** Transcriptional changes of genes that are considered as differentially expressed along the inferred trajectory. Heatmap shows scaled gene expression changing over the pseudo-time (x-axis, early to late). Selected marker genes of cMono and ncMono are highlighted. **(E)** Expression of CD14 and CD16 in relation to the estimated pseudo-time of cells. Cells are colored by their FACS annotation from map 1 or map 2.

**Figure 7. Backmapping strategy combining the new scRNA-seq data and different PBMC datasets. (A)** Annotation of cell types within the combined dataset (33-K PBMC and new scRNA-seq dataset). **(B)** Graphs in the left panel predict cell labels from the 33-K PBMC dataset by using the transcriptome information from the new scRNA-seq dataset. Graphs in the right panel show the visualization of the cells from the new scRNA-seq dataset after integration with the 33-k PBMC dataset. **(C)** UMAP dimensionality reduction of around 260.000 human cord blood cells and cell annotation based on markers obtained from the unrelated 33k-PBMC dataset (**Figure 7A-B**). **(D)** Reduction of the HCA dataset to cells, which were found within clusters associated with monocytes or dendritic cells. **(E)** Graphs within the left panel show the prediction scores calculated for the respective cell types of the new scRNA-seq data. Graphs in the right panel show the visualization of the cells from the new scRNA-seq data after “anchoring” together with the HCA dataset.

**Supplementary Figure Legends**

**Figure S1. Classical flow cytometry gating strategies applied for the generation of the legacy maps 1 and 2. (A)** Common part of the flow cytometry gating strategy applied for the generation of the legacy maps 1 and 2. **(B)** Map 1-specific part of the flow cytometry gating strategy with the resulting cell subsets (colored boxes). **(C)** Map 2-specific part of the flow cytometry gating strategy with the resulting cell subsets (colored boxes). Some cell types shown here are based on a priori definitions (e.g. monocytes) and were not part of the transcriptionally defined cells in the legacy map 2.

**Figure S2. Quality control criteria for the scRNA-seq data established for the new consensus map of the myeloid cell compartment. (A)** Visualization of the number of reads for all 8 384-well plates analyzed within this project. Violin plot of the number of genes observed to be present within all cells measured. **(B)** Visualization of the number of reads (left panel), the number of genes (middle panel), and the percent of aligned reads (right panel) as a violin plot for each of the 8 384-well plates individually. **(C)** Comparison and visualization of cell distribution across all identified clusters in relationship to the 8 384-well plates utilized within this experiment. **(D)** Mapping of single-cell information concerning the total number of reads, unaligned reads, number of genes, and number of transcripts onto the UMAP topology of the final consensus map based on scRNA-seq data. **(E)** Cluster relationship analysis **(F)** Distribution of cells within each of the identified cluster in relation to the 8 384-well plates used in this study.

**Figure S3. Overlay of phenotypic and transcriptomic data onto the new consensus map of the myeloid cell compartment. (A)** Visualization of cell surface markers detected by index sorting on the 2,509 cells which were used to define the UMAP topology of the index-sorted cells based on the single cell transcriptome data. **(B)** Visualization of gene-level expression of the

respective cell surface markers within the 2.509 cells which were used to define the UMAP topology of the index-sorted cells based on the single cell transcriptome data.

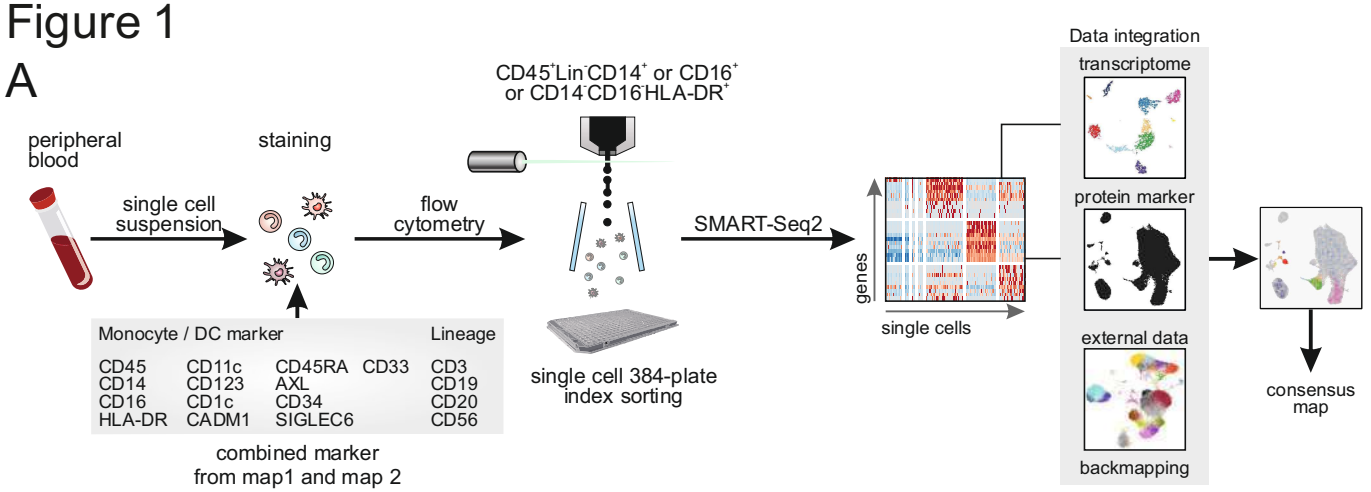
**Figure S4. Defining the relationship of map 1 DN-DC within the new consensus map of the myeloid cell compartment. (A)** Enrichment of map 1 mono 1,3,4 signatures in map 1 mono1-4 and DC1-6 subsets. **(B**) Violin plots for normalized gene expression of FCGR3A, TCF7L2, RHOC, MTSS1 in map 1 mono1-4 and DC1-6 subsets. **(C)** Overlay and visualization of map 1 mono 1-4 subset gene signature enrichment on the UMAP topology based on scRNA-seq of 2.509 index-sorted cells (see **Figure 2A**).

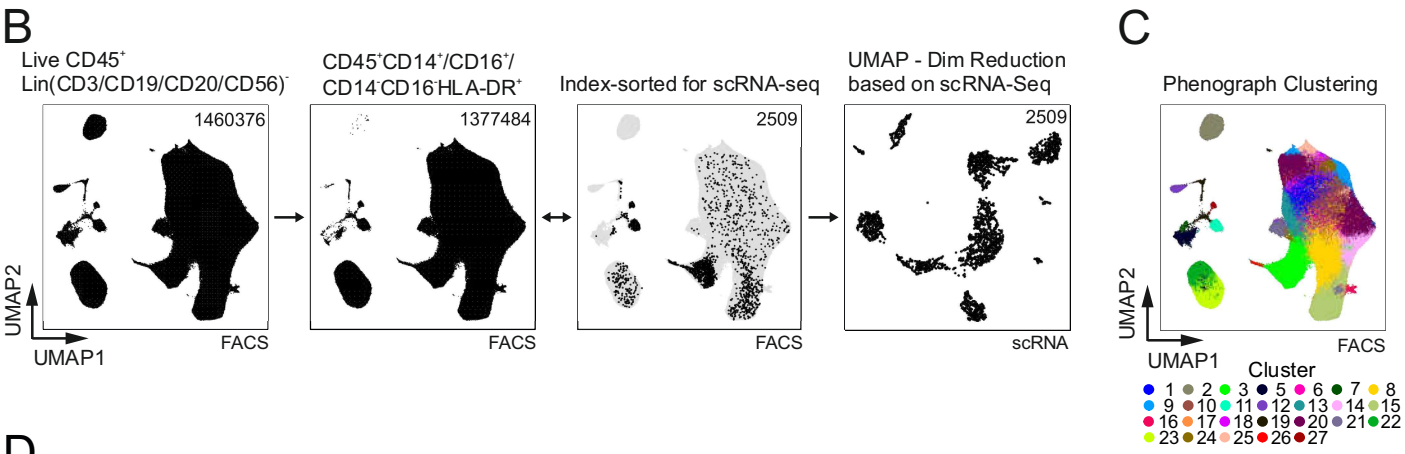
**Figure S5.** The relationship between the myeloid and the NK cell compartment in human peripheral blood. **(A)** Schematic representation of the development of a new focus strategy (panel adjustment) to define the relationship between the myeloid and the NK cell compartment in human peripheral blood. **(B)** Classical gating strategy to determine those cell populations that need to be placed either into the myeloid or the NK cell compartment followed by the development and visualization of the UMAP topology of both cellular compartments. **(C)** Color-coded visualization of markers used to define the mononuclear myeloid and NK cell compartments on the flow cytometry data-based UMAP topology. **(D)** Cytospins of cells sorted according to the gating strategy depicted in **Figure S5B**.

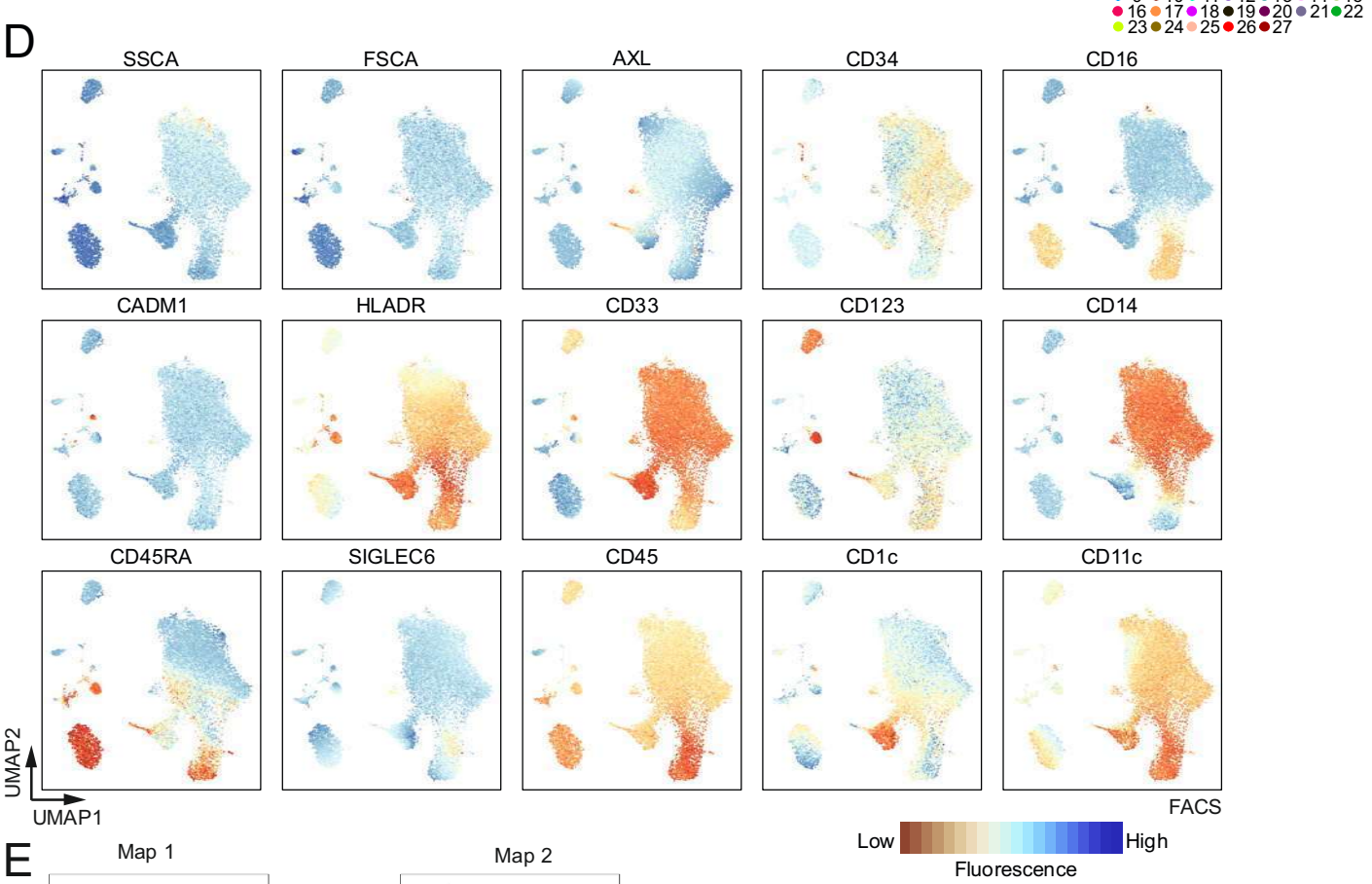
**Figure S6. Focused analysis of the monocyte compartment (A)** UMAP dimensionality reduction calculated on the focused subset of cluster 8 and 9 containing the monocyte populations. Overlaid are the cell annotations from the clustering and **(B)** the FACS annotation derived from Map 1 (upper panel) and Map 2 (lower panel)**.**

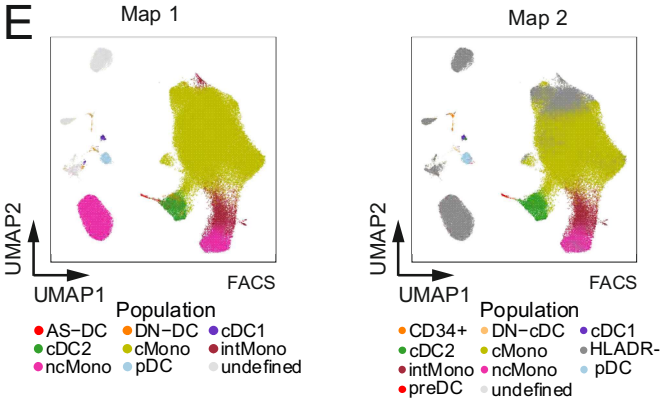
**Figure S7. Integration of the new scRNA-seq dataset with three individual PBMC datasets.** **(A)** UMAP dimensionality reduction based on the combined dataset of the new mononuclear myeloid scRNA-seq data and the external 33k-PBMC dataset. Cells originating from the new scRNA-seq data are colored black and cells from the external PBMC dataset are colored grey. **(B)** Overlay of cell annotations of cells from NK cells, $CD34^+$, plasma cells, $CD4^+$ T cells, $CD8^+$ T cells, and B cells from the 33-k PBMC dataset. **(C)** Clustering of the HCA dataset. **(D)** Prediction and backmapping of NK cells from the novel scRNA-seq dataset onto the complete HCA dataset. Left UMAP graph shows the computed prediction score for NK cells of the HCA dataset using the new scRNA-seq consensus map information. Red color indicates highest prediction score. Right UMAP graph shows the location of NK cells from the new scRNA-seq data within the combined dataset. **(E)** UMAP dimensionality reduction based on the combined dataset of the new mononuclear myeloid scRNA-seq data-based consensus map and a PBMC dataset processed by the BD Rhapsody technology. Cells originating from the new scRNA-seq consensus map are colored black and cells from the Rhapsody PBMC dataset are colored grey. **(F)** Overlay of cell annotations of the DC subsets from the new scRNA-seq data-based consensus map on the combined UMAP. **(G)** Overlay of cell annotations of NK cells and monocyte subsets identified in the new scRNA-seq data-based consensus map (top panel) and the respective cell annotations of NK cells, $CD14^+$ and $CD16^+$ monocytes from the Rhapsody PBMC dataset. **(H)** Overlay of cell annotations of $CD4^+$ T cells, $CD8^+$ T cells and B cells from the external PBMC dataset.

# Figure 1

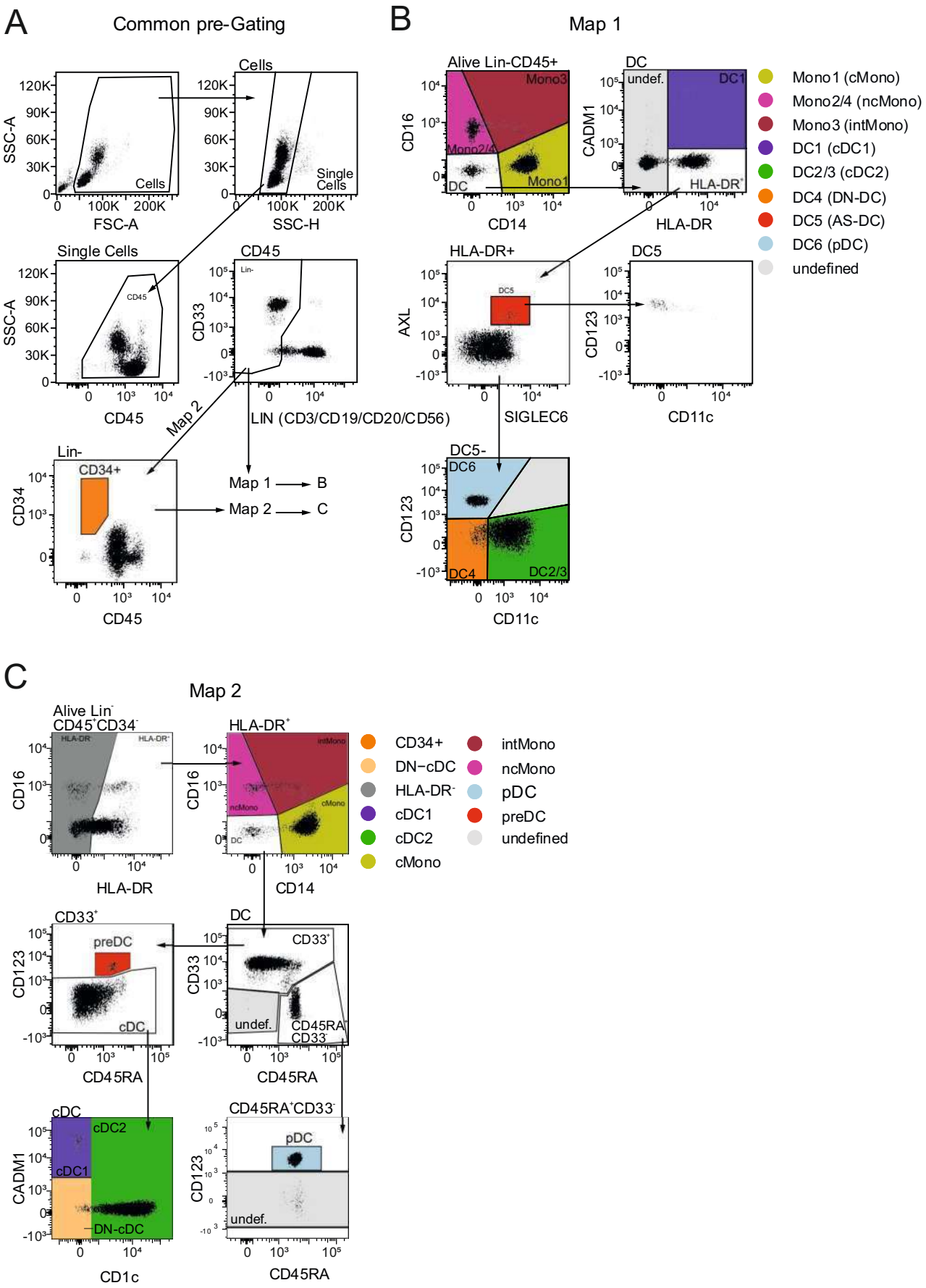

A
peripheral blood
staining
single cell suspension
flow cytometry
CD45+Lin-CD14+ or CD16+ or CD14-CD16-HLA-DR+
SMART-Seq2
genes
single cells
Data integration
transcriptome
protein marker
external data
backmapping
consensus map
single cell 384-plate index sorting
Monocyte / DC marker
Lineage
CD45 CD14 CD16 HLA-DR
CD11c CD123 CD1c CADM1
CD45RA AXL CD34 SIGLEC6
CD33
CD3 CD19 CD20 CD56
combined marker from map1 and map 2


B
Live CD45+ Lin(CD3/CD19/CD20/CD56)-
CD45+CD14+/CD16+/ CD14-CD16-HLA-DR+
Index-sorted for scRNA-seq
UMAP - Dim Reduction based on scRNA-Seq
1460376
1377484
2509
2509
FACS
FACS
FACS
scRNA
UMAP2
UMAP1
C
Phenograph Clustering
UMAP2
UMAP1
FACS
Cluster
1 2 3 5 6 7 8
9 10 11 12 13 14 15
16 17 18 19 20 21 22
23 24 25 26 27

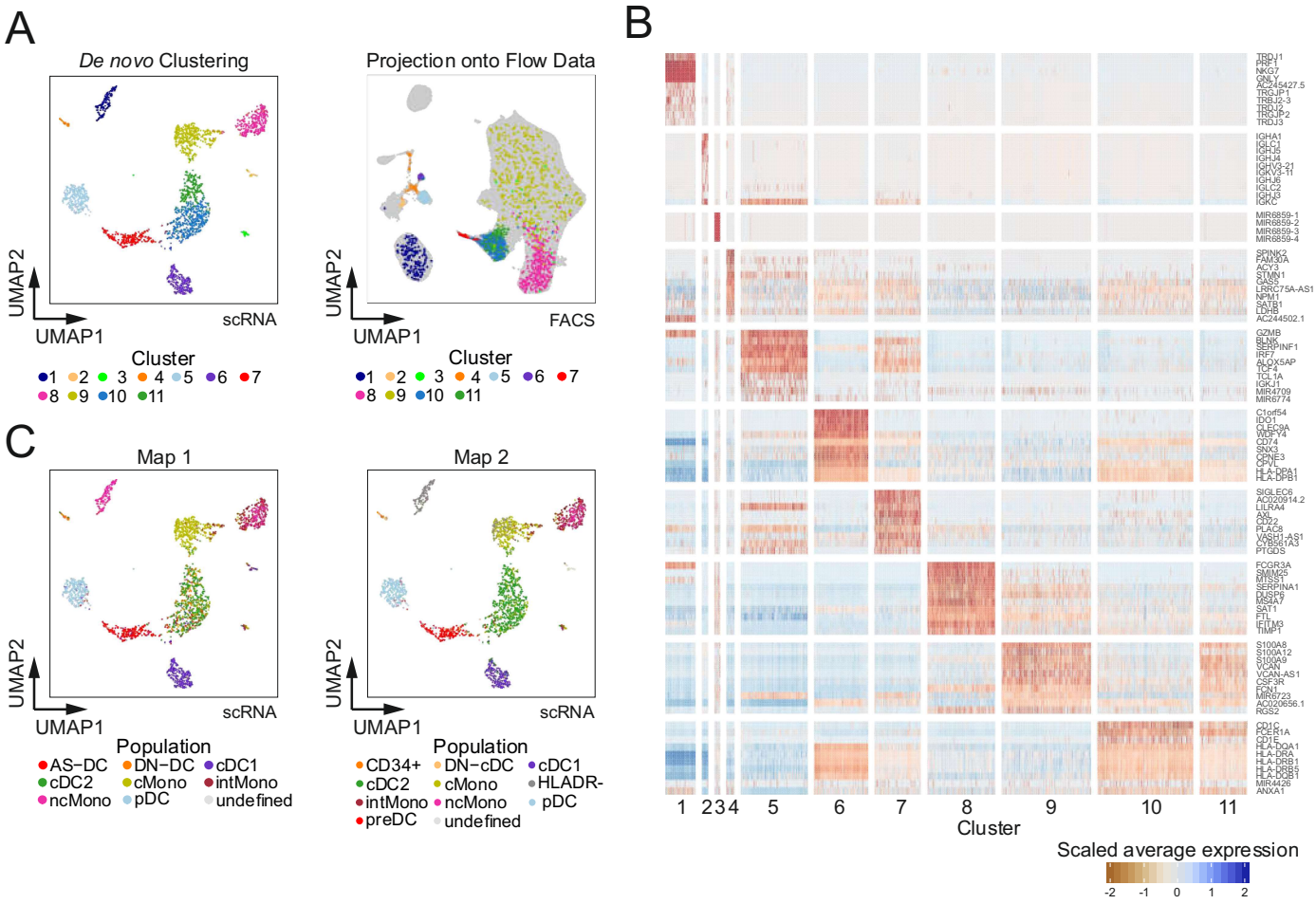

D
SSCA
FSCA
AXL
CD34
CD16
CADM1
HLADR
CD33
CD123
CD14
CD45RA
SIGLEC6
CD45
CD1c
CD11c
UMAP2
UMAP1
FACS
Low
High
Fluorescence


E
Map 1
Map 2
UMAP2
UMAP1
FACS
Population
AS-DC DN-DC cDC1
cDC2 cMono intMono
ncMono pDC undefined
CD34+ DN-cDC cDC1
cDC2 cMono HLADR-
intMono ncMono pDC
preDC undefined

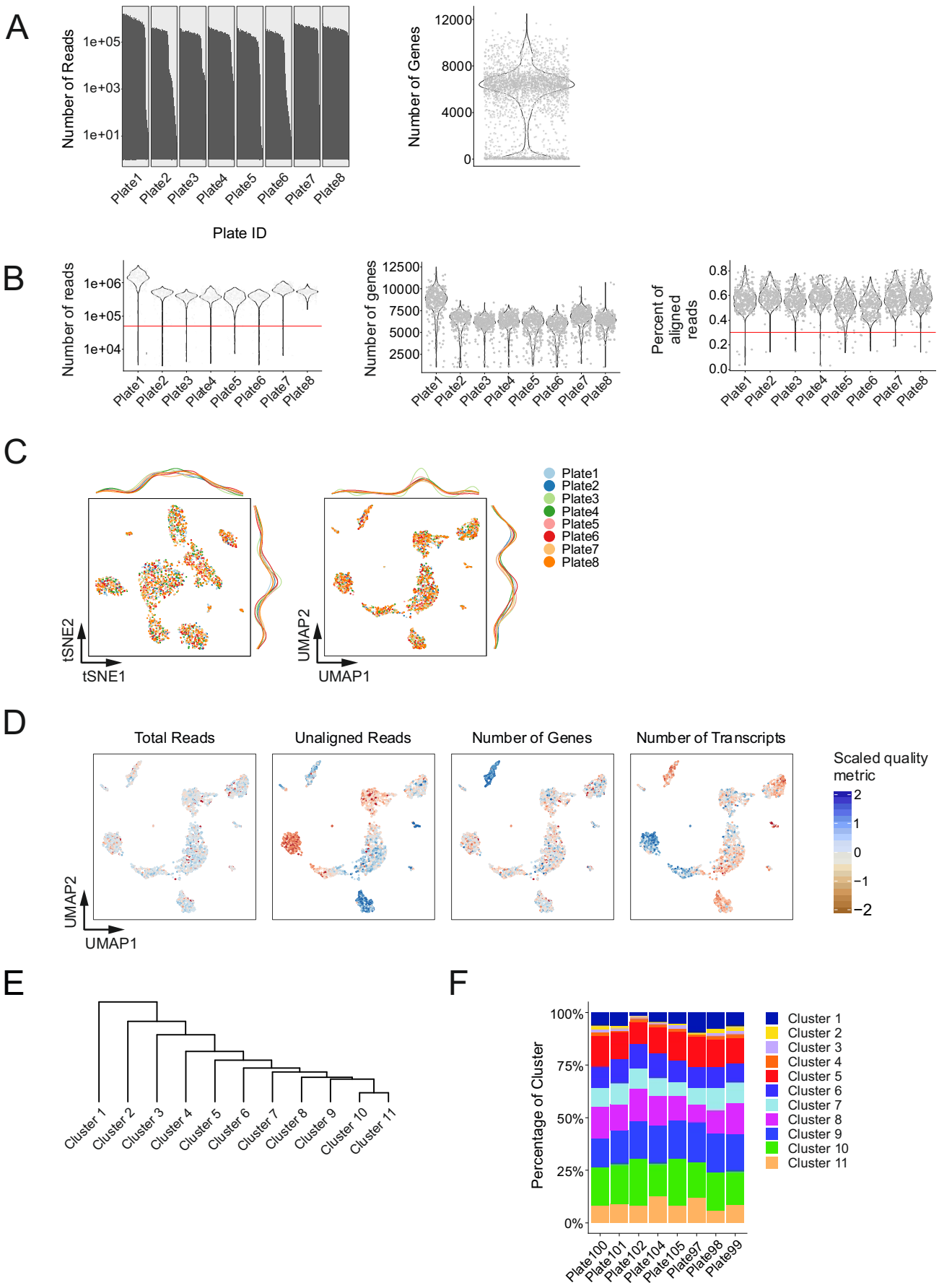
# Figure S1


A
Common pre-Gating
Cells
Single Cells
Single Cells
CD45
Lin-
CD34+
Map 1 → B
Map 2 → C
Map 2
LIN (CD3/CD19/CD20/CD56)
B
Map 1
Alive Lin-CD45+
Mono3
Mono2/4
Mono1
DC
DC
undef.
DC1
HLA-DR+
HLA-DR+
DC5
DC5-
DC6
DC4
DC2/3
Mono1 (cMono)
Mono2/4 (ncMono)
Mono3 (intMono)
DC1 (cDC1)
DC2/3 (cDC2)
DC4 (DN-DC)
DC5 (AS-DC)
DC6 (pDC)
undefined
C
Map 2
Alive Lin-
CD45+CD34-
HLA-DR+
CD34+
DN−cDC
HLA-DR-
cDC1
cDC2
cMono
intMono
ncMono
pDC
preDC
undefined
CD33+
preDC
cDC
DC
CD33+
undef.
CD45RA+
CD33-
cDC
cDC2
cDC1
DN-cDC
CD45RA+CD33-
pDC
undef.

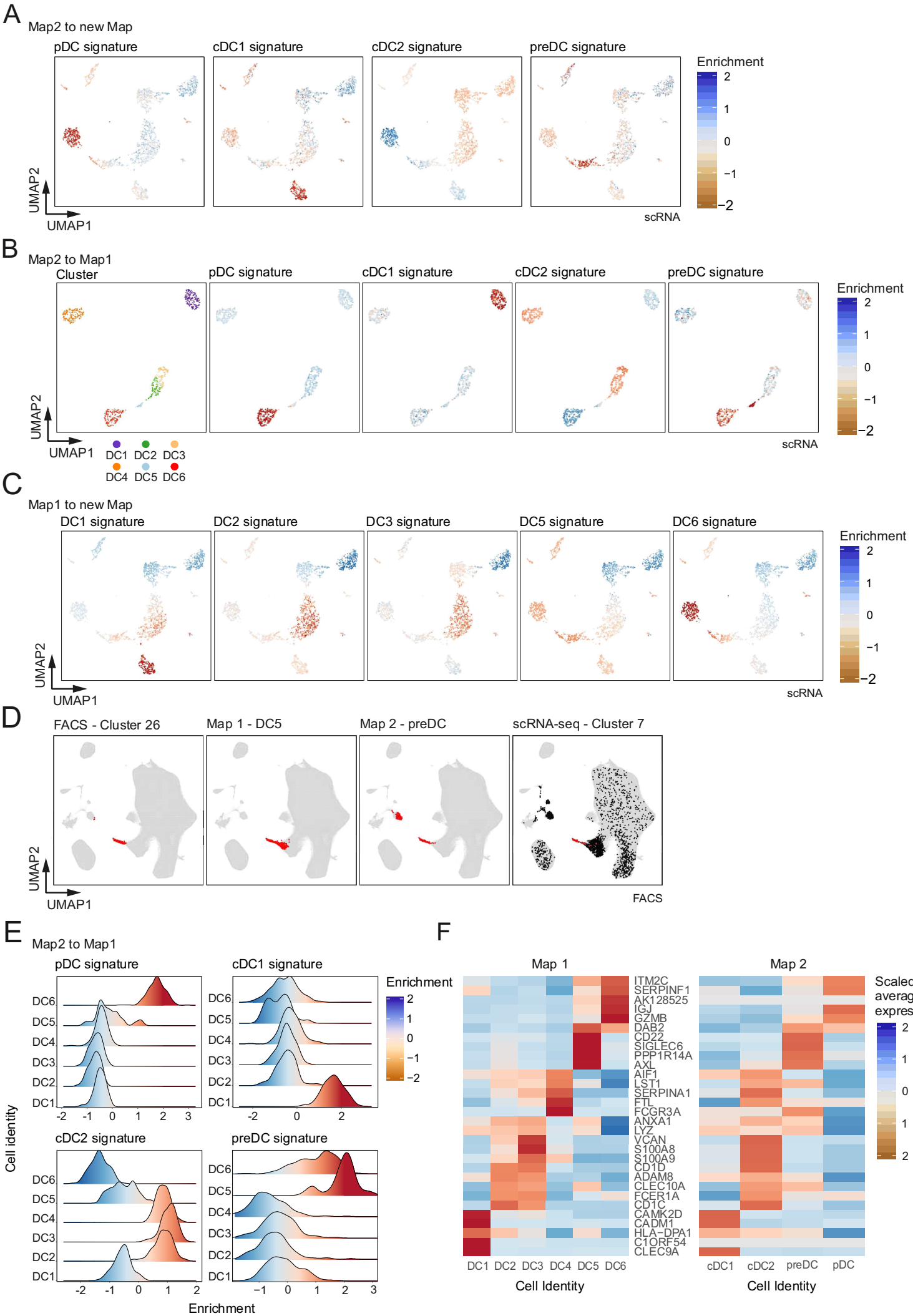
# Figure 2


A
De novo Clustering
Projection onto Flow Data
UMAP2
UMAP1
scRNA
FACS
Cluster
1 2 3 4 5 6 7
8 9 10 11
B
1 2 3 4 5 6 7 8 9 10 11
Cluster
Scaled average expression
-2 -1 0 1 2
C
Map 1
Map 2
Population
AS−DC
DN−DC
cDC1
cDC2
cMono
intMono
ncMono
pDC
undefined
CD34+
preDC
HLADR-

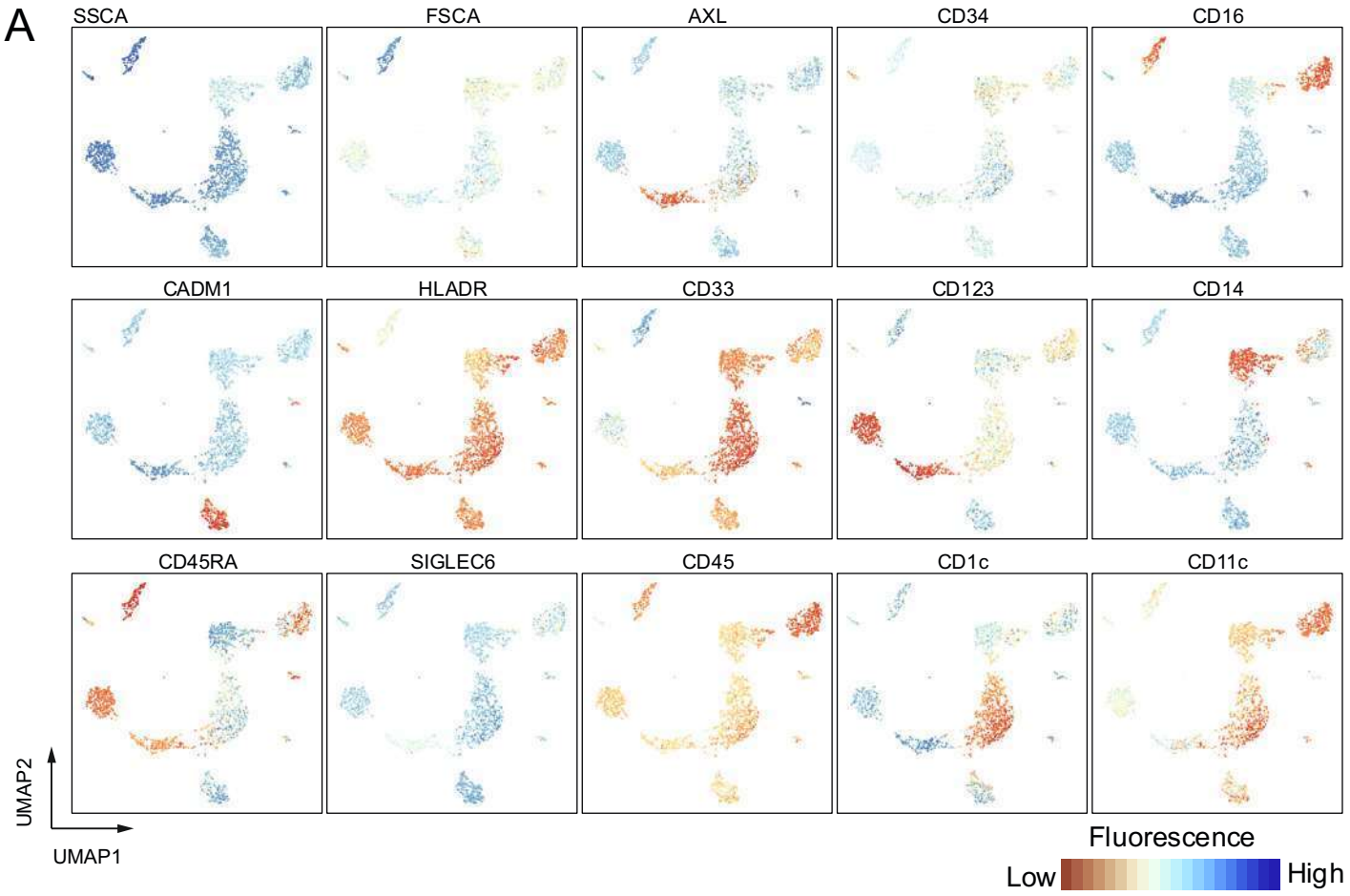
# Figure S2

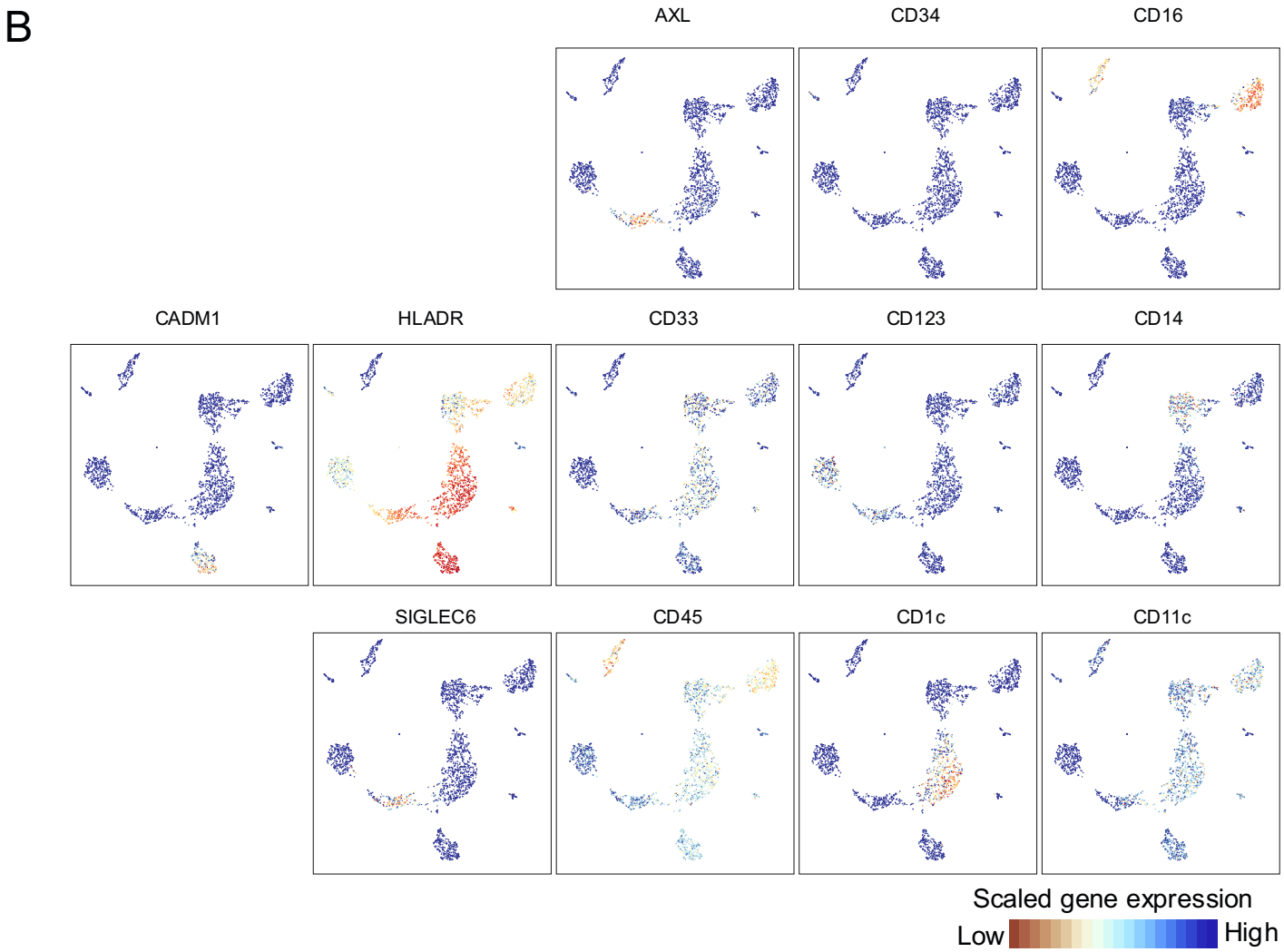

A
Number of Reads
1e+05
1e+03
1e+01
Plate1
Plate2
Plate3
Plate4
Plate5
Plate6
Plate7
Plate8
Plate ID
Number of Genes
12000
8000
4000
0
B
Number of reads
1e+06
1e+05
1e+04
Number of genes
12500
10000
7500
5000
2500
Percent of aligned reads
0.8
0.6
0.4
0.2
0.0
C
tSNE2
tSNE1
UMAP2
UMAP1
Plate1
Plate2
Plate3
Plate4
Plate5
Plate6
Plate7
Plate8
D
Total Reads
Unaligned Reads
Number of Genes
Number of Transcripts
Scaled quality metric
2
1
0
−1
−2
UMAP2
UMAP1
E
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
F
Percentage of Cluster
100%
75%
50%
25%
0%
Plate100
Plate101
Plate102
Plate104
Plate105
Plate97
Plate98
Plate99
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11

# Figure 3

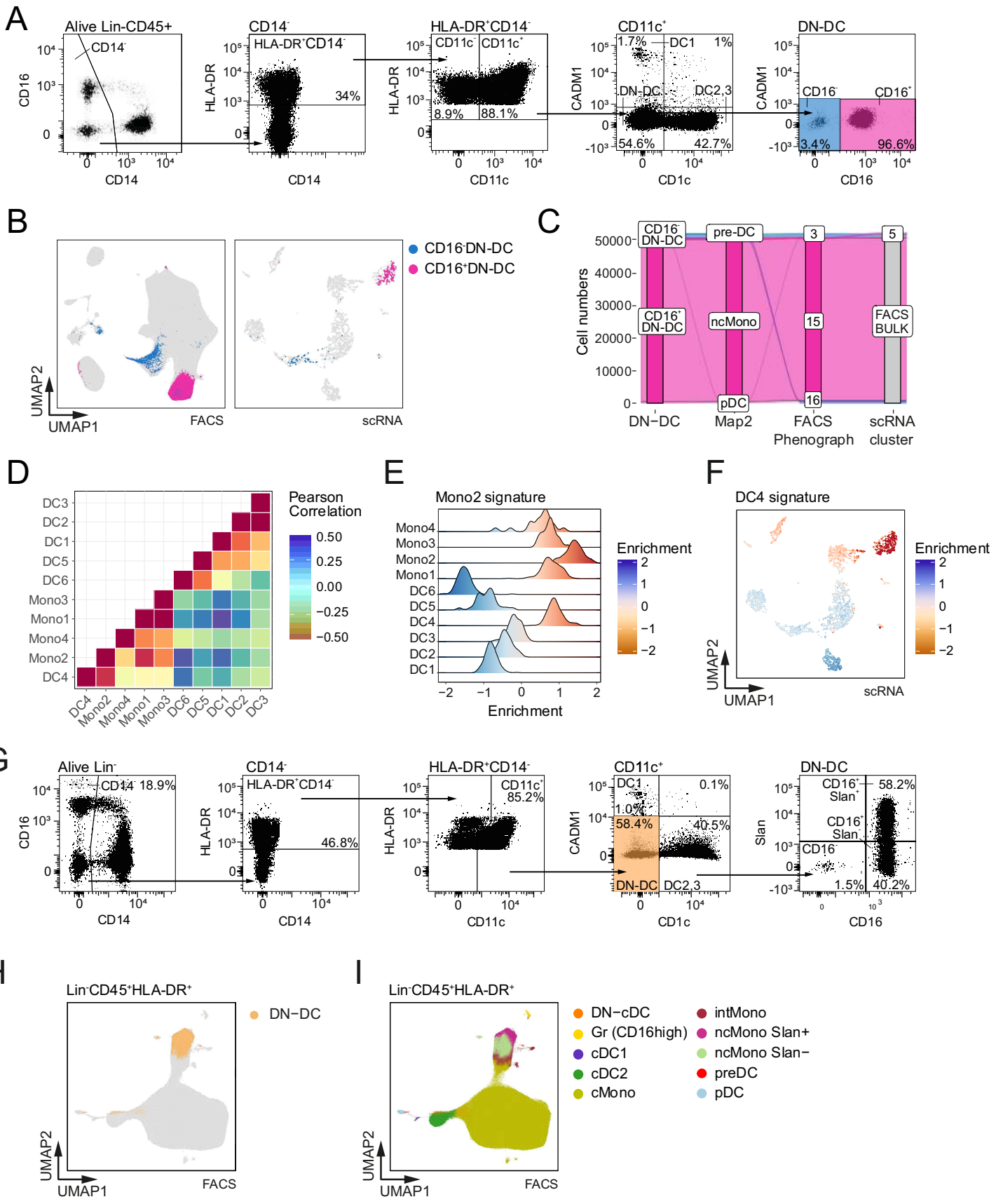

A
Map2 to new Map
pDC signature
cDC1 signature
cDC2 signature
preDC signature
Enrichment
2
1
0
−1
−2
UMAP2
UMAP1
scRNA
B
Map2 to Map1
Cluster
pDC signature
cDC1 signature
cDC2 signature
preDC signature
Enrichment
UMAP2
UMAP1
DC1 DC2 DC3
DC4 DC5 DC6
scRNA
C
Map1 to new Map
DC1 signature
DC2 signature
DC3 signature
DC5 signature
DC6 signature
Enrichment
UMAP2
UMAP1
scRNA
D
FACS - Cluster 26
Map 1 - DC5
Map 2 - preDC
scRNA-seq - Cluster 7
UMAP2
UMAP1
FACS
E
Map2 to Map1
pDC signature
cDC1 signature
cDC2 signature
preDC signature
Enrichment
DC6
DC5
DC4
DC3
DC2
DC1
Cell identity
Enrichment
F
Map 1
Map 2
ITM2C
SERPINF1
AK128525
IGJ
GZMB
DAB2
CD22
SIGLEC6
PPP1R14A
AXL
AIF1
LST1
SERPINA1
FTL
FCGR3A
ANXA1
LYZ
VCAN
S100A8
S100A9
CD1D
ADAM8
CLEC10A
FCER1A
CD1C
CAMK2D
CADM1
HLA−DPA1
C1ORF54
CLEC9A
DC1 DC2 DC3 DC4 DC5 DC6
cDC1 cDC2 preDC pDC
Cell Identity
Cell Identity
Scaled average expression

# Figure S3


A
SSCA
FSCA
AXL
CD34
CD16
CADM1
HLADR
CD33
CD123
CD14
CD45RA
SIGLEC6
CD45
CD1c
CD11c
UMAP2
UMAP1
Fluorescence
Low
High

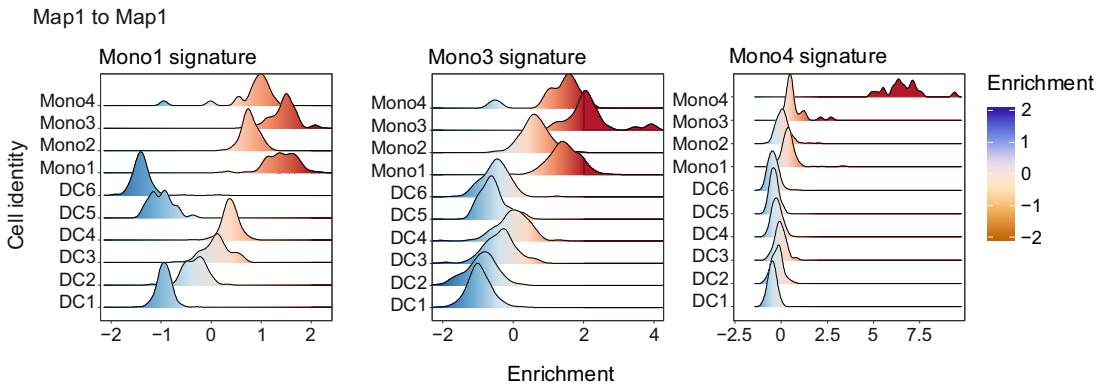

B
AXL
CD34
CD16
CADM1
HLADR
CD33
CD123
CD14
SIGLEC6
CD45
CD1c
CD11c
Scaled gene expression
Low
High

# Figure 4

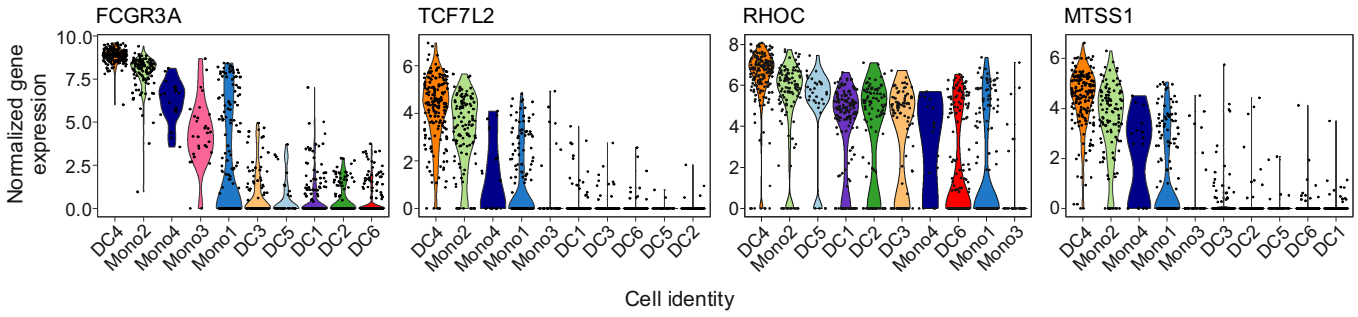

A
Alive Lin-CD45+
CD14-
CD16
CD14
CD14-
HLA-DR+CD14-
34%
HLA-DR
CD14
HLA-DR+CD14-
CD11c-
CD11c+
8.9%
88.1%
HLA-DR
CD11c
CD11c+
1.7%
DC1
1%
DN-DC
DC2,3
54.6%
42.7%
CADM1
CD1c
DN-DC
CD16-
CD16+
3.4%
96.6%
CADM1
CD16
B
CD16-DN-DC
CD16+DN-DC
UMAP2
UMAP1
FACS
scRNA
C
Cell numbers
CD16- DN-DC
CD16+ DN-DC
pre-DC
ncMono
pDC
3
15
16
5
FACS BULK
DN−DC
Map2
FACS Phenograph
scRNA cluster
D
DC3
DC2
DC1
DC5
DC6
Mono3
Mono1
Mono4
Mono2
DC4
Pearson Correlation
0.50
0.25
0.00
−0.25
−0.50
E
Mono2 signature
Mono4
Mono3
Mono2
Mono1
DC6
DC5
DC4
DC3
DC2
DC1
Enrichment
F
DC4 signature
Enrichment
UMAP2
UMAP1
scRNA
G
Alive Lin-
CD14- 18.9%
CD14
CD14-
HLA-DR+CD14-
46.8%
CD14
HLA-DR+CD14-
CD11c+
85.2%
CD11c
CD11c+
DC1
0.1%
1.0%
58.4%
40.5%
DN-DC
DC2,3
CD1c
DN-DC
CD16+ Slan+
58.2%
CD16+ Slan-
CD16-
1.5%
40.2%
CD16
H
Lin-CD45+HLA-DR+
DN−DC
UMAP2
UMAP1
FACS
I
Lin-CD45+HLA-DR+
DN−cDC
Gr (CD16high)
cDC1
cDC2
cMono
intMono
ncMono Slan+
ncMono Slan−
preDC
pDC
UMAP2
UMAP1
FACS

# Figure S4

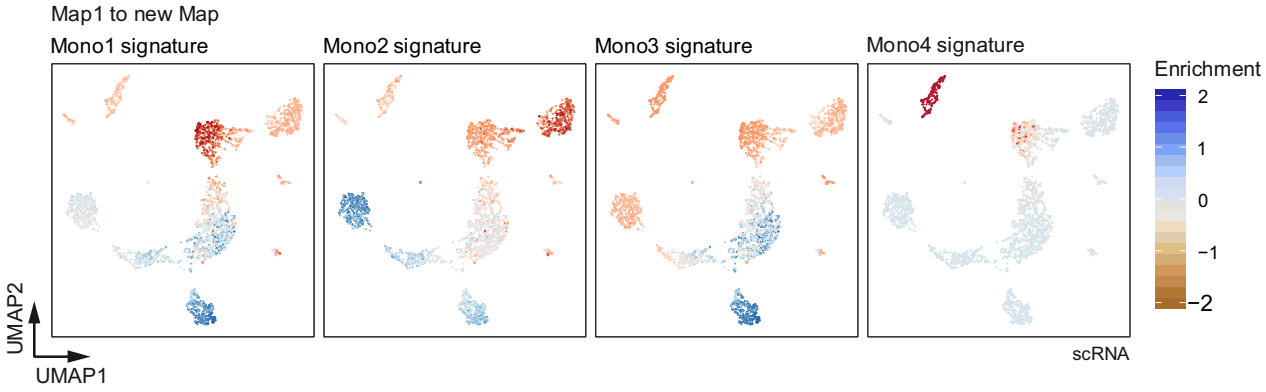
A



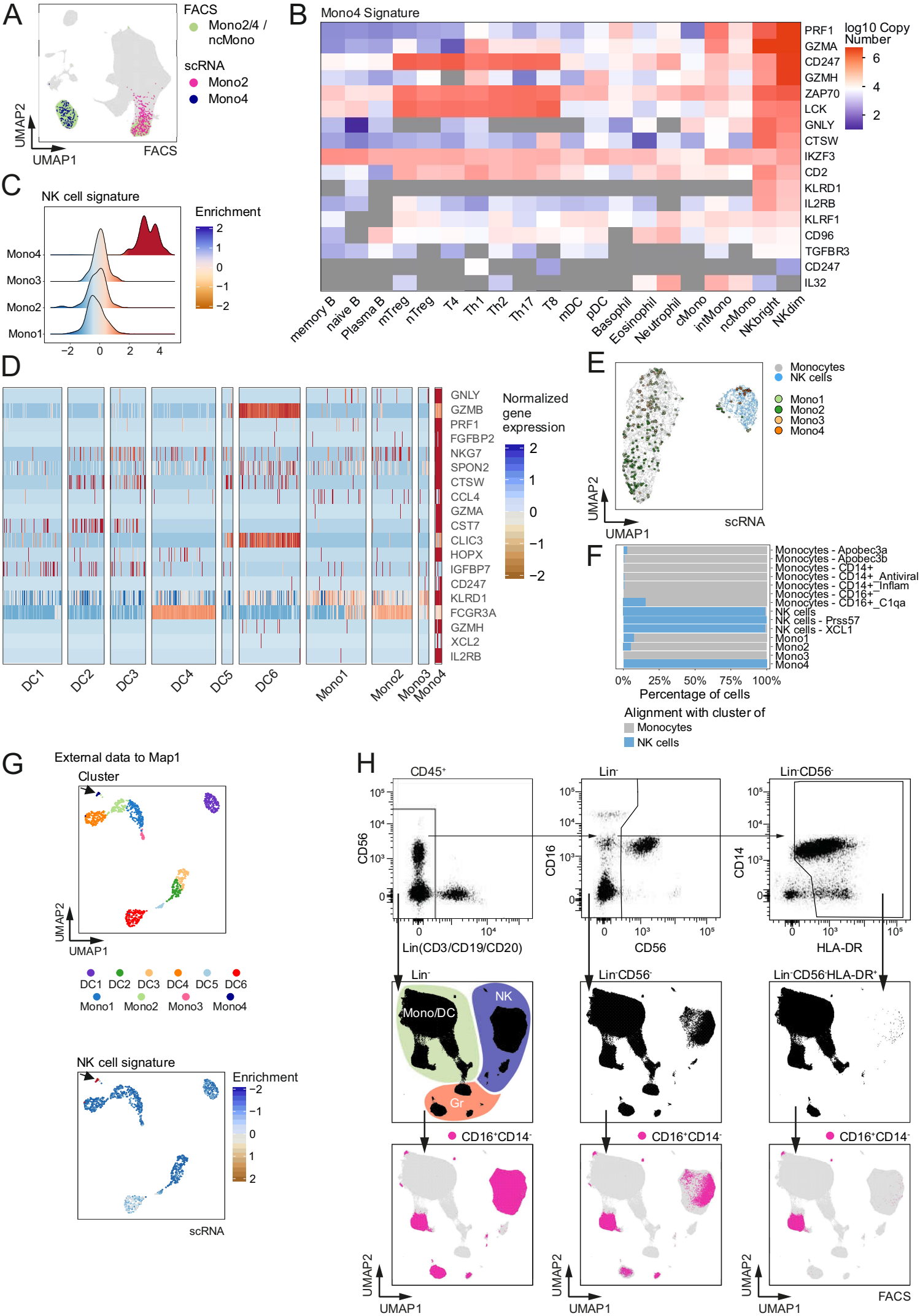

B



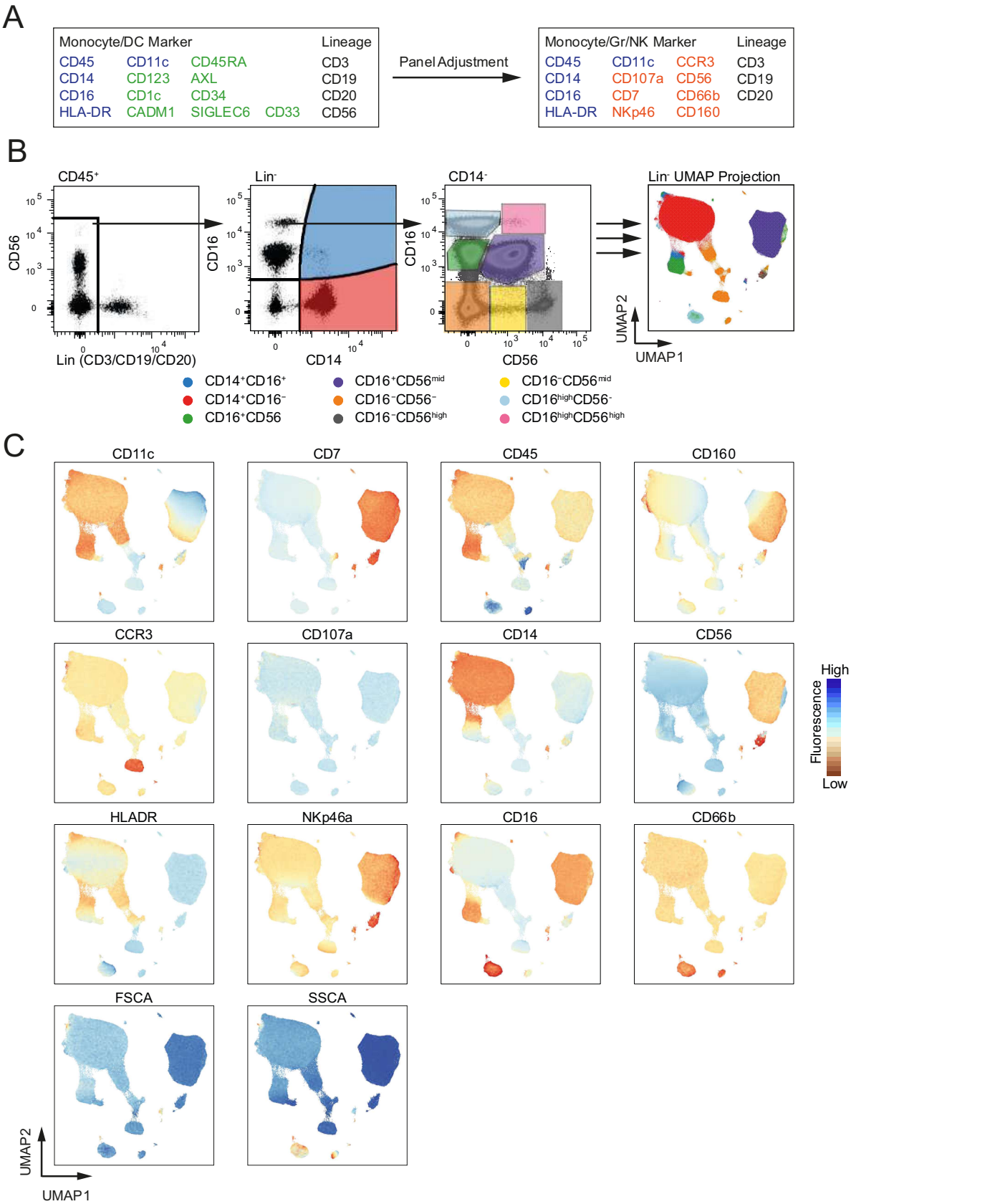

C



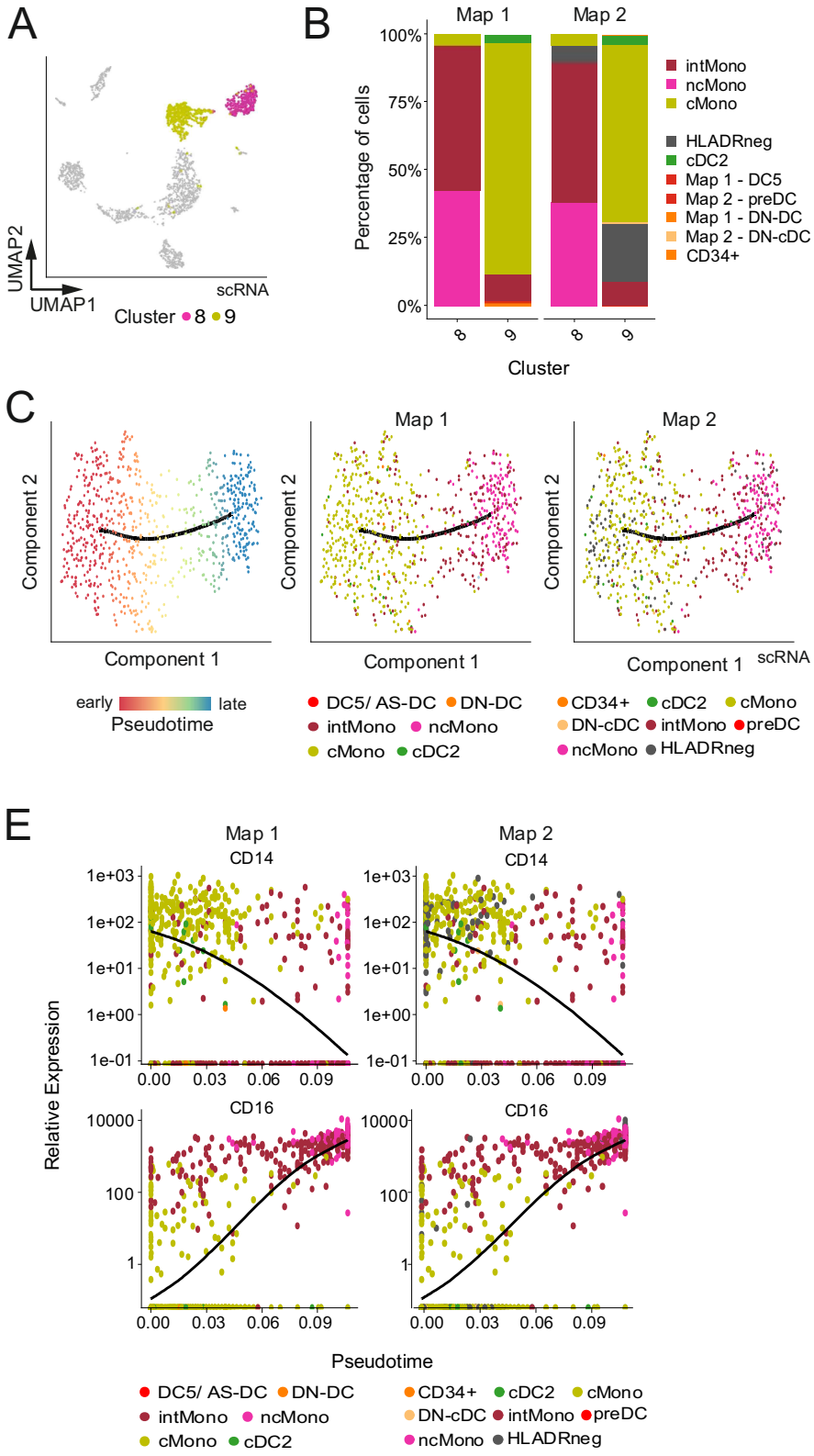
# Figure 5


A
FACS
Mono2/4 / ncMono
scRNA
Mono2
Mono4
UMAP2
UMAP1
FACS
B
Mono4 Signature
log10 Copy Number
PRF1
GZMA
CD247
GZMH
ZAP70
LCK
GNLY
CTSW
IKZF3
CD2
KLRD1
IL2RB
KLRF1
CD96
TGFBR3
CD247
IL32
memory B
naive B
Plasma B
mTreg
nTreg
T4
Th1
Th2
Th17
T8
mDC
pDC
Basophil
Eosinophil
Neutrophil
cMono
intMono
ncMono
NKbright
NKdim
C
NK cell signature
Enrichment
Mono4
Mono3
Mono2
Mono1
D
Normalized gene expression
GNLY
GZMB
PRF1
FGFBP2
NKG7
SPON2
CTSW
CCL4
GZMA
CST7
CLIC3
HOPX
IGFBP7
CD247
KLRD1
FCGR3A
GZMH
XCL2
IL2RB
DC1
DC2
DC3
DC4
DC5
DC6
Mono1
Mono2
Mono3
Mono4
E
Monocytes
NK cells
Mono1
Mono2
Mono3
Mono4
scRNA
UMAP2
UMAP1
F
Monocytes - Apobec3a
Monocytes - Apobec3b
Monocytes - CD14+
Monocytes - CD14+_Antiviral
Monocytes - CD14+_Inflam
Monocytes - CD16+
Monocytes - CD16+_C1qa
NK cells
NK cells - Prss57
NK cells - XCL1
Mono1
Mono2
Mono3
Mono4
Percentage of cells
Alignment with cluster of
Monocytes
NK cells
G
External data to Map1
Cluster
UMAP2
UMAP1
NK cell signature
Enrichment
scRNA
H
CD45+
Lin-
Lin-CD56-
CD56
CD16
CD14
Lin(CD3/CD19/CD20)
HLA-DR
Mono/DC
NK
Gr
Lin-CD56-HLA-DR+
CD16+CD14-
FACS

# Figure S5


A
Monocyte/DC Marker
CD45 CD11c CD45RA
CD14 CD123 AXL
CD16 CD1c CD34
HLA-DR CADM1 SIGLEC6 CD33
Lineage
CD3
CD19
CD20
CD56
Panel Adjustment
Monocyte/Gr/NK Marker
CD45 CD11c CCR3
CD14 CD107a CD56
CD16 CD7 CD66b
HLA-DR NKp46 CD160
Lineage
CD3
CD19
CD20
B
CD45+
Lin-
CD14-
Lin- UMAP Projection
CD56
CD16
CD14
Lin (CD3/CD19/CD20)
CD56
CD16
UMAP2
UMAP1
CD14+CD16+
CD14+CD16-
CD16+CD56
CD16+CD56mid
CD16-CD56-
CD16-CD56high
CD16-CD56mid
CD16highCD56-
CD16highCD56high
C
CD11c
CD7
CD45
CD160
CCR3
CD107a
CD14
CD56
HLADR
NKp46a
CD16
CD66b
FSCA
SSCA
High
Fluorescence
Low
UMAP2
UMAP1
D
CD16high
CD14+16+
CD16-CD56high
Mono2
Mono4

# Figure 6


A
UMAP2
UMAP1
scRNA
Cluster 8 9
B
Map 1
Map 2
100%
75%
50%
25%
0%
Percentage of cells
8
9
Cluster
intMono
ncMono
cMono
HLADRneg
cDC2
Map 1 - DC5
Map 2 - preDC
Map 1 - DN-DC
Map 2 - DN-cDC
CD34+
C
Component 2
Component 1
Map 1
Map 2
scRNA
early
late
Pseudotime
DC5/ AS-DC
DN-DC
intMono
ncMono
cMono
cDC2
CD34+
DN-cDC
preDC
HLADRneg
E
Map 1
Map 2
CD14
CD16
1e+03
1e+02
1e+01
1e+00
1e-01
10000
100
1
0.00
0.03
0.06
0.09
Relative Expression
Pseudotime
DC5/ AS-DC
DN-DC
intMono
ncMono
cMono
cDC2
CD34+
DN-cDC
preDC
HLADRneg

# Figure S6

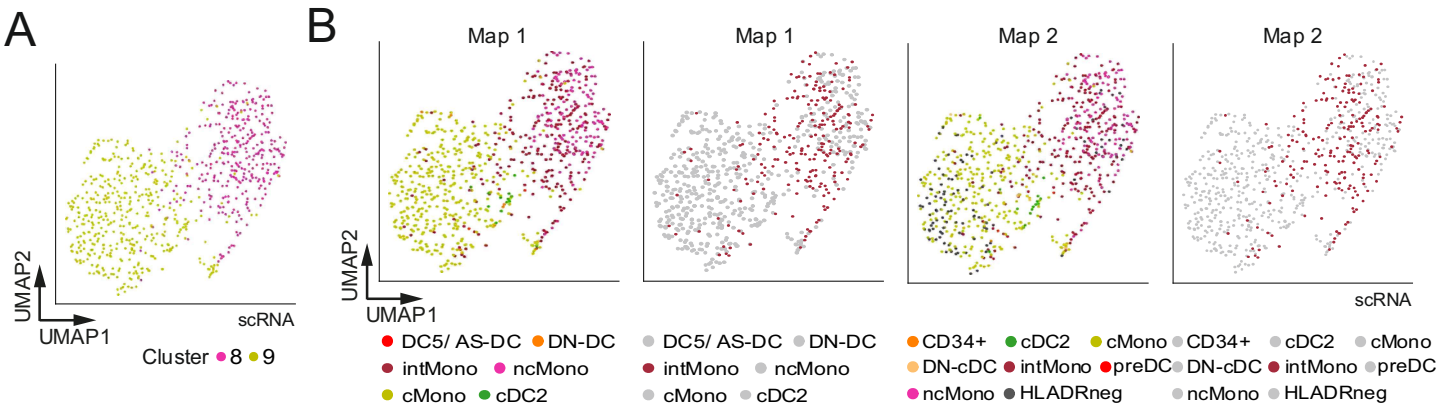
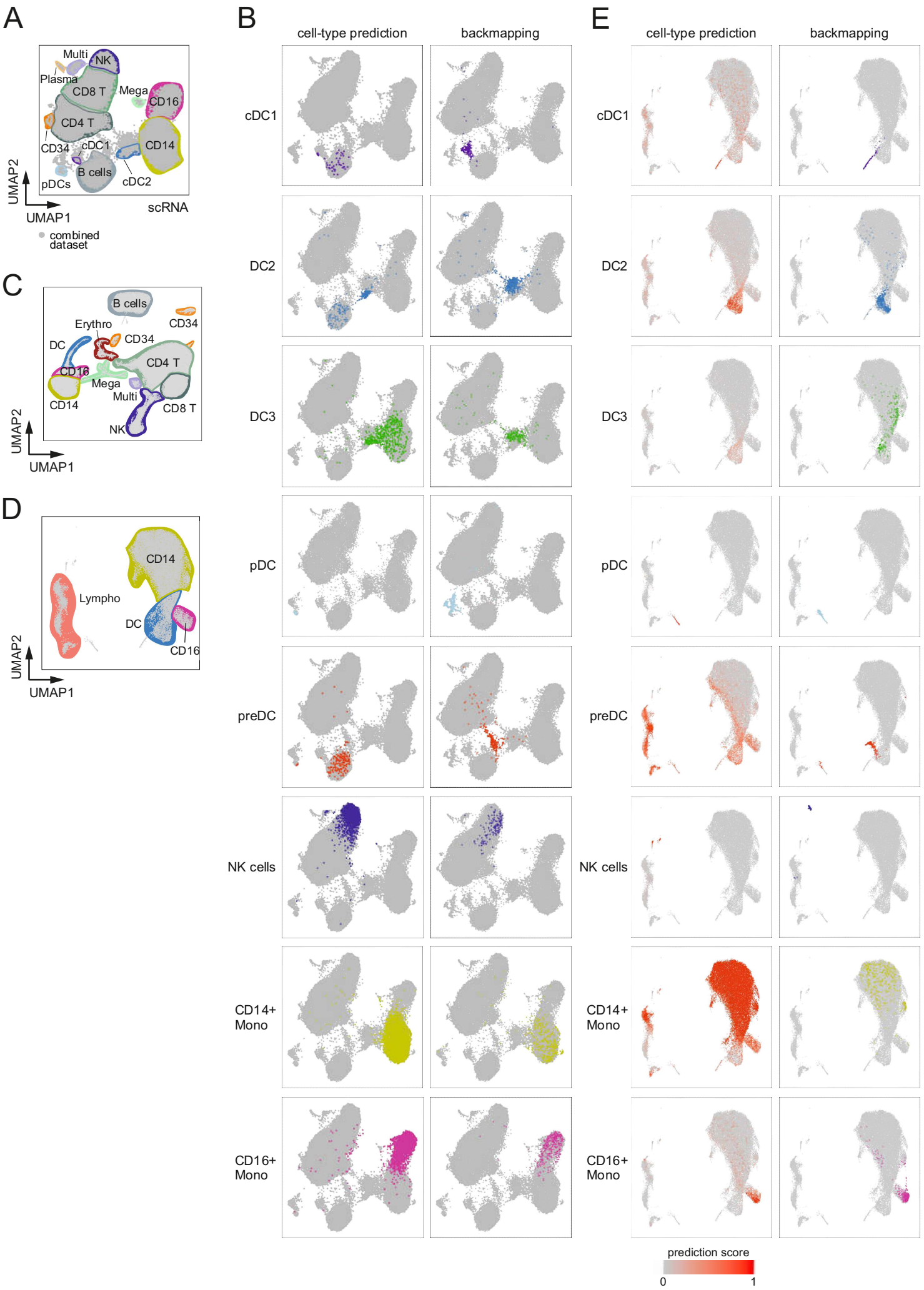

A
UMAP2
UMAP1
scRNA
Cluster 8 9
B
Map 1
Map 1
Map 2
Map 2
UMAP2
UMAP1
scRNA
DC5/ AS-DC DN-DC
intMono ncMono
cMono cDC2
DC5/ AS-DC DN-DC
intMono ncMono
cMono cDC2
CD34+ cDC2 cMono
DN-cDC intMono preDC
ncMono HLADRneg
CD34+ cDC2 cMono
DN-cDC intMono preDC
ncMono HLADRneg

Figure 7


A
Multi
NK
Plasma
CD8 T
Mega
CD16
CD4 T
CD34
cDC1
CD14
B cells
pDCs
cDC2
UMAP2
UMAP1
scRNA
combined dataset
C
B cells
CD34
Erythro
DC
CD34
CD4 T
CD16
Mega
Multi
CD14
CD8 T
NK
UMAP2
UMAP1
D
CD14
Lympho
DC
CD16
UMAP2
UMAP1
B
cell-type prediction
backmapping
cDC1
DC2
DC3
pDC
preDC
NK cells
CD14+ Mono
CD16+ Mono
E
cell-type prediction
backmapping
cDC1
DC2
DC3
pDC
preDC
NK cells
CD14+ Mono
CD16+ Mono
prediction score
0
1

# Figure S7

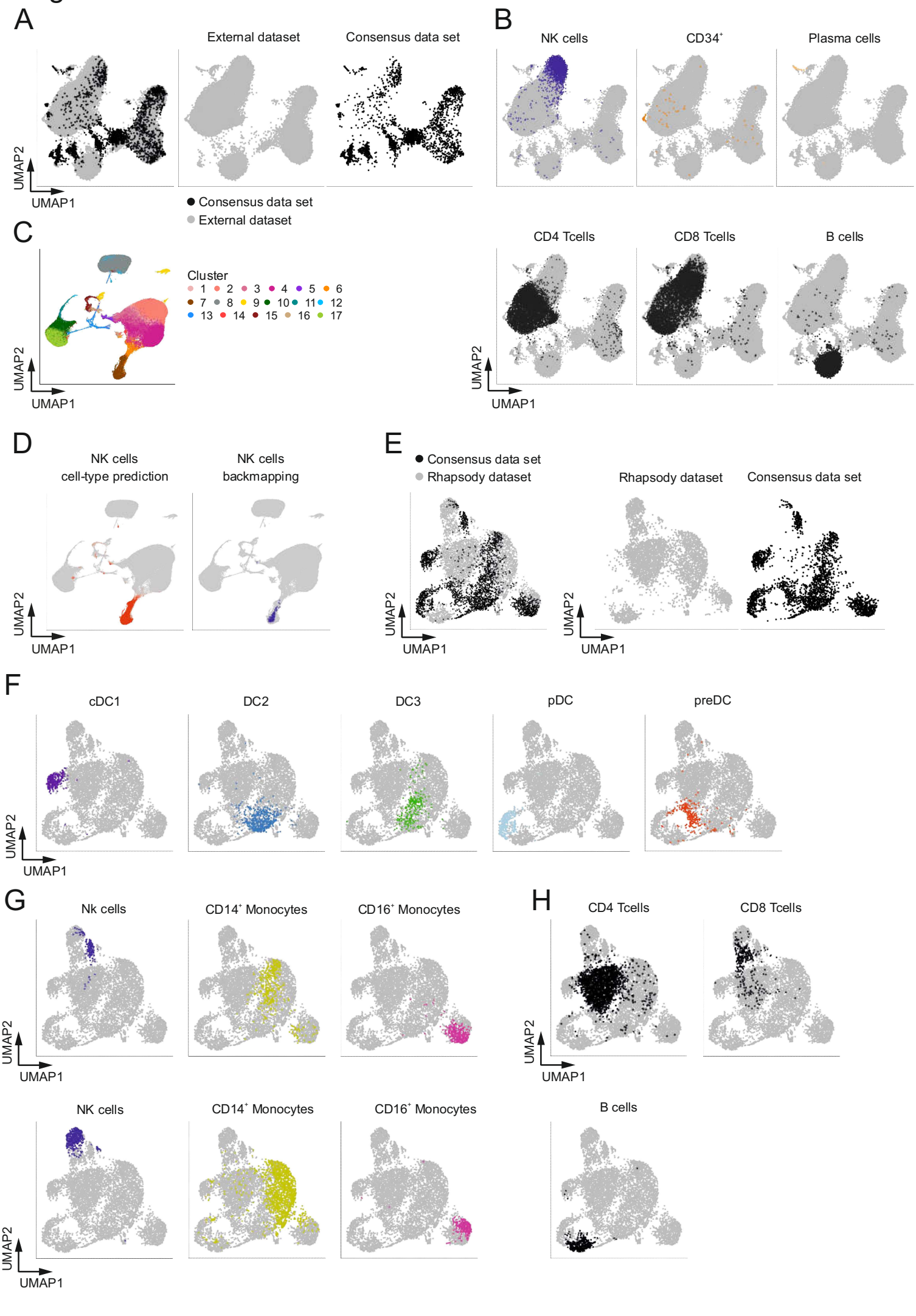

A
External dataset
Consensus data set
UMAP2
UMAP1
Consensus data set
External dataset
B
NK cells
CD34+
Plasma cells
C
Cluster
1 2 3 4 5 6
7 8 9 10 11 12
13 14 15 16 17
CD4 Tcells
CD8 Tcells
B cells
D
NK cells
cell-type prediction
NK cells
backmapping
E
Consensus data set
Rhapsody dataset
Rhapsody dataset
Consensus data set
F
cDC1
DC2
DC3
pDC
preDC
G
Nk cells
CD14+ Monocytes
CD16+ Monocytes
NK cells
CD14+ Monocytes
CD16+ Monocytes
H
CD4 Tcells
CD8 Tcells
B cells

**Table legends**

**Tables S1:**

Cell types classified in the respective studies

**Data Table S1:**

Data Table S1.csv. Gene signatures of the 11 clusters identified in our new scRNA-seq consensus map

**Data Table S2:**

Data Table S2.xlsx. Gene signatures derived from map 2

**Table S1. Cell types classified in the respective studies**

| Cell types identified in Map1 (*12*) | Cell types in Map2 (*13*) | Cluster identified in the consensus Map | Cell types in the consensus Map |
|---|---|---|---|
| DC1 | cDC1 | Cluster 6 | cDC1 |
| DC2 | cDC2 | Cluster 10 | DC2 |
| DC3 | cDC2 | Cluster 11 | DC3 |
| DC4 | - | - | ncMonos, Cd34 |
| DC5 (AS-DC) | preDC | Cluster 7 | preDC |
| DC6 | pDC | Cluster 5 | pDC |
| Mono1 | - | Cluster 9 | CD14 Mono |
| Mono2/ DC4 | - | Cluster 8 | CD16 mono |
| Mono3 | - | - | intMonos |
| Mono4 | - | Cluster 1 | NK cells |
| | | | |
| - | | | Plasma cells |
| - | | | Undef. |

## EXPERIMENTAL PROCEDURE

### CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact jschultze@uni-bonn.de

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Peripheral blood mononuclear cells (PBMC)

Buffy coats or venipuncture blood were obtained from healthy donors (University hospital Bonn, local ethics vote 203/09) after written consent was given according to the Declaration of Helsinki. Peripheral blood mononuclear cells (PBMC) were isolated by Pancoll (PAN-Biotech) density centrifugation from buffy coats.

### METHOD DETAILS

#### Flow cytometric analysis

Whole blood or buffy coat was diluted in room temperature PBS (1:2 or 1:5, respectively) and layered onto polysuccrose solution (Pancoll; PAN Biotech, Germany) for the enrichment of mononuclear cells by density gradient centrifugation according to the manufacturer's instructions. After three times washing in cold PBS, cells were resuspended in FACS-buffer (0.5% BSA, 2 mM EDTA in PBS) for surface marker staining (Table S2). Human FcR-Block (Miltenyi Biotec, Germany) was included to reduce unspecific staining. After 1 h incubation at 4º in the dark, cells were washed and optionally stained for additional 20 min with 1:400 anti-biotin BV421 in FACS-buffer for CADM1-biotin secondary staining. Washed cells were incubated with L/D Marker DRAQ7 (BioLegend, USA) for 5 min at room temperature before acquisition and sorting of the cells using a BD FACSARIA III (BD BioSciences, USA). Single antibody staining was prepared in

parallel to assess fluorescence spillover. Fluorescence-minus-one (FMO) controls were prepared in addition for critical markers to set sorting gates. Post-sort data analysis was performed using FlowJo software (FlowJo, Tree Star Inc., USA). The packages "flowCore" and "flowWorkspace" were used to import raw data into R. For dimensionality reduction with UMAP fluorescence parameters were transformed with logicleTransform (Becht et al., 2018).

**Library preparation and sequencing using Smart-Seq2**

Our new index-sorted single cell transcriptome dataset was based on the Smart-Seq2 protocol (Picelli et al., 2013). For single cell sorting into 384-well plates and to ensure sufficient cell numbers and balanced representation from each main myeloid subset, loose sorting gates have been set covering the entire space of alive $CD45^+Lin$ (CD3, CD19, CD20, $CD56)^-CD14^+$, $CD16^+$ or $CD14^-CD16^-HLA\text{-}DR^+$ cells. To achieve this, the alive $CD45^+Lin^-$ compartment was divided to sort 24 cells per plate each of $CD14^+CD16^-$, $CD14^-CD16^+$, $CD14^+CD16^+$, $HLA\text{-}DR^+CADM1^+$, $HLA\text{-}DR^+CADM^-AXL^+SIGLEC6^+$, $HLA\text{-}DR^+CADM1^-AXL^-SIGLEC6^-CD123^+CD11c^-$, $HLA\text{-}DR^+CADM1^-AXL^-SIGLEC6^-CD123^-CD11c^-$ or $HLA\text{-}DR^+CADM^-AXL^-SIGLEC6^-CD123^-CD11c^-$ cells. Cells were FACS sorted into eight 384-well plates containing 2.3µl lysis buffer (Guanidine Hydrochloride (50 mM), dNTPs (17.4mM), 2.2µM SMART dT30VN primer) retaining protein expression information for every well to subsequently match with the respective single-cell transcriptomic data in an index sorting approach. Plates were sealed and stored at -80°C until further processing. Smart-Seq2 libraries were finally generated on a Tecan Freedom EVO and Nanodrop II (BioNex) system as previously described (Picelli et al., 2013).

In short, lysed cells were incubated at 95°C for 3 min. 2.7 µl RT mix containing SuperScript II buffer (Invitrogen), 9.3mM DTT, 370mM Betaine, 15mM MgCl2, 9.3U SuperScript II RT (Invitrogen), 1.85U recombinant RNase Inhibitor (Takara), 1.85 µM template-switching oligo was aliquoted to each lysed cell using a Nanodrop II liquid handling system (BioNex) and incubating

at 42°C for 90 min and at 70°C for 15min. 7.5µl preamplification mix containing KAPA HiFi HotStart ReadyMix and 2µM ISPCR primers was added to each well and full-length cDNA was amplified for 16 cycles. cDNA was purified with 1X Agencourt AMPure XP beads (Beckman Coulter) and eluted in 14µl nuclease-free water. Concentration and cDNA size were checked for select representative wells using a High Sensitivity DNA5000 assay for the Tapestation 4200 (Agilent). cDNA was diluted to an average of 200pg/µl and 100pg cDNA from each cell was tagmented by adding 1µl TD and 0.5µl ATM from a Nextera XT DNA Library Preparation Kit (Illumina) to 0.5µl diluted cDNA in each well of a fresh 384-well plate. The tagmentation reaction was incubated at 55°C for 8min before removing the Tn5 from the DNA by adding 0.5µl NT buffer per well. 1µl well-specific indexing primer mix from Nextera XT Index Kit v2 Sets A-D and 1.5µl NPM was added to each well and the tagmented cDNA was amplified for 14 cycles according to manufacturer's specifications. PCR products from all wells were pooled and purified with 1X Agencourt AMPure XP beads (Beckman Coulter) according to manufacturer's protocol. The fragment size distribution was determined using a High Sensitivity DNA5000 assay for the Tapestation 4200 (Agilent) and library concentration was determined using a Qubit dsDNA HS assay (Thermo Fischer). Libraries were clustered at 1.4pM concentration using High Output v2 chemistry and sequenced on a NextSeq500 system SR 75bp with 2*8bp index reads. Single-cell data was demultiplexed using bcl2fastq2 v2.20.

**Proteomic Data**

To validate the gene signature associated with the mono 4 subset as described by Villani et al. on the protein level, we extracted copy numbers from key signature proteins from the publicly accessible proteomic resource (http://www.immprot.org/) described by Rieckmann et al. containing quantitative high-resolution mass-spectrometry data derived from FACS-enriched

human primary blood cells (Rieckmann et al., 2017). Copy numbers were visualized as a heatmap using the *pheatmap* package (v1.0.10) in R.

**Cytospin preparation and May-Grünwald/Giemsa staining**

Cell populations of interest were sorted into 1.5 ml reaction tubes containing 200 µl FACS-buffer using a BD FACSARIA III (BD BioSciences, USA). The cell suspension was centrifuged onto SuperFrost Plus glass slides (Thermo Scientific, USA) at 1000 rpm for 5 min using a Universal 16A slide centrifuge (Andreas Hettich GmbH & Co.KG, Germany). Slides were air-dried overnight and subsequently stained with May-Grünwald/Giesma solution (Carl Roth GmbH, Germany) according to the manufacturer's guidelines. Images were acquired with a BZ-9000 (Keyence, Japan).

**Targeted sequencing of human PBMC with the BD Rhapsody™ system**

Whole blood was diluted in room temperature PBS (1:2) and layered onto polysuccrose solution (Pancoll; PAN Biotech, Germany) for the enrichment of mononuclear cells by density gradient centrifugation according to the manufacturer's instructions. Granulocytes were isolated using erythrocyte lysis buffer (ELB, 0.15M NH4Cl, 0.01M KHCO3, 0.1mM EDTA, pH 7.4 at ca. 2-8°C) and mononuclear cells and granulocytes are mixed in a ratio of 2:1. After washing in cold PBS 10.000 cells were loaded onto a BD Rhapsody™ cartridge and processed according to manufacturer's instructions for targeted single-cell RNA-seq using the predesigned Immune Response Panel (Human). The library was clustered at 1.75pM on a NextSeq500 system (Illumina) to generate ~40.000 paired end (2*75bp) reads per cell using High Output v2 chemistry. Sequenced single-cell data was demultiplexed using bcl2fastq2 v2.20.

**Single-cell RNA-Seq raw data processing**

Following sequencing by the Smart-Seq2 method (Picelli et al., 2013), RNA-Seq libraries were subjected to initial quality control using FASTQC (http://www.bioinformatics.babraham.ac.uk /projects/fastqc, v0.11.7) implemented in a scRNA pre-processing pipeline (docker image and scripts available at https://hub.docker.com/r/pwlb/rna-seq-pipeline-base/, v0.1.1; https://bitbucket.org/limes_bonn/bulk-rna-kallisto-qc/src/master/, v0.2.1). Next, raw reads were pseudoaligned to the human transcriptome (GRCh38, Gencode v27 primary assembly) using Kallisto with default settings (v0.44.0) (Bray et al., 2016). Based on the pseudoalignment estimated by Kallisto, transcript levels were quantified as transcripts per million reads (TPM). TPM counts were imported into R using tximport (Soneson et al., 2015) and transcript information was summarized on gene-level. We imported the resulting dataset of 43,612 features across 3,072 samples and performed the downstream analysis using the R package Seurat (v.2.3.4, (Butler et al., 2018)).

For processing of the single-cell data obtained by the BD Rhapsody™ system, we run the recommended BD Rhapsody™ Analysis Pipeline of Seven Bridges Genomics (sbgenomics.com/bdgenomics) with standard settings. The resulting count table that was accounted for UMI sequencing and amplification errors, was comprised of 488 features across 7,873 cells. Normalization and further downstream analysis were conducted in R using Seurat (v.2.3.4, (Butler et al., 2018)).

**Quality control**

Concerning our new index-sorted and Smart-Seq2-based single cell transcriptome dataset the following quality control scheme using various meta information was performed to obtain high-quality transcriptome data: 1) We removed genes that are detected in less than 6 cells (0.2 percent of cells), 2) and removed cells that have less than 1,000 uniquely detected genes. Next,

we filtered further outlier cells with 3) less than 50,000 unique reads, 4) less than 30% pseudoalignment of reads to the transcriptome, 5) a lower rate of endogenous-to-mitochondrial count rate of 2, 6). This quality control scheme results in a dataset of 29,240 genes across 2,509 cells.

**Normalization of single-cell transcriptomic data**

To reduce the influence of variation of sequencing depth among samples we applied a log-normalization to the data and scaled each cells gene expression profile to a total count of 10,000. In addition, we corrected for other technical effects including differences in the fraction of mitochondrial counts as well of unique detected genes using a linear regression model for these factors. The residuals of this regression are scaled and centered and used for further downstream analysis.

**Dimensionality reduction and clustering**

In order to reduce the dimensionality of the dataset, we selected highly variable genes as genes with an average expression of at least 0.0125 and a scaled dispersion of at least 1. This resulted in a total of 2491 genes, which were used as input for a principal component (PC) analysis. We visualized the standard deviation of the first 20 PCs and identified the first 10 principal components with a minimum standard deviation of at least 2 as significant PCs. Next, we utilized Uniform Manifold Approximation and Projection (UMAP) to further reduce the data into a two-dimensional representation (Becht et al., 2018). To test for cellular heterogeneity, we used a shared nearest neighbor (SNN)-graph based clustering algorithm implemented in the Seurat package. We used the first 10 principal components for constructing the SNN-graph and set the resolution to 1. Monocle was used to infer differentiation trajectories by using the Louvain clustering method, umap dimensionality reduction and the SimplePPT algorithm (Qiu et al., 2017)

**Additional analysis**

Differentially expressed (DE) genes were defined using a Wilcoxon-based test for differential gene expression built in the Seurat pipeline (v.2.3.4) (Data Table S1). Unless otherwise stated genes have been considered as differentially expressed, if the adjusted p-value is smaller than 0.1. Top10 DE genes have been visualized using heatmap of hierarchical clustered gene expression profiles. DE genes have been verified with current literature.

**Gene signature enrichment analysis**

Single-cell RNA-Seq data is inherently sparse and a high-dropout rate is limiting the use of single marker genes to identify cell populations. In order to unambiguously identify the different cell types, we have used an updated version of a gene signature score analysis described earlier (Mass et al., 2016). A cell population is always characterized by genes that are significantly upregulated in comparison to other populations and genes that show significantly lower expression in comparison to the background populations. In order to increase the power, we use both up and downregulated gene signatures for the calculation of the gene expression scores. A cell *i* may be described by a gene expression profile *A[i,j]* as the combination of gene expression values of all genes *j*. To calculate a signature score for a cell *i*, we first calculate the scaled average expression of all genes $j_{up}$ from an upregulated list and of all genes $j_{down}$ from a downregulated gene list. The difference between these two is scaled and visualized. The visualization is performed as color-coded overlay on the UMAP dimensionality reduction or as density distribution.

**Data analysis of external single-cell RNA-Seq datasets**

To assess the single-cell RNA-Seq data of human dendritic cells and monocytes publicly available under the Gene Expression Omnibus accession number GSE94820, we applied the processing

steps previously described (Villani et al., 2017). We focused on the “discovery” dataset and performed downstream analysis with the R software package Seurat (https://github.com/satijalab/seurat; http://satijalab. org/seurat/, v.2.3.4). Uniform manifold approximation (UMAP) algorithm integrated into the Seurat package was used as a dimensionality reduction method with standard settings. To define cell-type specific gene signatures for all cell populations, a Wilcoxon-based test was used. We considered genes as differentially expressed with an adjusted p-value of smaller 0.1 and a log2-fold change of higher than 1 or lower than -1, respectively. A global comparison of all cell types was performed by calculating the Pearson Correlation coefficients between the average expression profiles of all clusters. Scaled gene expression profiles have been used.

In order to have a comprehensive single-cell RNA-Seq dataset of human PBMCs, we downloaded a dataset containing transcriptome data of 33,148 PBMCs from a healthy donor (short 33k-PBMC dataset), which is publicly available on the 10x Genomics webpage (https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc33k). Next, we followed the general data analysis scheme described at the Seurat package webpage (https://satijalab.org/seurat/get_started_v1_4.html). Briefly, we used the filtered cell-gene matrix provided by 10x Genomics and imported the data and performed the analysis with the Seurat package. We filtered genes that are expressed in less than three cells and removed cells from the data set that have gene counts for less than 500 genes or for more than 2500 genes. In addition, we removed cells that have more than 5% mitochondrial counts. This resulted in a dataset of 17943 genes across 28.823 cells. Next, a log-normalization was applied, and highly variable genes were identified by applying a dispersion cutoff of 0.8 (2.281 variable genes). To account for technical variability in the dataset, a linear model was used to regress out the effects of the number of measured molecules per cell, the fraction of mitochondrial counts as well as the effect introduced by processing the cells in different sets. The first 25 principal components were

used for a graph-based clustering approach. NK cell specific genes were identified by a Wilcoxon-based test for differential gene expression (adj. p-value < smaller 0.1, |log2-fold change| > 1).

**Backmapping**

In order to compare the transcriptome profiles of monocytes isolated from the dataset derived from GSE94820 (Villani et al., 2017) with the comprehensive PBMC dataset, we used the previously introduced canonical correlation alignment to combine datasets (Butler et al., 2018). First, we isolated all monocyte populations from Villani et al. and all monocyte and NK cell populations of the 10x Genomics dataset. Both datasets are normalized, scaled and a linear regression was performed to account for differences in the number of detected genes. In both datasets, a feature selection was performed to identify genes with high dispersion. We determined the mutual highly variable genes as the overlap of the 4.000 genes from each dataset with highest dispersion. Next, we combined both datasets by performing the canonical correlation alignment, which resulted in an integrated dataset comprising 41.620 genes across 8.846 cells. UMAP dimensionality reduction was applied to the dataset using the first 8 canonical correlation alignment components and 40 neighbor points as well as a minimal distance of 0.01.

In addition, we downloaded from the data portal (https://preview.data.humancellatlas.org/) of the HCA consortium a single-cell dataset comprised of immune cells from human cord blood samples. When analyzing this dataset, we observed a donor dependent batch effect and thus decided to use an “anchoring” approach to harmonize the different batches of the single-cell dataset and to integrate the new consensus map. To this end, we took advantage of the R package Seurat (v. 3.0.0.9000). After filtering genes that were expressed in less than 10 cells of the HCA dataset with a cell being kept when 500 genes were detected, we ended up with a large dataset that contained 21,409 genes expressed across 254,937 cells. Next, we merged this Seurat object with the Seurat object of the new consensus map. We treated the different batches of the HCA dataset as individual datasets and normalized them and the expression table of the consensus map

separately. For each dataset, we calculated the top 2,000 most variable genes based on a variance stabilizing transformation followed by data integration by leaving the standard settings unaltered. The integrated dataset was visualized using UMAP based on the top 30 computed PCs. For cell type prediction of the cord blood cells based on the calculated clusters of the new consensus map, we followed the recommendations of the Seurat vignette for the 'FindTransferAnchors' and the 'TransferData' functions. First, we repeated the steps above but without integration of the new consensus map data. We used the resulting integrated HCA dataset as query dataset and the new consensus map as reference dataset. Because of the large cell number of the HCA dataset, we projected the PCA from the query dataset onto the reference dataset. The remaining standard settings were left unaltered. Finally, we transferred the cluster information of the new consensus map onto the query dataset. The resulting prediction scores were visualized as color code onto the UMAP graph by coloring the highest prediction score red. Clustering of the dataset was done based on the construction of an SNN-graph by setting the resolution to 0.6. Cluster 5, 7, 9 and 13 were found to be associated with monocytes or DCs and thus the HCA dataset was filtered on these cells followed by repetition of the abovementioned steps.

**Population-based gene signatures of pDC, pre-DC, cDC1 and cDC2**

Specific gene signatures of up or downregulated in the comparison of human DC subsets have been identified as described earlier using the publicly available dataset (See et al., 2017) (GEO accession number: GSE80171). Gene signatures are available in supplementary Data Table S2.

**Data visualization**

In general, the ggplot2 package was used to generate figures (Wickham, 2016).

## QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using the R programming language. Statistical tests used are described in the figure legend or methods part, respectively. Differentially expressed genes have been identified using a Wilcoxon-based test for differential gene expression. If not otherwise stated a significance level of 0.1 was applied to adjusted p-values (Benjamini Hochberg).

## DATA AND SOFTWARE AVAILABILITY

Processed and raw scRNA-seq datasets are available through the Gene Expression Omnibus (GSE126422). Additional Data tables are provided in form of EXCEL Tables (Data S1, S2)

**Data Table S1: Data Table S1.csv**
Gene signatures of the 11 clusters identified in our new scRNA-seq consensus map

**Data Table S2: Data Table S2.xlsx**
Gene signatures derived from map 2

## ADDITIONAL RESOURCES

In addition, we provide an interactive web tool to visualize the single-cell RNA-Seq data together with the flow cytometry data at https://paguen.shinyapps.io/DC_MONO/ (external database S1).

## References


Alcantara-Hernandez, M., Leylek, R., Wagar, L.E., Engleman, E.G., Keler, T., Marinkovich, M.P., Davis, M.M., Nolan, G.P., and Idoyaga, J. (2017). High-Dimensional Phenotypic Mapping of Human Dendritic Cells Reveals Interindividual Variation and Tissue Specialization. Immunity *47*, 1037-1050 e1036.

Bassler, K., Schulte-Schrepping, J., Warnat-Herresthal, S., Aschenbrenner, A.C., and Schultze, J.L. (2019). The myeloid cell compartment - cell by cell. Ann Rev Immunol https://doi.org/10.1146/annurev-immunol-042718-041728.

Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol *34*, 525-527.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature biotechnology *36*, 411-420.

Calzetti, F., Tamassia, N., Micheletti, A., Finotti, G., Bianchetto-Aguilera, F., and Cassatella, M.A. (2018). Human dendritic cell subset 4 (DC4) correlates to a subset of CD14(dim/-)CD16(++) monocytes. J Allergy Clin Immunol *141*, 2276-2279 e2273.

Cavaillon, J.M. (2011). The historical milestones in the understanding of leukocyte biology initiated by Elie Metchnikoff. Journal of leukocyte biology *90*, 413-424.

Costanzo, M.C., Kim, D., Creegan, M., Lal, K.G., Ake, J.A., Currier, J.R., Streeck, H., Robb, M.L., Michael, N.L., Bolton, D.L., *et al.* (2018). Transcriptomic signatures of NK cells suggest impaired responsiveness in HIV-1 infection and increased activity post-vaccination. Nat Commun *9*, 1212.

Dick, S.A., Macklin, J.A., Nejat, S., Momen, A., Clemente-Casares, X., Althagafi, M.G., Chen, J., Kantores, C., Hosseinzadeh, S., Aronoff, L., *et al.* (2019). Self-renewing resident cardiac macrophages limit adverse remodeling following myocardial infarction. Nature immunology *20*, 29-39.

Dutertre, C.A., Becht, E., Irac, S.E., Khalilnezhad, A., Narang, V., Khalilnezhad, S., Ng, P.Y., van den Hoogen, L.L., Leong, J.Y., Lee, B., *et al.* (2019). Single-cell omics reveal human mononuclear phagocyte heterogeneity and inflammatory DC in health and disease. submitted.

Dutertre, C.A., Jourdain, J.P., Rancez, M., Amraoui, S., Fossum, E., Bogen, B., Sanchez, C., Couedel-Courteille, A., Richard, Y., Dalod, M., *et al.* (2014). TLR3-responsive, XCR1+, CD141(BDCA-3)+/CD8alpha+-equivalent dendritic cells uncovered in healthy and simian immunodeficiency virus-infected rhesus macaques. Journal of immunology *192*, 4697-4708.

Edney, M.H. (2019). Cartography. The Ideal and Its History (Chicago: University of Chicago Press).

Guilliams, M., Dutertre, C.A., Scott, C.L., McGovern, N., Sichien, D., Chakarov, S., Van Gassen, S., Chen, J., Poidinger, M., De Prijck, S., *et al.* (2016). Unsupervised High-Dimensional Analysis Aligns Dendritic Cells across Tissues and Species. Immunity *45*, 669-684.

Guilliams, M., Ginhoux, F., Jakubzick, C., Naik, S.H., Onai, N., Schraml, B.U., Segura, E., Tussiwand, R., and Yona, S. (2014). Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. Nat Rev Immunol *14*, 571-578.

Haniffa, M., Shin, A., Bigley, V., McGovern, N., Teo, P., See, P., Wasan, P.S., Wang, X.N., Malinarich, F., Malleret, B., *et al.* (2012). Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells. Immunity *37*, 60-73.

Hansel, A., Gunther, C., Ingwersen, J., Starke, J., Schmitz, M., Bachmann, M., Meurer, M., Rieber, E.P., and Schakel, K. (2011). Human slan (6-sulfo LacNAc) dendritic cells are inflammatory dermal dendritic cells in psoriasis and drive strong TH17/TH1 T-cell responses. The Journal of allergy and clinical immunology *127*, 787-794 e781-789.

Hulett, H.R., Bonner, W.A., Barrett, J., and Herzenberg, L.A. (1969). Cell sorting: automated separation of mammalian cells as a function of intracellular fluorescence. Science *166*, 747-749.

Hussein, I.H., Raad, M., Safa, R., Jurjus, R., and Jurjus, A. (2015). Once Upon a Microscopic Slide: The Story of Histology. Journal of Cytology & Histology *6*.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods *11*, 163-166.

Kiselev, V.Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. Nature methods *15*, 359-362.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739-1740.

Macosko, Evan Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, Allison R., Kamitaki, N., Martersteck, Emily M., *et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Mass, E., Ballesteros, I., Farlik, M., Halbritter, F., Günther, P., Crozet, L., Jacome-Galarza, C.E., Händler, K., Klughammer, J., Kobayashi, Y., *et al.* (2016). Specification of tissue-resident macrophages during organogenesis. Science *353*, 1114-aaf4238-1111.

Mildner, A., Schonheit, J., Giladi, A., David, E., Lara-Astiaso, D., Lorenzo-Vivas, E., Paul, F., Chappell-Maor, L., Priller, J., Leutz, A., *et al.* (2017). Genomic Characterization of Murine Monocytes Reveals C/EBPbeta Transcription Factor Dependence of Ly6C(-) Cells. Immunity *46*, 849-862 e847.

Monmonier, M. (2015). The History of Cartography, Volume 6, Vol 6 (Chicago: University of Chicago Press).

Perfetto, S.P., Chattopadhyay, P.K., and Roederer, M. (2004). Seventeen-colour flow cytometry: unravelling the immune system. Nat Rev Immunol *4*, 648-655.

Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods *10*, 1096-1098.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nature methods *14*, 979-982.

Ridpath, I. (2007). Norton's Star Atlas and Reference Handbook: And Reference Handbook, Vol 20, 20 edn (New York: Penguin Group).

Rieckmann, J.C., Geiger, R., Hornburg, D., Wolf, T., Kveler, K., Jarrossay, D., Sallusto, F., Shen-Orr, S.S., Lanzavecchia, A., Mann, M., *et al.* (2017). Social network architecture of human immune cells unveiled by quantitative proteomics. Nature immunology *18*, 583-593.

Rozenblatt-Rosen, O., Stubbington, M.J.T., Regev, A., and Teichmann, S.A. (2017). The Human Cell Atlas: from vision to reality. Nature *550*, 451-453.

Satoh, Y., Yokota, T., Sudo, T., Kondo, M., Lai, A., Kincade, P.W., Kouro, T., Iida, R., Kokame, K., Miyata, T., *et al.* (2013). The Satb1 protein directs hematopoietic stem cell differentiation toward lymphoid lineages. Immunity *38*, 1105-1115.

Schakel, K., Poppe, C., Mayer, E., Federle, C., Riethmuller, G., and Rieber, E.P. (1999). M-DC8+ leukocytes--a novel human dendritic cell population. Pathobiology *67*, 287-290.

Schlitzer, A., and Ginhoux, F. (2014). Organization of the mouse and human DC network. Curr Opin Immunol *26*, 90-99.

Schultze, J.L. (2019). Myocardial infarction cell by cell. Nature immunology *20*, 7-9.

Schultze, J.L., Mass, E., and Schlitzer, A. (2019). Emerging principles in myelopoiesis at homeostasis and during infection and inflammation. Immunity *50*.

See, P., Dutertre, C.A., Chen, J., Gunther, P., McGovern, N., Irac, S.E., Gunawan, M., Beyer, M., Handler, K., Duan, K., *et al.* (2017). Mapping the human DC lineage through the integration of high-dimensional techniques. Science.

Shi, W., Liao, Y., Willis, S.N., Taubenheim, N., Inouye, M., Tarlinton, D.M., Smyth, G.K., Hodgkin, P.D., Nutt, S.L., and Corcoran, L.M. (2015). Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. Nature immunology *16*, 663.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research *4*, 1521.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., *et al.* (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nature methods *6*, 377-382.

van der Aa, E., van Montfoort, N., and Woltman, A.M. (2015). BDCA3(+)CLEC9A(+) human dendritic cell function and development. Semin Cell Dev Biol *41*, 39-48.

Villani, A.C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., *et al.* (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis, 2 edn (Springer International Publishing).

Will, B., Vogler, T.O., Bartholdy, B., Garrett-Bakelman, F., Mayer, J., Barreyro, L., Pandolfi, A., Todorova, T.I., Okoye-Okafor, U.C., Stanley, R.F., *et al.* (2013). Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. Nature immunology *14*, 437-445.

Zhang, J.G., Czabotar, P.E., Policheni, A.N., Caminschi, I., Wan, S.S., Kitsoulis, S., Tullett, K.M., Robin, A.Y., Brammananth, R., van Delft, M.F., *et al.* (2012). The dendritic cell receptor Clec9A binds damaged cells via exposed actin filaments. Immunity *36*, 646-657.

Ziegler-Heitbrock, L., Ancuta, P., Crowe, S., Dalod, M., Grau, V., Hart, D.N., Leenen, P.J., Liu, Y.J., MacPherson, G., Randolph, G.J., *et al.* (2010). Nomenclature of monocytes and dendritic cells in blood. Blood *116*, e74-80.

Publication V:

Mass, E.*, Ballesteros, I.*, Farlik, M.*, Halbritter, F.*, **Günther, P.***, Crozet, L., Jacome-Galarza, C.E., Händler, K., Klughammer, J., Kobayashi, Y., Gomez-Perdiguero, E., Schultze, J.L., Beyer, M., Bock, C. and Geissmann, F. (2016) 'Specification of Tissue-Resident Macrophages during Organogenesis.' *Science (New York, N.Y.)*, 353(6304), p. aaf4238. DOI: 10.1126/science.aaf4238.



My contribution to this publication includes:

assistance in library preparation of single-cell MARS-Seq libraries; development and improvement of a bioinformatics pipeline for the preprocessing of MARS-Seq data; quality control and data cleaning of raw data; creating a downstream analysis pipeline including normalization and filtering of data; performing initial analysis of single-cell data; performing advanced analysis of scRNA-seq data; benchmarking different dimensionality reduction and clustering methods; developing a strategy for signature score calculation; applying trajectory analysis of single-cell data; identification of DC-specific signatures from microarray data; interpretation of scRNA-seq results; visualizations for all figures of the scRNA-seq data; writing manuscript

Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238

# Differentiation and specification of tissue-resident macrophages

Elvira Mass[1,#], Ivan Ballesteros[1,#], Matthias Farlik[2,#], Florian Halbritter[2,#], Patrick Günther[3,#], Lucile Crozet[1,4], Christian E. Jacome-Galarza[1], Kristian Händler[3], Johanna Klughammer[2], Yasuhiro Kobayashi[5], Elisa Gomez-Perdiguero [6,7], Joachim L. Schultze[3,8], Marc Beyer[3,8], Christoph Bock [2,9,10], Frederic Geissmann [1,4,6]

[1] Immunology Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA.
[2] CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria
[3] Genomics & Immunoregulation, Life and Medical Sciences Institute, University of Bonn, Bonn, Germany
[4] Weill Cornell Graduate School of Medical Sciences, New York, New York, USA.
[5] Institute for Oral Science, Matsumoto Dental University, 1780 Hiro-Oka Gobara Shiojiri, Nagano, 390-0781 Japan.
[6] Centre for Molecular and Cellular Biology of Inflammation (CMCBI), King's College London, London SE1 1UL, UK
[7] present address: Macrophages and Endothelial Cells group, Department of Developmental and Stem Cell Biology, CNRS UMR 3738, Institut Pasteur, Paris, France
[8] Single Cell Genomics and Epigenomics Unit at the German Center for Neurodegenerative Diseases and the University of Bonn, Bonn, Germany
[9] Department of Laboratory Medicine, Medical University of Vienna, 1090 Vienna, Austria
[10] Max Planck Institute for Informatics, 66123 Saarbrücken, Germany

[#] These authors contributed equally to this work
MB, CB and FG are co-senior authors of this study

***Correspondence***: Frederic Geissmann

**One Sentence summary:** Differentiation of 'core' macrophages is initiated in the E9.5 embryo and followed by their rapid specification into tissue-specific macrophages.

**Abstract:**

**Tissue-resident macrophages are functionally diverse cells that share an embryonic mesodermal origin. However, the mechanism(s) that control their specification remain unclear. We performed transcriptional, molecular and *in situ* spatio-temporal analyses of macrophage development in mice. We report that Erythro-Myeloid Progenitors generate pre-macrophages (pMacs) that simultaneously colonize the head and caudal embryo from embryonic day (E)9.5 in a chemokine-receptor dependent manner, to further differentiate into tissue F4/80$^+$ macrophages. The core macrophage program initiated in pMacs is rapidly diversified as expression of transcriptional regulators becomes tissue-specific in early macrophages from E10.25. This process appears essential for macrophage specification and maintenance, as for example inactivation of *Id3* impairs the development of liver macrophages and results in selective Kupffer cell deficiency in adults. We propose a two-step model where colonization of developing tissues by differentiating macrophages is immediately followed by their specification as they establish residence, hereby generating the macrophage diversity observed in postnatal tissues.**

**Main text**

Tissue-resident macrophages are a diverse family of cells found in most organs, such as brain microglia, liver Kupffer cells, lung alveolar macrophages, and epidermal Langerhans cells. They share an embryonic origin and differentiate, at least in part, from yolk sac (YS) Erythro-Myeloid Progenitors (EMPs) (*1, 2*), and are self-maintained in adult tissues independently of hematopoietic stem cells (HSCs) under steady state conditions (*3-6*). However, the mechanisms responsible for the generation of macrophage diversity observed in adult mice remain unclear. It was proposed that resident macrophage diversity reflects their exposure to specialized tissue environments (*7-10*), or the contribution of distinct embryonic or fetal progenitors to distinct subsets (*2, 11-13*). The preferential expression of transcription factors in macrophage subsets was also noted (*7*), and appears functionally important. Several such cases have been functionally validated by knockout mice, including *Gata6* for large peritoneal macrophages *(9, 10, 14)*, *Runx3* for Langerhans cells *(15)*, *Nr1h3* for splenic marginal zone macrophages *(16)*, *SpiC* for splenic red pulp macrophages *(17)* and *Pparg* for alveolar macrophages (*18*). To better understand how macrophage diversity is generated, we performed a molecular and spatio-temporal analysis of macrophage development in mice.

EMPs (Csf1r$^{+}$ Kit$^{+}$ CD45$^{low}$ AA4.1$^{+}$) are first detected in the YS at E8.5 (*2, 19*) and subsequently colonize the fetal liver (*1, 2*). Previous fate mapping analysis of EMP differentiation indicated that their progeny loses Kit expression and increases CD45 expression as they invade the embryo, before acquiring F4/80 expression to give rise to fetal and postnatal tissue-resident macrophages (*2*) (Fig. 1A). To explore the spatio-temporal and molecular determinants of macrophage differentiation and diversification, we first performed whole transcriptome sequencing of sorted CD45$^{low}$ Kit$^{+}$ EMPs, CD45$^{+}$ Kit$^{-}$ Lin$^{-}$ (Ter119, Gr1, F4/80) cells, and F4/80$^{+}$ macrophages from embryonic and postnatal tissues up to 3 weeks after birth (Fig. 1A, S1A, B). We identified genes that were significantly upregulated between EMPs and CD45$^{+}$ Kit$^{-}$ Lin$^{-}$ cells (adjusted p-value≤0.05, DESeq2 (*20*), Benjamini Hochberg (BH)-correction, Table S1). Subsequent summarization and visualization of these genes via scorecard analysis (*21*) (Fig. 1B), indicated that the CD45$^{+}$ Kit$^{-}$ Lin$^{-}$ cells signature was also present in F4/80$^{+}$ macrophages across tissues and over the entire time course analysis in the embryo and postnatal

mice in the kidney, liver, and brain, albeit epidermal Langerhans cells and lung alveolar macrophage signatures were modified after birth (Fig. 1B). A second scorecard analysis of genes upregulated between EMPs and early (E10.25-E10.5) F4/80$^+$ macrophages identified a signature that was already detectable in CD45$^+$ Kit$^-$ Lin$^-$ cells, and conserved in later macrophages across tissues (Fig. S1C, Table S1). Unsupervised principal component analysis (PCA) showed a distinct grouping of EMPs, CD45$^+$ Kit$^-$ Lin$^-$ cells and macrophages, irrespectively of their tissue of origin *i.e.* YS, liver, head or caudal region (Fig. S1D). Morphologically, CD45$^+$ Kit$^-$ Lin$^-$ cells from the YS, fetal liver, head and caudal embryo displayed a similar morphology, as did macrophages from the same tissues (Fig. 1C). CD45$^+$ Kit$^-$ Lin$^-$ cells resembled EMPs, albeit with the presence of occasional phagocytic vacuoles, while phagocytic features become prominent in F4/80$^+$ cells (Fig. 1C).

These data suggested that a macrophage differentiation program was initiated simultaneously in the whole embryo in CD45$^+$ Kit$^-$ Lin$^-$ cells, which will be referred to below as pre-Macrophage (pMac). To further test this hypothesis, we performed independent unbiased whole transcriptome single-cell RNA sequencing (scRNA-seq) of CD45$^{low/+}$ cells purified from the whole embryo at E10.25 (30-34 somite pairs). Nonlinear dimensionality reduction in combination with unsupervised clustering of cells indicated that these cells are best described by three major clusters (Fig. 1D, Fig. S2). Overlay of EMP, pMac, and macrophage signatures from differentially expressed genes in bulk RNA-seq analysis (Table S1) indicated superimposition on cluster 1, 2, and 3 respectively (Fig. 1E) albeit intermediate differentiation states in EMPs, pMacs, and macrophages were clearly apparent, which suggested a gradual differentiation path from EMP to macrophages via pMacs (Fig. 1E), consistent with the scorecard analysis.

Analysis of genes differentially expressed in EMPs, pMacs, and early macrophages by scRNA-seq (Fig. 1F, G, Table S2) and bulk RNA-seq analysis (Fig. 1H, Table S1) confirmed that a 'core macrophage' transcriptional program was initiated in pMacs. As *Kit*, *Gata1*, and *Gata2* expression was lost, pMacs upregulated expression of *Csf1r*, and the transcription factors *Maf, Batf3, Pparg, Irf8,* and *Zeb2*, the chemokine receptor *Cx3cr1*, cytokine receptors, complement and complement receptors, pattern-recognition receptors, phagocytic receptors, Fc gamma receptors, the inhibitory receptor *Sirpa*, *MerTK*, cathepsins, *Aif1* (Iba1), *Emr1* (F4/80) and *Grn* (Granulin). Expression of

selected cytokine receptors for Il-4, Il-13, Interferon gamma and Tumor necrosis factor, phagocytic and activating receptors Mrc1 (CD206), Trem2, Dectin1 (*Clec7a*), Fc-gamma receptors (Fcgr1, Fcgr2/3, Fcgr4), Iba1, and Grn was confirmed at the protein level by flow cytometry and by immunofluorescence *in situ* on EMP-derived pMacs and early macrophages from the YS as well as from the head, caudal, limbs, and liver of the embryo proper, fetal macrophages and adult tissue macrophages in *Csf1r*$^{MeriCreMer}$*; Rosa26*$^{LSL\text{-}YFP}$ mice pulsed with OH-TAM at E8.5 (Fig. 2A, B, Fig. S3, S4). Expression of these proteins was first detected in pMacs and increased as pMacs differentiated into F4/80$^{+}$ macrophages (Fig. 2A,B, Fig. S3). Of note, pMacs and macrophages represented 70-90% of EMP-derived cells in the head and caudal embryo, while 80-90% of YFP$^{+}$ cells in the E10.25 fetal liver represented progenitors (Fig. 2B, Fig. S4).

Detection of the early expression of the cytokine receptor *Tnfrsf11a* in pMacs by RNA-seq and scRNA-seq analyses (Fig. 1H, 2C) predicted that pMacs or their progeny may express YFP in *Tnfrsf11a*$^{Cre}$ (*22*)*; Rosa26*$^{LSL\text{-}YFP}$ mice. Indeed, we observed YFP labeling by flow cytometry in ~80% pMacs and early macrophages from the YS and embryo of *Tnfrsf11a*$^{Cre}$*; Rosa26*$^{LSL\text{-}YFP}$ mice (Fig. 2D, S5). Fetal macrophages in all tissues also expressed YFP at comparable levels at E10.25 and E14.5 (Fig. 2D). In addition, ~80% of brain, lung, epidermis, kidney, and liver macrophages from 6-week old mice expressed YFP (Fig. 2D). Therefore, the whole resident macrophage lineage is labeled in *Tnfrsf11a*$^{Cre}$*; Rosa26*$^{LSL\text{-}YFP}$ mice, although *Tnfrsf11a* expression itself is lost in postnatal Langerhans cells and alveolar macrophages (Fig. 1H). Moreover, we noted that YFP expression was observed only in ~15% of fetal HSCs, adult HSC, and HSC-derived cells in the blood and tissues of adult mice (Fig. 2D, S5). *Tnfrsf11a*$^{Cre}$*; Rosa26*$^{LSL\text{-}YFP}$ mice thus represent an efficient and relatively specific model for genetic labeling of tissue-resident macrophages in fetuses and adult mice.

Early *Cx3cr1* expression by pMacs (Fig. 1H, 2C) is in line with previous reports showing *Cx3cr1* expression in macrophage precursors (*23*), and suggested that this chemokine receptor may be involved in colonization of embryo tissues by pMacs. Time course analysis of GFP expression in *Cx3cr1*$^{gfp/+}$ mice from E8.5 (19-21 somite pairs - sp) to E10.5 (38-39sp) indicated that GFP expression is not detected in Kit$^{+}$ progenitors, but is upregulated in Kit$^{low}$ Dectin$^{+}$ pMacs and is highest in F4/80$^{+}$ macrophages that appear at E10.25 (Fig. 2E). Next, we found that colonization of the head and caudal

tissues is delayed in *Cx3cr1*-deficient embryos as pMacs and macrophages numbers are decreased in the head and caudal/limbs tissues of E9.5 and E10.5 embryos in comparison with *Cx3cr1*$^{+/-}$ littermates, while they accumulate in the YS and fetal liver (Fig. 2F). Nevertheless, tissue macrophage numbers even out in the consecutive days of embryonic development in most tissues, with the exception of the kidney where a 50% lower number in resident macrophages is still observed in adult mice in line with previous research (*24*) (Fig. S6).

A progressive enrichment of pMac and macrophage specific gene expression signatures was observed in gene set enrichment analysis (GSEA) of bulk RNA-seq data (Fig. S7A,B). Moreover, the 'core' macrophage signature identified in adult mice by the Immgen consortium (*25*) was already enriched in the genes upregulated in pMacs and macrophages compared to EMPs (Fig. S7C,D). *In silico* investigation of transcription factor binding sites identified using ChIP-seq in the proximity of upregulated genes (TSS +/- 20kb) in pMacs and macrophages further supports the proposition that pMacs undergo a coordinated macrophage differentiation program: a LOLA analysis (*26*) yielded a statistically significant association (adjusted p-value≤0.001, Benjamini Yekutieli-correction) with binding sites for *Spi1*, *Egr1, Irf1, Irf8, Maf, Jun, Stat1, Stat3, Stat5b, Stat6, Rela,* and *Relb* (Fig. S8A). These factors are expressed in our dataset (Fig. S8B, Table S1, S5) and their binding sites align at the same loci in enhancers and super-enhancers associated with differentially regulated genes, such as Thrombospondin1 (*Thbs1)* (*27*), *Cx3cr1* (*8*), F4/80 (*Emr1*), and *Mrc1*, but not with control genes such as *Gata1* and *MyoD* (Fig. S8C). Comparable results, with a higher statistical significance, were found for genes differentially upregulated in early macrophages when compared to EMPs (Fig. S8A).

Altogether, these results characterize at the cellular and molecular level the EMP-derived macrophage precursors (pMacs) that acquire a ‘core’ macrophage transcriptional program as they colonize the head and caudal embryo from E9.5 in a *Cx3cr1*-dependent manner to give rise to tissue-resident macrophages. These data are consistent with our previous demonstration that the vast majority of resident macrophages in these tissues originate from a progenitor that does not express the stem cell and endothelial marker *Tie2* after E10.5 (*2*), but do not exclude the possibility that later precursors may also contribute to the resident macrophage pool.

Following the acquisition of this 'core' program, we observed that F4/80$^{+}$ macrophages soon display heterogeneity between different tissues (Fig. 1H). For example, expression of *Timd4* is lost at E12.5 at the transcriptional and protein level in brain macrophages and later microglia, but maintained in liver macrophages and in adult Kupffer cells throughout development, while expression of the Il-4 and Il-13 receptors is lost in postnatal Kupffer cells but maintained in microglia (Fig. 1H, Fig. S3). To systematically investigate the kinetics and molecular determinants of macrophage diversification, we first characterized tissue-specific signatures of genes differentially upregulated in postnatal microglia, kidney macrophages, Langerhans cells, alveolar macrophages and Kupffer cells in the bulk RNA-seq dataset. Unsupervised clustering analysis suggested that macrophages in different tissues undergo characteristic differentiation trajectories (Fig. S9A). Supervised analyses identified lists of genes differentially upregulated in each cell population (Table S3, Fig. S9B, C). The signatures of adult tissue-resident macrophages of the brain, lung, and liver previously defined by independent research (*7, 25, 28*) were progressively enriched in our developing resident macrophage populations (Fig. S10A-D). Finally, scorecard (*21*) and GSEA analysis of differentially upregulated genes for each postnatal macrophage population in all tissues and over time from E9 to P21 (Fig. 3A,B, S10C) revealed that the tissue-specific signature of postnatal microglia, Kupffer cells, and kidney macrophages could be traced back to fetal macrophages, as early as E12.5 (Fig. 3A). The signatures of Langerhans cells and alveolar macrophages reflected important postnatal changes in gene expression (Fig. 3B), noted in previous studies (*29*), and which may reflect their anatomical location at epithelial barriers.

A heat-map visualization of all transcriptional regulators present in the postnatal tissue-specific signatures (2-fold change, adj. p-value<0.05, BH-correction, Fig. 3C) confirmed the tissue-specific expression of the transcription factors *Sall1* and *Sall3* in microglia (*7*), *Nr1h3* (*Lxra*) in Kupffer cells (*30*), *Pparg* in lung alveolar macrophages (*18, 31*), and *Runx3* and *Ahr* in Langerhans cells (*15, 32*) (Fig. 3C), and identified additional tissue-specific transcriptional regulators such as *Id1* and *Id3* in Kupffer cells (Fig. 3C). In addition, this analysis showed that many of these transcriptional regulators start to be differentially expressed in early tissue macrophages, as early as E10.25, for example in *Sall1* and *Sall3* in head macrophages, *Nr1h3, Id1* and *Id3* in the liver, and *Ahr* in limb

macrophages (Fig. 3C). Some genes, like *Id1* and *Sall3* are expressed by progenitors and pMacs before their expression becomes restricted to macrophages in the liver and the head, respectively (Fig. S10E), while expression of other genes such as *Id3* and *Sall1* is low in progenitors and upregulated in pMacs (Fig. S10E). *In situ* immunofluorescence confirmed expression of Id3 by E10.25 macrophages in the liver and head and of Id1 in liver macrophages (Fig. 3D). These data altogether suggested that the transcriptional programs of tissue-specific resident macrophages start to be established early on, as soon as macrophages or pMacs are present in tissues and identified a number of novel 'candidate' tissue-specific transcriptional regulators.

We thus investigated whether specification of tissue macrophages takes place in F4/80$^{+}$ macrophages or at the level of their pMac or EMP precursors. We plotted transcriptional co-expression of *Id1*, *Id3*, and *Sall3* onto the tSNE representation of single CD45$^{low/+}$ cells from our scRNA-seq dataset (see Fig.1D, Fig. 3E, Fig. S11). Co-expression of *Id1* and *Id3* was found in pMacs and macrophages. However, *Id1* and *Sall3* were co-expressed in EMPs and pMacs and *Id3* and *Sall3* were co-expressed in pMacs, suggesting that cells at the pMac state are not completely committed to exclusive expression of tissue-specific transcription factors. These data confirmed that *Id1* and *Sall3* are expressed by progenitors and pMacs (Fig. S10E), while their expression is ultimately lost by tissue macrophages outside the liver and the brain, respectively (Fig. 3C,E and Fig. S10E). When common and tissue-specific E14.5-E18.5 macrophage signatures (Table S2) were superimposed on the pMac population, we found that the common macrophage signature was expressed in pMacs with a gradual enrichment within this population (Fig. 3F). However, we did not observe pMac subsets expressing these tissue-specific signatures (Fig. 3F). The lack of tissue specificity within the pMacs was confirmed using an alternative bioinformatic approach based on analysis of multimodal expression followed by hierarchical clustering of genes with subsequent analysis of enrichment of tissue signatures within these clusters in pMacs (Fig. S11E). Taken together, these data suggest that expression of *Id1/Id3* or *Sall3* does not specify pMac subsets pre-committed to give rise to either Kupffer cells or microglia, and more generally that diversification of tissue macrophage takes place after pMacs have colonized tissues and differentiated into F4/80$^{+}$ macrophages.

Next, to functionally validate the role of early tissue-specific transcriptional regulators on the development and specification of tissue macrophages, and thereby the significance of our findings, we studied the role of *Id3* expression in Kupffer cell differentiation. E10.25 *Id3*-deficient embryos had normal or increased numbers of pMacs/early macrophages in the YS, but macrophages were reduced in numbers in the embryo proper (liver and head) in comparison with littermate controls (Fig. 4A). The further development of liver macrophages was severely impaired in E14.5 and E18.5 *Id3*-deficient embryos as determined by flow cytometry and histology, and 4-week old *Id3*-deficient mice still presented with a marked Kupffer cell deficiency, while development of microglia and kidney macrophages appeared normal (Fig. 4B-D, S12C). The role of *Id3* in Kupffer cells appears cell autonomous as targeted deletion of an *Id3* floxed allele in pMacs (*Tnfrsf11a*$^{Cre+}$*; Id3*$^{f/f}$*)* recapitulated the phenotype of the *Id3*-deficient mice in embryo and postnatal mice (Fig. 4C, E). Expression of *Id3* in postnatal Kupffer cells was confirmed by qRT-PCR (Fig. S12A), and by immunofluorescence in Kupffer cells from *Csf1r*$^{MeriCreMer}$*; Rosa26*$^{LSL\text{-}YFP}$ mice pulse-labeled with OH-Tamoxifen at E8.5 (Fig. 4F, Fig. S12B). In contrast, fate-mapped microglia do not express *Id3* (Fig. S12B), in line with our RNA-seq data (Fig. 3C). Of note, the partial Kupffer cell deficiency observed in *Id3*-deficient and *Tnfrsf11a*$^{Cre+}$*; Id3*$^{f/f}$ embryo and adult was not associated with an abnormal liver lobular architecture or vasculature (Fig. 4C,E, Fig. S12D). Kupffer cell proliferation in the steady state was not affected by *Id3* deficiency (Fig. S12E), but RNA-seq analysis of *Id3*$^{-/-}$ and *Id3*$^{+/-}$ adult Kupffer cells indicated the upregulation (>3-fold) of *Id1* expression, and GO-term analysis evidenced that *Id3*$^{-/-}$ cells overexpressed genes involved in the control of cell death and cytokine responses and down-regulated genes involved in metabolic processes (Fig. 4G, Fig. S12F,G, Table S6). These data suggest that *Id3* is important for the development and maintenance of Kupffer cells in the liver, but dispensable for other macrophage subsets. As *Id1* and *Id3* are co-expressed, it will be interesting to investigate whether up-regulation of *Id1* partially compensates for *Id3*-deficiency (*33, 34*).

We show here that EMPs rapidly differentiate into a population of cells that we call pMacs, because they simultaneously colonize the whole embryo from E9.5 in a *Cx3cr1*-dependent manner while differentiating into macrophages. pMacs do not yet have macrophage morphology but are in the process of establishing a full ‘core’ macrophage differentiation program that includes cytokine receptors, phagocytic and

pattern recognition receptors, and complement. Starting from E10.5, almost immediately after colonization of embryonic tissues, tissue-specific expression of transcriptional regulators in F4/80$^{+}$ tissue macrophages initiates their specification into adult type resident macrophages (Fig. 4H). Our data therefore suggest that a broad or 'core' macrophage program is progressively restricted or refined to a tissue-specific one in response to the absence and presence of tissue-specific cues, rather than the alternative possibility that 'committed' subsets of early macrophages, pMacs or even EMPs choose their future tissue of residence. The present results are consistent with an EMP origin of tissue-resident macrophages (*2*) albeit they do not exclude the additional contribution of 'primitive' precursors (*23*), fetal HSCs or of a second wave of EMPs (*11, 12*), which would adopt a similar differentiation program when entering tissues. However, we note that the novel transcriptional dataset we provide here may lead to re-interpretation of fate-mapping models used to characterize the contribution of such later waves. For example, the progressive YFP expression by fetal macrophages in *S100a4$^{Cre}$; Rosa26$^{LSL\text{-}YFP}$* mice (*12*) is compatible with our present findings, and do not require the existence of a 'second wave' of precursors, because expression of *S100a4* and several other S100a family members was found to be part of the pMac transcriptional profile (Fig. S13).

In summary, we propose a two-step model for resident macrophage differentiation and specification where the establishment of a core macrophage differentiation program in EMP-derived pMacs as they colonize the embryo is followed by the initiation of macrophage specification via the acquisition of tissue specific transcriptional regulators, such as *Id3* in Kupffer cells. This process initiated simultaneously in the whole embryo during the first two days of organogenesis, from E8.5 to E10.5. Differentiation of resident macrophages is thus a developmental process and an integral part of organogenesis, independent of postnatal changes in the environment in kidney, liver, and brain, but not in skin and lung. These results also identify a developmental window where the molecular mechanisms of macrophage specification can be best studied, tools to selectively label resident macrophages, and sets of tissue specific transcriptional regulators expressed by developing and adult macrophages that may control their differentiation, maintenance and function, as well as a molecular road map that will support efforts to differentiate specialized macrophages: microglia, Kupffer cells, kidney macrophages, alveolar macrophages and Langerhans

cells *in vitro* from multipotent progenitors. Finally, our work provides a framework to analyze and understand the consequence(s) of genetic variation for macrophage contribution to disease pathogenesis in different tissues.

## References


1. K. E. McGrath *et al.*, Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell reports* **11**, 1892-1904 (2015).
2. E. Gomez Perdiguero *et al.*, Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature* **518**, 547-551 (2015).
3. D. Hashimoto *et al.*, Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity* **38**, 792-804 (2013).
4. S. Yona *et al.*, Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* **38**, 79-91 (2013).
5. C. Schulz *et al.*, A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science* **336**, 86-90 (2012).
6. L. C. Davies *et al.*, A quantifiable proliferative burst of tissue macrophages restores homeostatic macrophage populations after acute inflammation. *European journal of immunology* **41**, 2155-2164 (2011).
7. Y. Lavin *et al.*, Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312-1326 (2014).
8. D. Gosselin *et al.*, Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327-1340 (2014).
9. Y. Okabe, R. Medzhitov, Tissue-specific signals control reversible program of localization and functional polarization of macrophages. *Cell* **157**, 832-844 (2014).
10. M. Rosas *et al.*, The transcription factor Gata6 links tissue macrophage phenotype and proliferative renewal. *Science* **344**, 645-648 (2014).
11. J. Sheng, C. Ruedl, K. Karjalainen, Most Tissue-Resident Macrophages Except Microglia Are Derived from Fetal Hematopoietic Stem Cells. *Immunity* **43**, 382-393 (2015).
12. G. Hoeffel *et al.*, C-Myb(+) erythro-myeloid progenitor-derived fetal monocytes give rise to adult tissue-resident macrophages. *Immunity* **42**, 665-678 (2015).
13. F. Ginhoux *et al.*, Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330**, 841-845 (2010).
14. E. L. Gautier *et al.*, Gata6 regulates aspartoacylase expression in resident peritoneal macrophages and controls their survival. *The Journal of experimental medicine* **211**, 1525-1531 (2014).

15. O. Fainaru *et al.*, Runx3 regulates mouse TGF-beta-mediated dendritic cell function and its absence results in airway inflammation. *The EMBO journal* **23**, 969-979 (2004).
16. A.-G. N *et al.*, The nuclear receptor LXRalpha controls the functional specialization of splenic macrophages. *Nature immunology* **14**, 831-839 (2013).
17. M. Kohyama *et al.*, Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis. *Nature* **457**, 318-321 (2009).
18. E. L. Gautier *et al.*, Systemic analysis of PPARgamma in mouse macrophage populations reveals marked diversity in expression with critical roles in resolution of inflammation and airway immunity. *J Immunol* **189**, 2614-2624 (2012).
19. J. Y. Bertrand *et al.*, Characterization of purified intraembryonic hematopoietic stem cells as a tool to define their site of origin. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 134-139 (2005).
20. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
21. C. Bock *et al.*, Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439-452 (2011).
22. K. Maeda *et al.*, Wnt5a-Ror2 signaling between osteoblast-lineage cells and osteoclast precursors enhances osteoclastogenesis. *Nature medicine* **18**, 405-412 (2012).
23. J. Y. Bertrand *et al.*, Three pathways to mature macrophages in the early mouse yolk sac. *Blood* **106**, 3004-3011 (2005).
24. M. S. Lionakis *et al.*, CX3CR1-dependent renal macrophage survival promotes Candida control and host survival. *The Journal of clinical investigation* **123**, 5035-5051 (2013).
25. E. L. Gautier *et al.*, Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nature immunology* **13**, 1118-1128 (2012).
26. N. C. Sheffield, C. Bock, LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*, (2015).
27. W. A. Whyte *et al.*, Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319 (2013).
28. N. N. Gorgani, Y. Ma, H. F. Clark, Gene signatures reflect the marked heterogeneity of tissue-resident macrophages. *Immunology and cell biology* **86**, 246-254 (2008).
29. L. Chorro *et al.*, Langerhans cell (LC) proliferation mediates neonatal development, homeostasis, and inflammation-associated expansion of the epidermal LC network. *The Journal of experimental medicine* **206**, 3089-3100 (2009).
30. M. Hoekstra, J. K. Kruijt, M. Van Eck, T. J. Van Berkel, Specific gene expression of ATP-binding cassette transporters and nuclear hormone receptors in rat

liver parenchymal, endothelial, and Kupffer cells. *The Journal of biological chemistry* **278**, 25448-25453 (2003).

31. C. Schneider *et al.*, Induction of the nuclear receptor PPAR-gamma by the cytokine GM-CSF is critical for the differentiation of fetal monocytes into alveolar macrophages. *Nature immunology* **15**, 1026-1037 (2014).
32. B. Jux, S. Kadow, C. Esser, Langerhans cell maturation and contact hypersensitivity are impaired in aryl hydrocarbon receptor-null mice. *J Immunol* **182**, 6709-6717 (2009).
33. Y. Jen, K. Manova, R. Benezra, Expression patterns of Id1, Id2, and Id3 are highly related but distinct from that of Id4 during mouse embryogenesis. *Developmental dynamics : an official publication of the American Association of Anatomists* **207**, 235-252 (1996).
34. D. Lyden *et al.*, Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature* **401**, 670-677 (1999).

**Author contribution**. FG, EM designed the study and wrote the manuscript. EM, IB performed cell sorting, flow cytometry, fate-mapping, immunostaining experiments and RNA-seq analysis on *Id3*-deficicient and control Kupffer cells. LC and CEJG helped with analysis of *Tnfrsf11a*$^{Cre+}$*; Rosa26*$^{LSL-YFP}$ mice. EGP assisted with the design of cell sorting experiments. MF designed and prepared bulk RNA-seq libraries. FH, JK, CB performed primary and differential analysis of the bulk RNA-seq data. KH and MB generated single-cell RNA-seq libraries and performed single-cell RNA-seq, PG, MB and JLS analyzed single cell RNA-seq data. All authors contributed to the manuscript.

**Acknowledgement.** This work was supported by the National Cancer Institute of the US National Institutes of Health (P30CA008748), and by Wellcome Trust investigator (WT101853MA), and European Research Council investigator awards (2010-StG-261299) to FG. EM is supported by an EMBO long-term Fellowship. FH is supported by a postdoctoral fellowship of the German Research Foundation (DFG). CB is supported by a New Frontiers Group award of the Austrian Academy of Sciences and by an ERC Starting Grant. MB and JLS are members of the Excellence Cluster ImmunoSensation. M.B. is supported by a DFG grant (BE 4427/3-1), JLS is supported by DFG grants SFB 704, SFB 645, INST 217/575-1, INST 217/576-1 and INST 217/577-1. The authors are indebted to R. Benezra, Memorial Sloan Kettering Cancer Center, New York for the *Id3*$^{-/-}$ and *Id3*$^{f/f}$ strains. We thank the Biomedical Sequencing Facility at CeMM for assistance with next generation sequencing. The authors wish to acknowledge assistance from the MSKCC Bioinformatics Core, which is funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748.

## Figures legends

**Fig. 1. A 'core' macrophage program is initiated simultaneously in pMacs in all tissues. (A)** Summary of surface phenotype used for EMPs, pMacs, and macrophages. **(B)** Scorecard visualization of differentially upregulated genes (DESeq2 Wald test, adjusted p-value < 0.05, BH-correction) in pMacs (E9.5 and E10.25) in comparison to EMPs. The table shows the relative enrichment of differentially upregulated genes in pMacs across cell types and tissues (y-axis) and developmental time points (x-axis, from E9 to P21). See Table S1, Fig. S1, and Methods for details of the scorecard. **(C)** May-Gruenwald-Giemsa stained cytospin preparations of sorted EMPs, pMacs and early macrophages from yolk sac (YS), head, limbs and fetal liver (FL) at E10.25 and E12.5. n=3 independent experiments. **(D)** tSNE plot of scRNA-seq data showing distribution of CD45$^{low/+}$ cells from E10.25 embryos into three clusters (see also Fig. S2). Cluster distribution based on DBScan is overlaid onto the graph. **(E)** Superimposition of EMP-, pMac-, or macrophage-specific signatures defined by the bulk RNA sequencing on the tSNE plot shown in D. **(F)** tSNE plot as in (D) overlaid with the relative expression values for *Kit* and *Maf*. **(G)** (**upper panel**) Developmental pseudotime diagram (q-value<0.05) showing down regulation of EMP-specific genes (differentially expressed compared to the macrophage and pMac cluster, p-value<0.05, FC>1.4) over the differentiation path from EMP to pMacs and macrophages. (**lower panel**) Similar plot depicting macrophage-specific genes significantly regulated over pseudotime (q-value<0.05) and differentially expressed compared to the EMP and pMac cluster (p-value<0.05, FC>1.4)). See Fig. S2G. **(H)** Heatmap representation of selected genes differentially regulated between EMPs vs. pMacs and EMPs vs. early macrophages in bulk RNA-seq analysis. Black boxes were drawn around those samples used for differential expression analysis. See also Table S1, S2.

**Fig. 2. EMP-derived pMacs colonize the embryo to generate macrophages. (A)** Flow cytometry analysis of E10.25 *Csf1r*$^{MeriCreMer}$; *Rosa26*$^{LSL-YFP}$ (OH-TAM at E8.5) tissues showing expression of Il4ra, Il13ra1, CD16.2, CD64, Ifngr, Tnfr2, Tim4, and CD206 on YFP$^{+}$ Kit$^{+}$ progenitors, pMacs, and macrophages. MFI: mean fluorescent intensity. Data are representative of n=4 independent experiments with 4-6 embryos per marker. See also Fig. S3A. **(B)** Quantification of immunostainings on cryosection from E10.25 *Csf1r*$^{MeriCreMer}$; *Rosa26*$^{LSL-YFP}$ embryos, pulse-labeled with OH-TAM at E8.5 with antibodies against YFP, Iba1 and CD16/32, Dectin-1, Trem2, F4/80, CD206 or Granulin. n=2-4 embryos and 2 sections per embryo per marker. See Fig. S4. **(C)** tSNE plot as in (1D) overlaid with the relative expression values for *Tnfrsf11a* and *Cx3cr1*. **(D)** YFP labeling efficiency of *Tnfrsf11a*$^{Cre+}$; *Rosa26*$^{LSL-YFP}$ in pMacs and F4/80$^{+}$ macrophages in YS and whole embryo at E10.25, fetal liver HSCs (long term (LT, Lin$^{-}$ Kit$^{+}$Sca1$^{+}$CD150$^{+}$CD48$^{-}$), short term (ST, Lin$^{-}$Kit$^{+}$Sca1$^{+}$CD150$^{-}$CD48$^{-}$) and multipotent progenitor (MPP, Lin$^{-}$Kit$^{+}$Sca1$^{+}$CD150$^{-}$CD48$^{+}$)) and tissue macrophages at E14.5 and 6 weeks, and blood leukocytes (B-cells (CD19$^{+}$), T-cells (CD19$^{-}$Ly6G$^{-}$CD115$^{-}$CD3$^{+}$), NK cells (CD19$^{-}$Ly6G$^{-}$CD115$^{-}$CD3$^{-}$NKp46$^{+}$), neutrophils (CD19$^{-}$Ly6G$^{+}$) and Ly6C$^{hi}$ monocytes (CD19$^{-}$CD115$^{+}$Ly6G$^{-}$Ly6C$^{hi}$), and tissue CD11b$^{high}$ myeloid cells from 6 week-old mice. Circles represent individual mice. n=4 independent experiments. See Fig. S5. **(E)** Expression of GFP and Dectin-1 in *Cx3cr1*$^{gfp/+}$ mice during development (E8.5-10.5) in Kit$^{+}$ cells (CD45$^{low}$, Kit$^{+}$), pMacs and macrophages. sp: somite pairs. Data are representative of n=9 independent experiments. Biological replicates have been aggregated per cell type, time point and tissue. **(F)** Flow cytometry analysis in *Cx3cr1*$^{+/-}$ and *Cx3cr1*$^{-/-}$ of pMacs and macrophages from yolk sac (YS), head, and caudal at E9.5 (upper panel) and liver, YS, head, and limbs at E10.5 (lower panel). Circles represent individual mice. Data are representative of n=6 independent experiments. sp: somite pairs.

**Fig. 3. Early specification of tissue-resident macrophages**. **(A, B)** Scorecard analysis of all differentially upregulated genes in postnatal macrophages. The scorecards show the relative enrichment of each set of upregulated genes across each cell type (y-axis) and developmental time point (x-axis). See Methods for details of the score card. Numbers for each population indicate differentially up-regulated transcripts in postnatal (P2-P21) brain, liver, kidney, epidermis or lung macrophages when comparing one population vs. the others. See also Table S3. **(C)** Heatmap representing all differentially upregulated transcriptional regulators (2-fold change, adj. p-value<0.05, BH-correction) between postnatal macrophages from brain, liver, kidney, skin and lung macrophages, and their relative expression in tissue macrophages from E10.25 to P21. **(D)** Immunostaining with antibodies against Id1 or Id3 (red), F4/80 (cyan) and YFP (green) on cryosections from E10.25 *Csf1r*$^{MeriCreMer}$; *Rosa26*$^{LSL-YFP}$ embryos (OH-TAM at E8.5) (upper panel). Nuclei are counterstained with DAPI (white). Scale bar represents 2 µm. **(E)** tSNE plots of scRNA-seq data from CD45$^{low/+}$ cells from E10.25 embryo showing co-expression of *Id1, Id3*, and *Sall3*. See Fig. S11. **(F)** PCA plot of scRNA-seq data of cells from cluster 2 (pMacs) with superimposed fetal tissue macrophage-specific signatures. See Fig. S11, Table S2 and methods.

**Fig. 4. Role of Id3 in development of Kupffer cells. (A)** Flow cytometry analysis of pMacs and macrophages in yolk sac (YS) and liver from E10.25 *Id3*$^{-/-}$ and *Id3*$^{+/-}$ embryos. Circles represent individual mice. n=3 independent experiments. **(B)** Flow cytometry analysis of F4/80$^{+}$ macrophages in liver, brain and kidney from E14.5 and E18.5 *Id3*$^{-/-}$ and *Id3*$^{+/-}$ mice. Circles represent individual mice. n=4 independent experiments. **(C)**

Immunostaining with antibodies against CD31 and F4/80 on liver cryosections from E14.5 *Id3*$^{-/-}$ and *Id3*$^{+/-}$, and *Tnfrsf11a*$^{Cre-}$*;Id3*$^{f/+}$ and *Tnfrsf11a*$^{Cre+}$*; Id3*$^{f/f}$ mice. The figure displays isovolume-rendered images. Bar graphs represent F4/80$^{+}$ cells/mm$^2$. Circles represent individual images. **(D)** Flow cytometry analysis of F4/80$^{+}$ macrophages in liver, brain and kidney from 4 week-old *Id3*$^{-/-}$ and *Id3*$^{+/-}$ mice. Circles represent individual mice. n=2 independent experiments. **(E)** Immunostaining with antibodies against CD31 and F4/80 on liver cryosections from 2 week-old *Tnfrsf11a*$^{Cre+}$*;Id3*$^{f/+}$ and *Tnfrsf11a*$^{Cre+}$*; Id3*$^{f/f}$ mice. The figure displays isovolume-rendered images. Bar graphs represent F4/80$^{+}$ cells/mm$^2$. Circles represent individual images. **(F)** Immunostaining with antibodies against Id3 (red), F4/80 (cyan) and YFP (green) on cryosections from on livers from 4 week-old *Csf1r*$^{MeriCreMer}$; *Rosa26*$^{LSL-YFP}$ (OH-TAM at E8.5) mice. Nuclei are counterstained with DAPI (white). Scale bar represents 5 um. **(G)** Scatterplot comparison of gene expression of 3 week-old *Id3*$^{-/-}$ and *Id3*$^{+/-}$ Kupffer cells. Both axes (in $\log_2$ scale) represent normalized gene expression values (average value from three *Id3*$^{+/-}$ and two *Id3*$^{-/-}$ replicates). Red circles mark the 3-fold cut-off in both directions in gene expression level. Top GO terms for genes enriched in either *Id3*$^{+/-}$ or *Id3*$^{-/-}$ are indicated. See also Fig. S12. **(H)** Graphic summary of the establishment of the core macrophage program and subsequent specification.

Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238

Figure 1

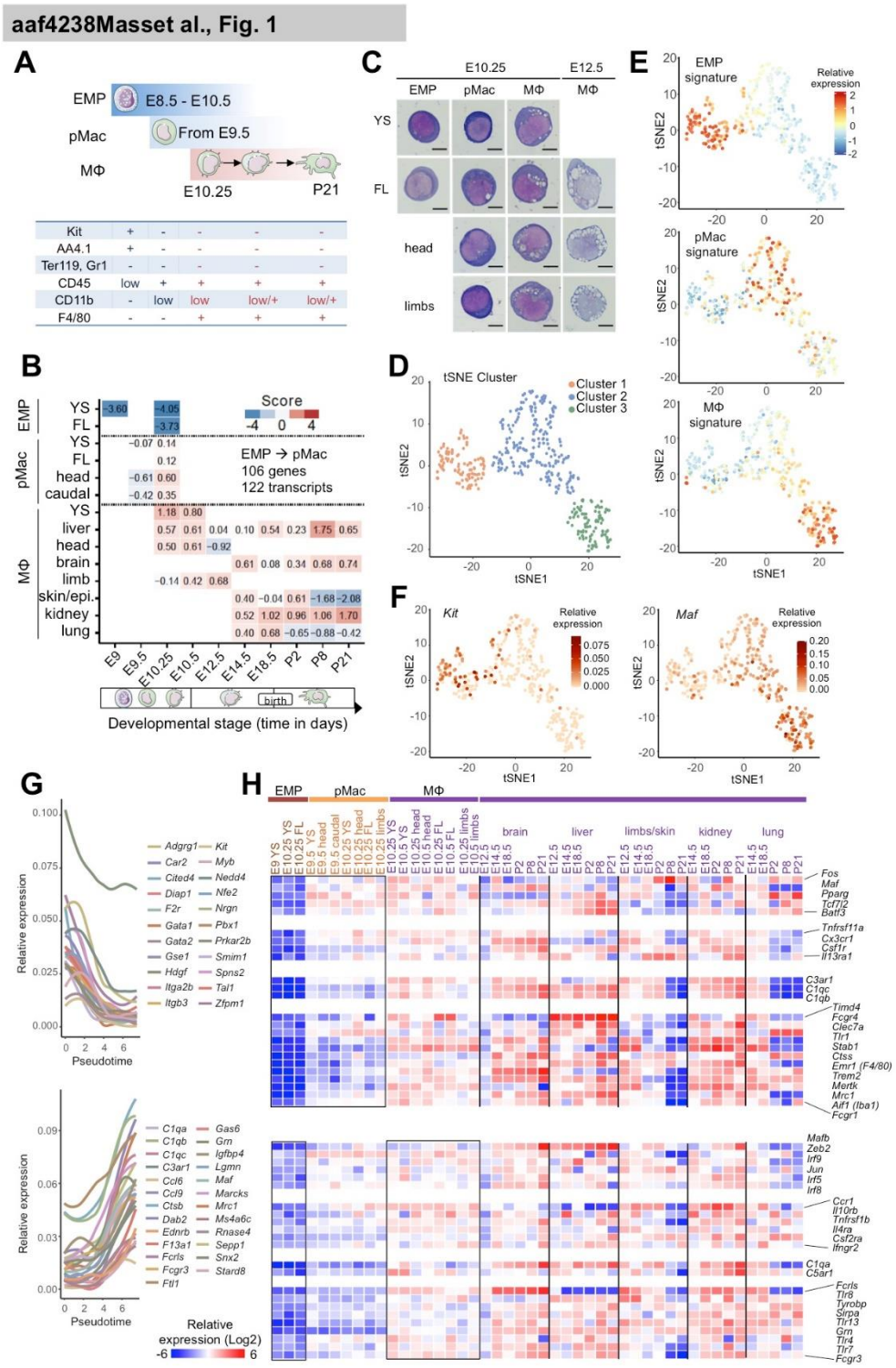
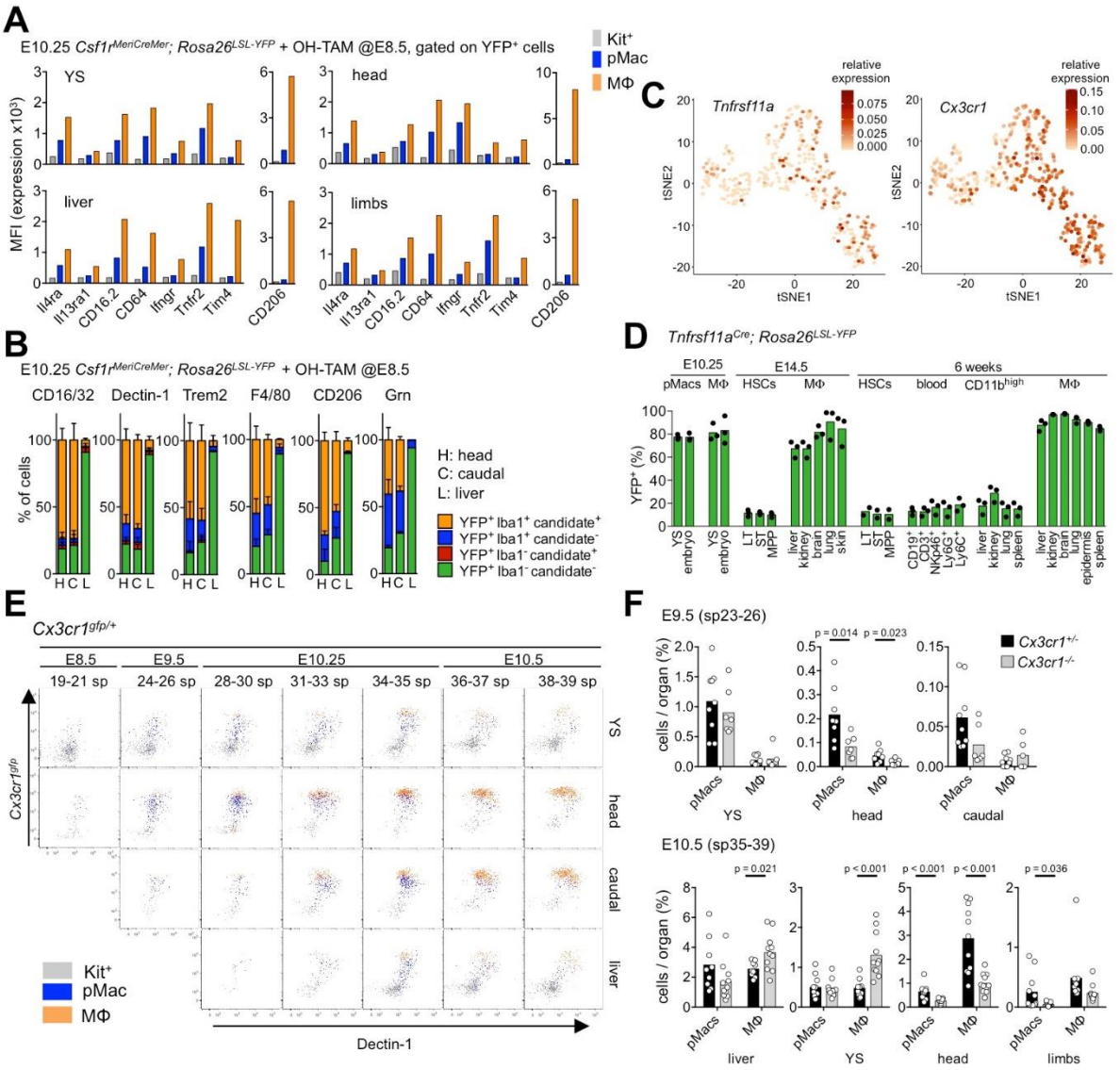


Figure 2


aaf4238Masset al., Fig. 2
A
E10.25 Csf1r MeriCreMer; Rosa26 LSL-YFP + OH-TAM @E8.5, gated on YFP+ cells
Kit+
pMac
MΦ
YS
head
liver
limbs
MFI (expression x10³)
Il4ra
Il13ra1
CD16.2
CD64
Ifngr
Tnfr2
Tim4
CD206
B
E10.25 Csf1r MeriCreMer; Rosa26 LSL-YFP + OH-TAM @E8.5
CD16/32
Dectin-1
Trem2
F4/80
CD206
Grn
% of cells
H C L
H: head
C: caudal
L: liver
YFP+ Iba1+ candidate+
YFP+ Iba1+ candidate-
YFP+ Iba1- candidate+
YFP+ Iba1- candidate-
C
Tnfrsf11a
Cx3cr1
relative expression
tSNE1
tSNE2
D
Tnfrsf11a Cre; Rosa26 LSL-YFP
E10.25
E14.5
6 weeks
pMacs MΦ
HSCs
MΦ
HSCs
blood
CD11b high
MΦ
YFP+ (%)
YS embryo
LT HSC
ST HSC
MPP
liver
kidney
brain
lung
skin
CD19+
CD3+
NK46+
Ly6G+
Ly6C+
spleen
epidermis
E
Cx3cr1 gfp/+
E8.5
E9.5
E10.25
E10.5
19-21 sp
24-26 sp
28-30 sp
31-33 sp
34-35 sp
36-37 sp
38-39 sp
YS
head
caudal
liver
Cx3cr1 gfp
Dectin-1
Kit+
pMac
MΦ
F
E9.5 (sp23-26)
Cx3cr1 +/-
Cx3cr1 -/-
p = 0.014
p = 0.023
cells / organ (%)
pMacs
MΦ
YS
head
caudal
E10.5 (sp35-39)
p = 0.021
p < 0.001
p < 0.001
p < 0.001
p = 0.036
liver
YS
head
limbs

Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238

Figure 3

**aaf4238Masset al., Fig. 3**

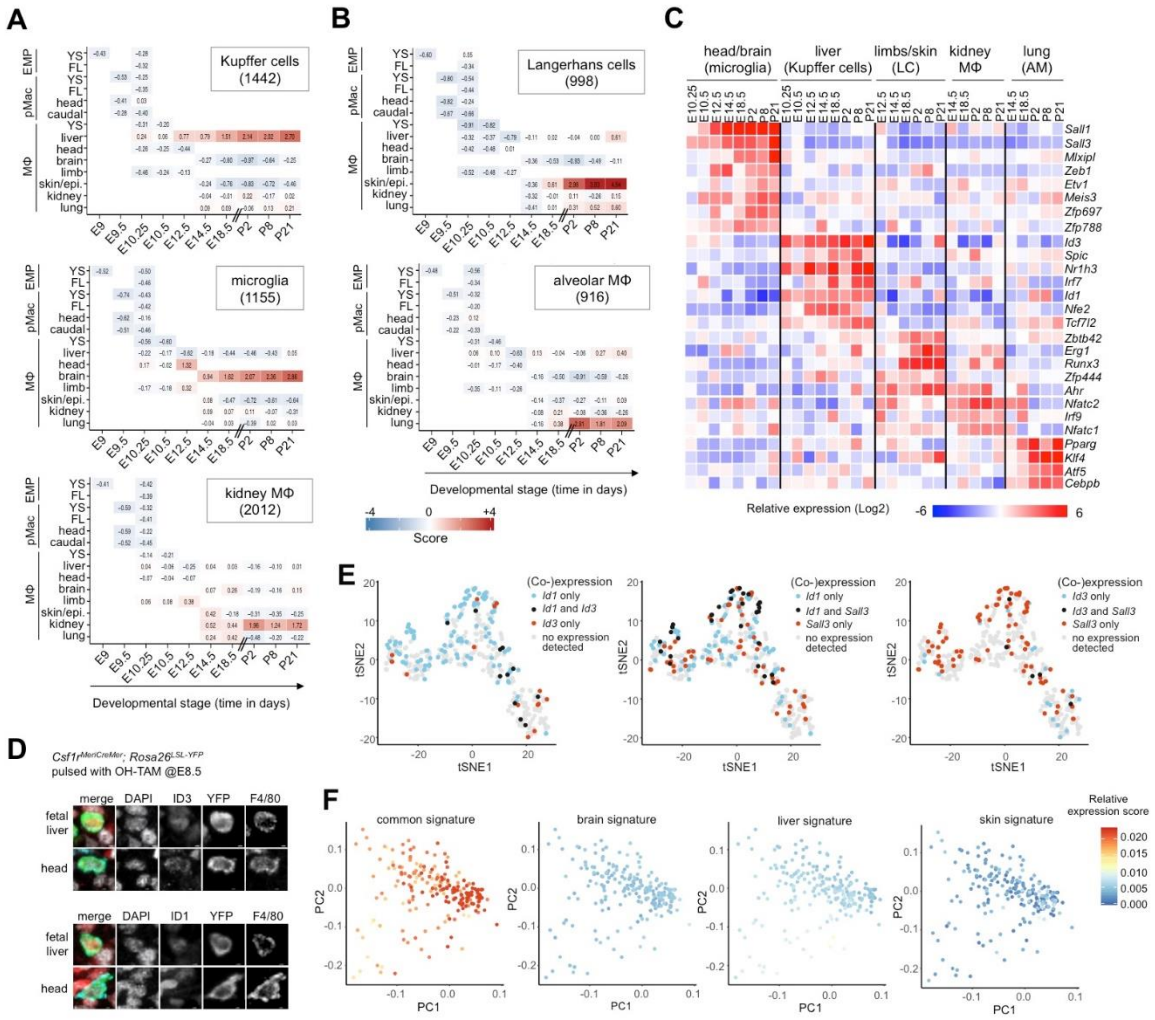
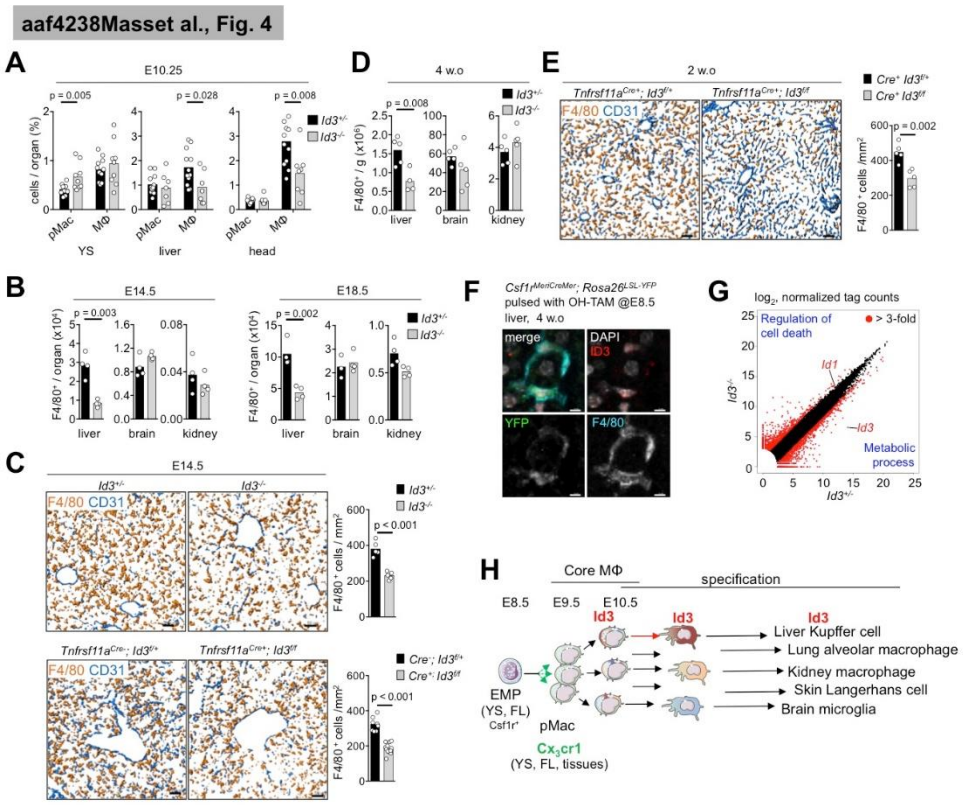
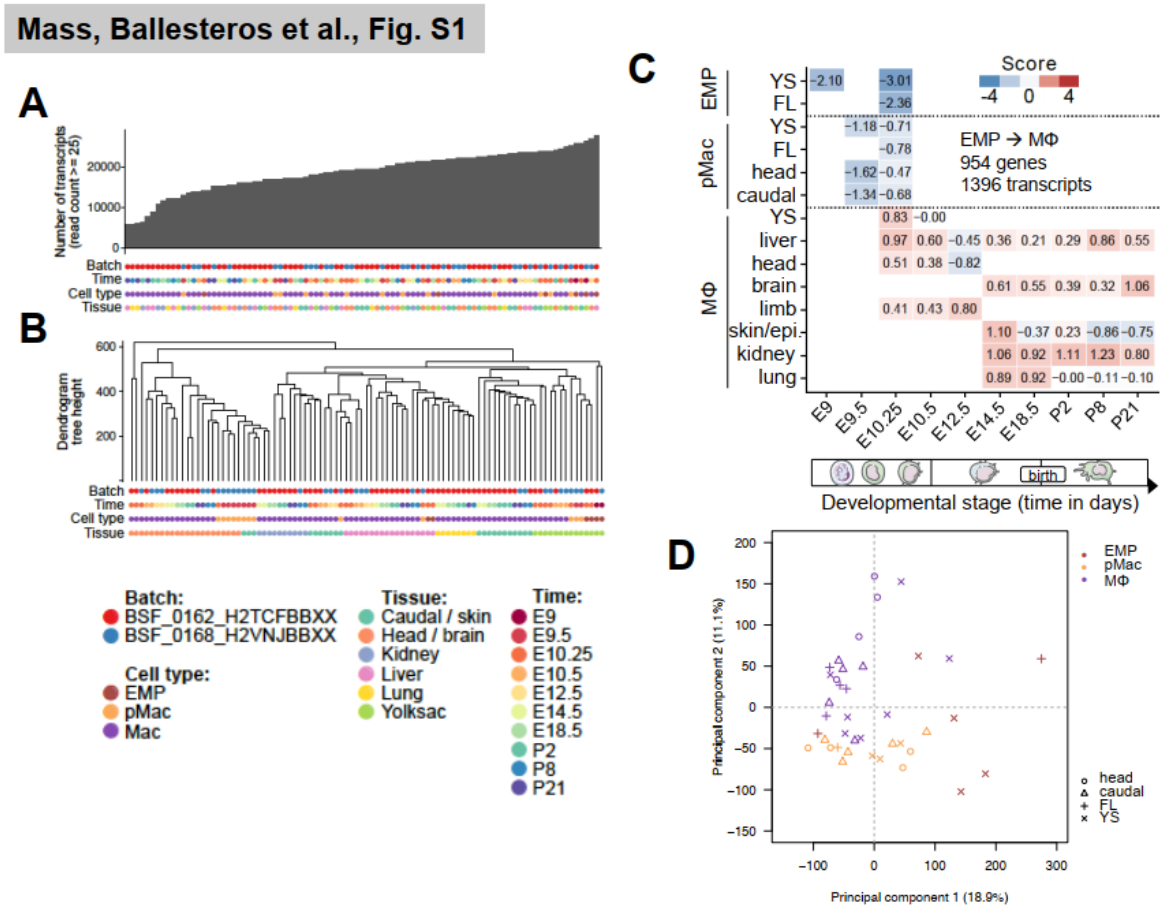


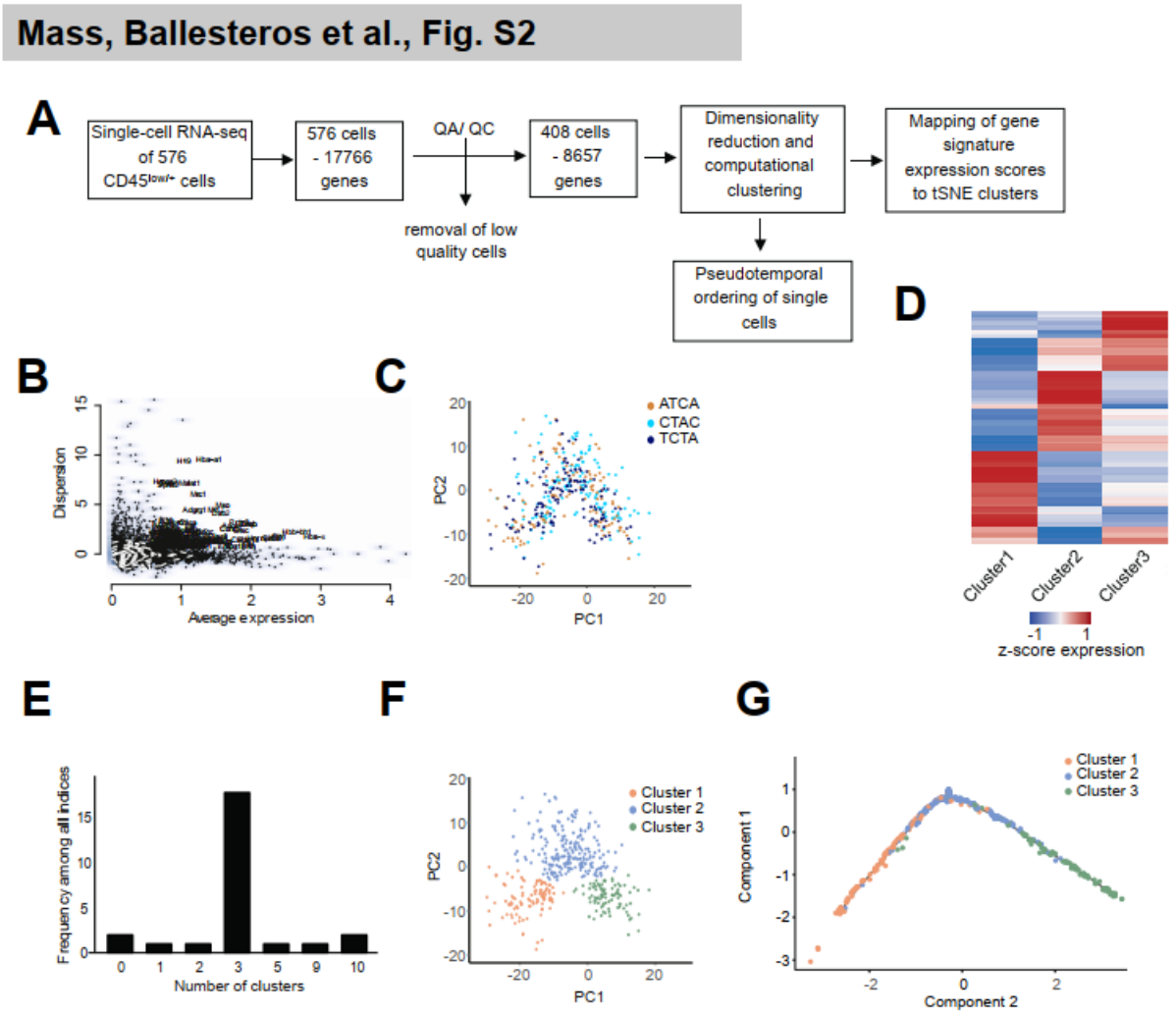
Figure 4


aaf4238Masset al., Fig. 4
A
E10.25
p = 0.005
p = 0.028
p = 0.008
cells / organ (%)
pMac
Mφ
YS
liver
head
D
4 w.o
p = 0.008
F4/80+ / g (x10⁶)
liver
brain
kidney
E
2 w.o
F4/80 CD31
p = 0.002
F4/80+ cells /mm²
B
E14.5
E18.5
p = 0.003
p = 0.002
F4/80+ / organ (x10⁴)
C
E14.5
p < 0.001
F4/80+ cells / mm²
F
merge
DAPI
Id3
YFP
F4/80
pulsed with OH-TAM @E8.5
liver, 4 w.o
G
log2, normalized tag counts
Regulation of cell death
> 3-fold
Id1
Id3
Metabolic process
H
Core MΦ
specification
E8.5
E9.5
E10.5
EMP
(YS, FL)
Csf1r+
pMac
Cx3cr1
(YS, FL, tissues)
Liver Kupffer cell
Lung alveolar macrophage
Kidney macrophage
Skin Langerhans cell
Brain microglia

Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238

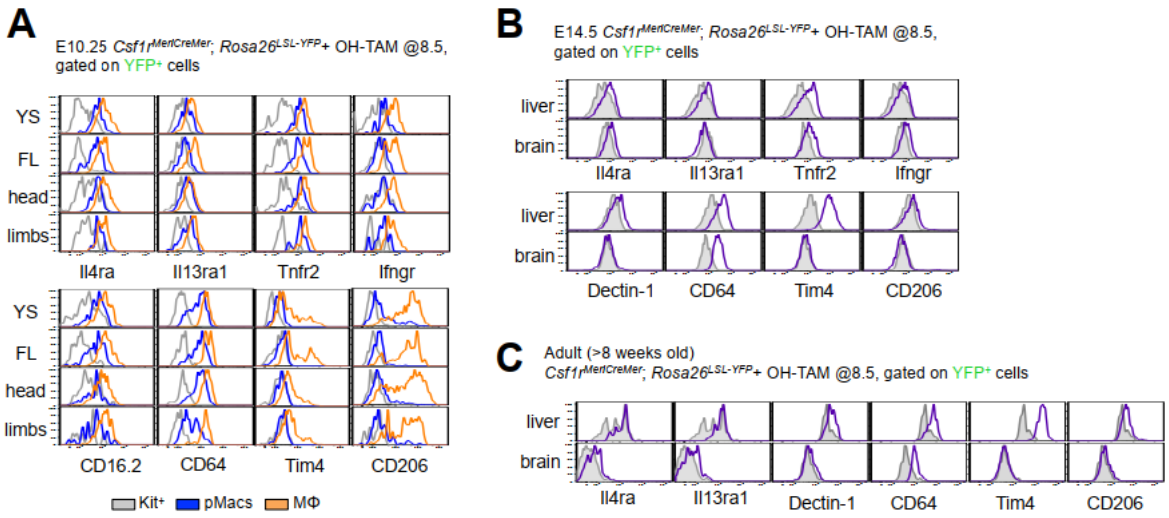


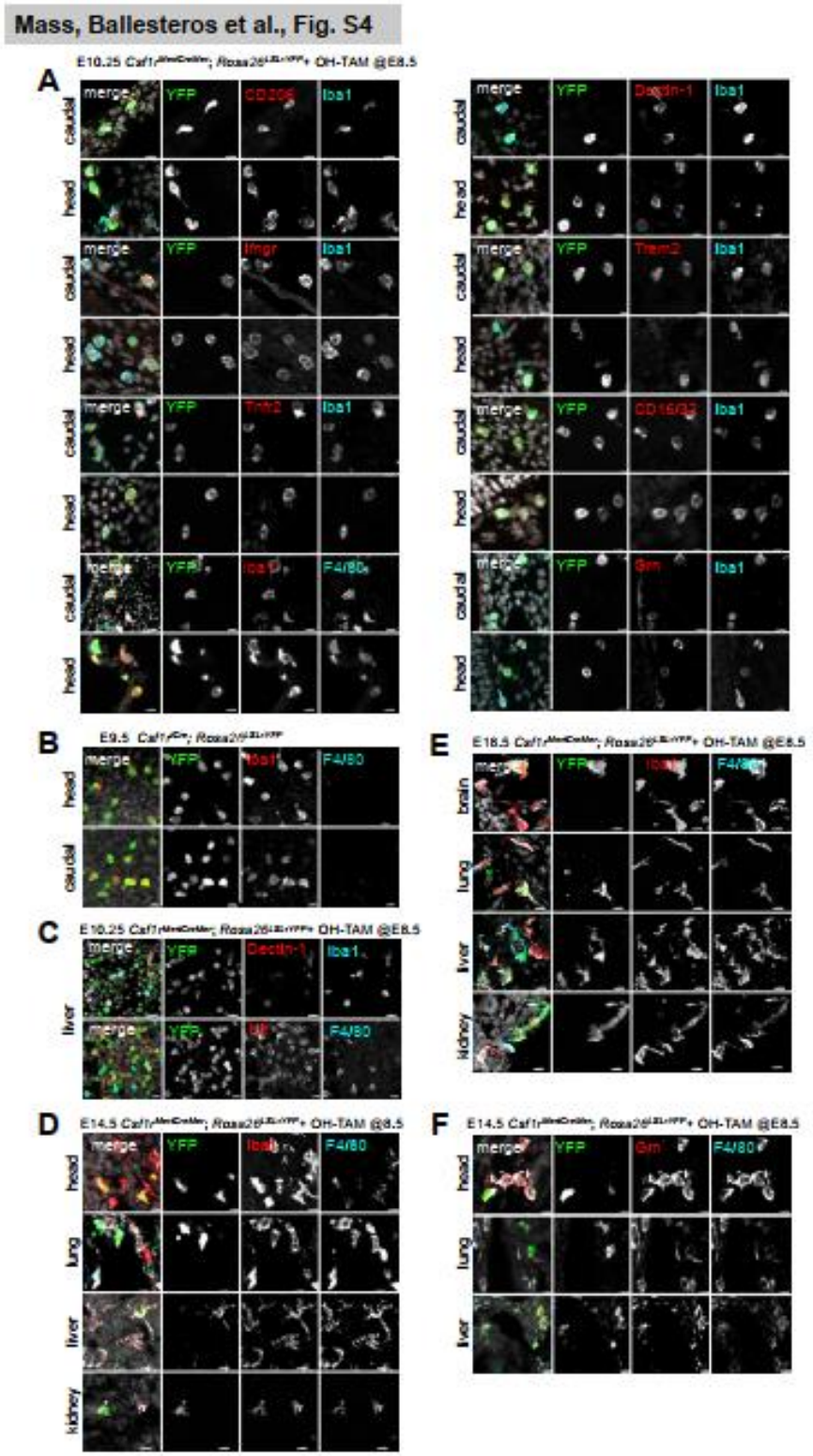
Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238

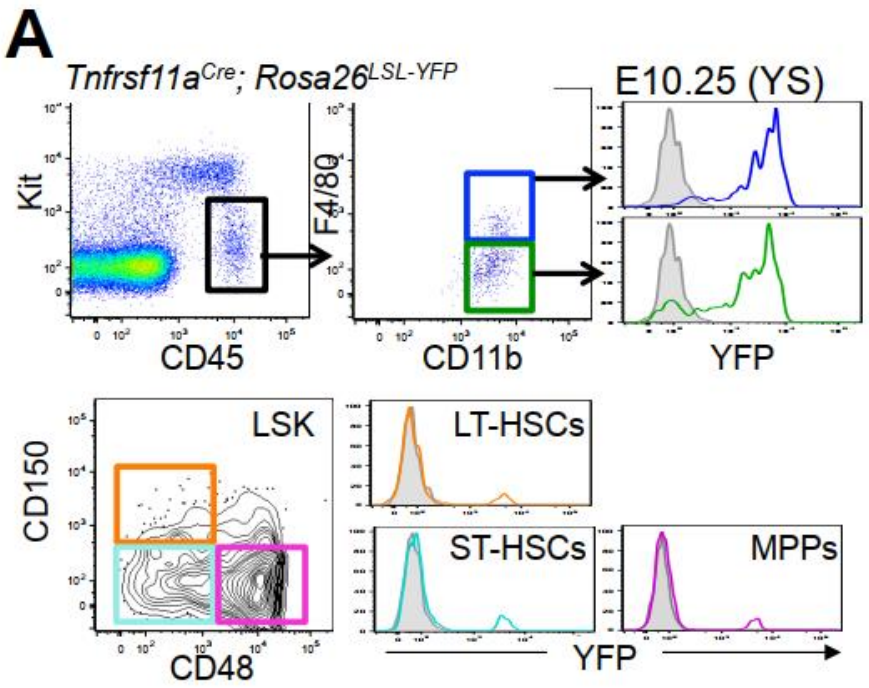


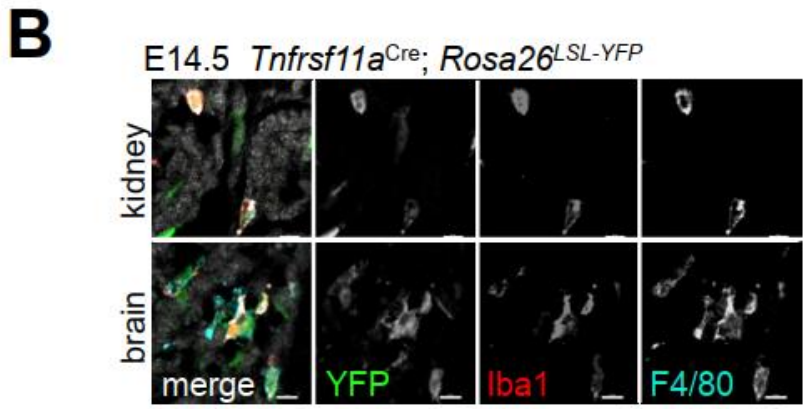
Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238

**Mass, Ballesteros et al., Fig. S3**



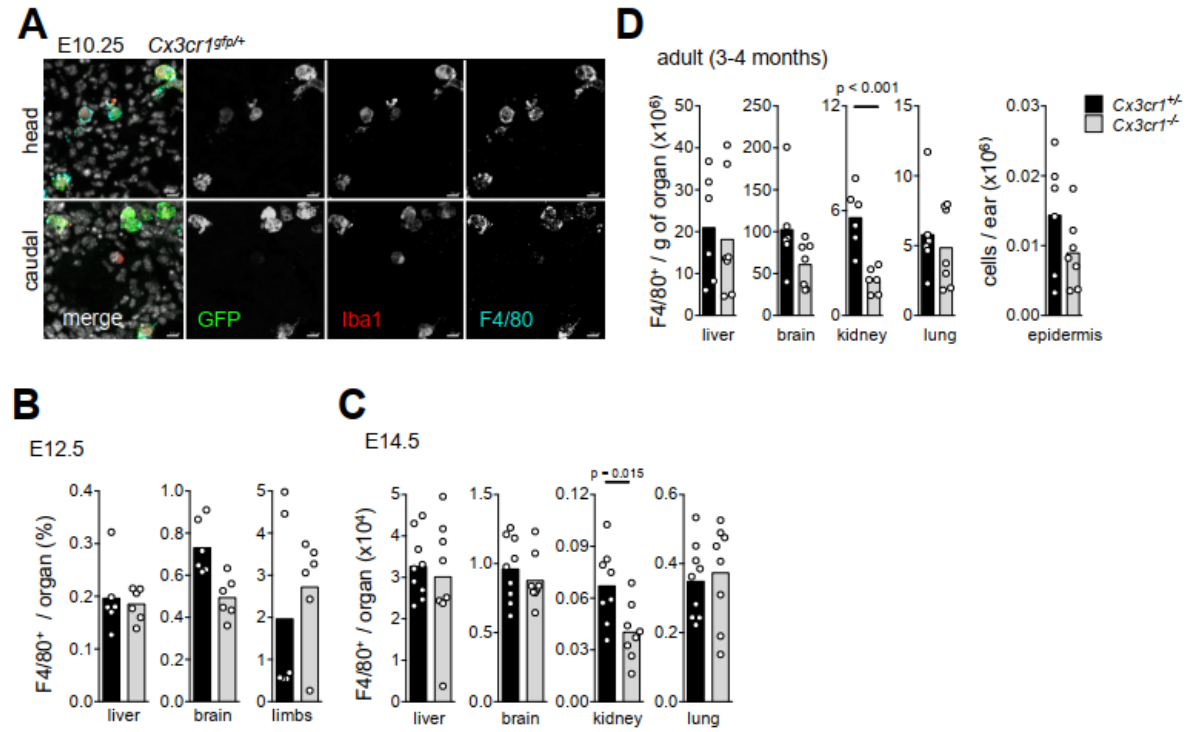

Mass, Ballesteros et al., Fig. S4
A
B
C
D
E
F
merge
YFP
Iba1
F4/80
Dectin-1
Ifngr
Trem2
CD169
Sirp
brain
lung
liver
kidney
head
caudal


A
Tnfrsf11a^Cre; Rosa26^LSL-YFP
E10.25 (YS)
Kit
CD45
F4/80
CD11b
YFP
LSK
CD150
CD48
LT-HSCs
ST-HSCs
MPPs
YFP

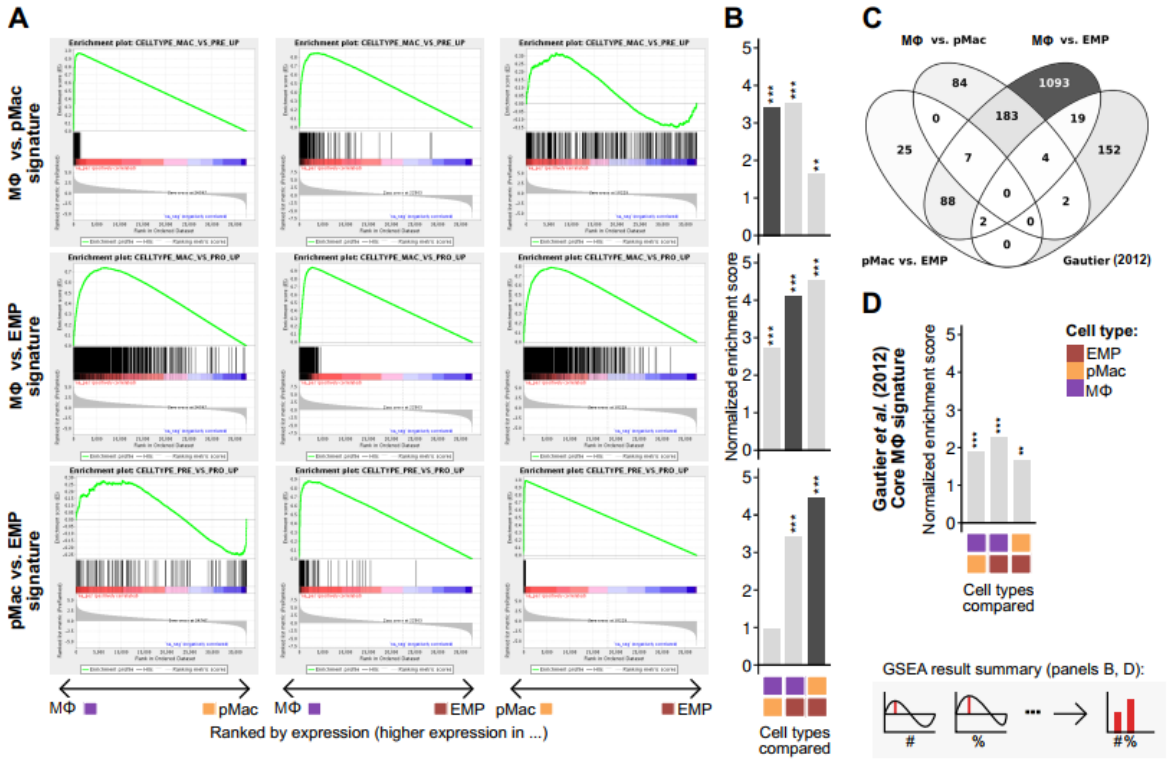

B
E14.5 Tnfrsf11a^Cre; Rosa26^LSL-YFP
kidney
brain
merge
YFP
Iba1
F4/80

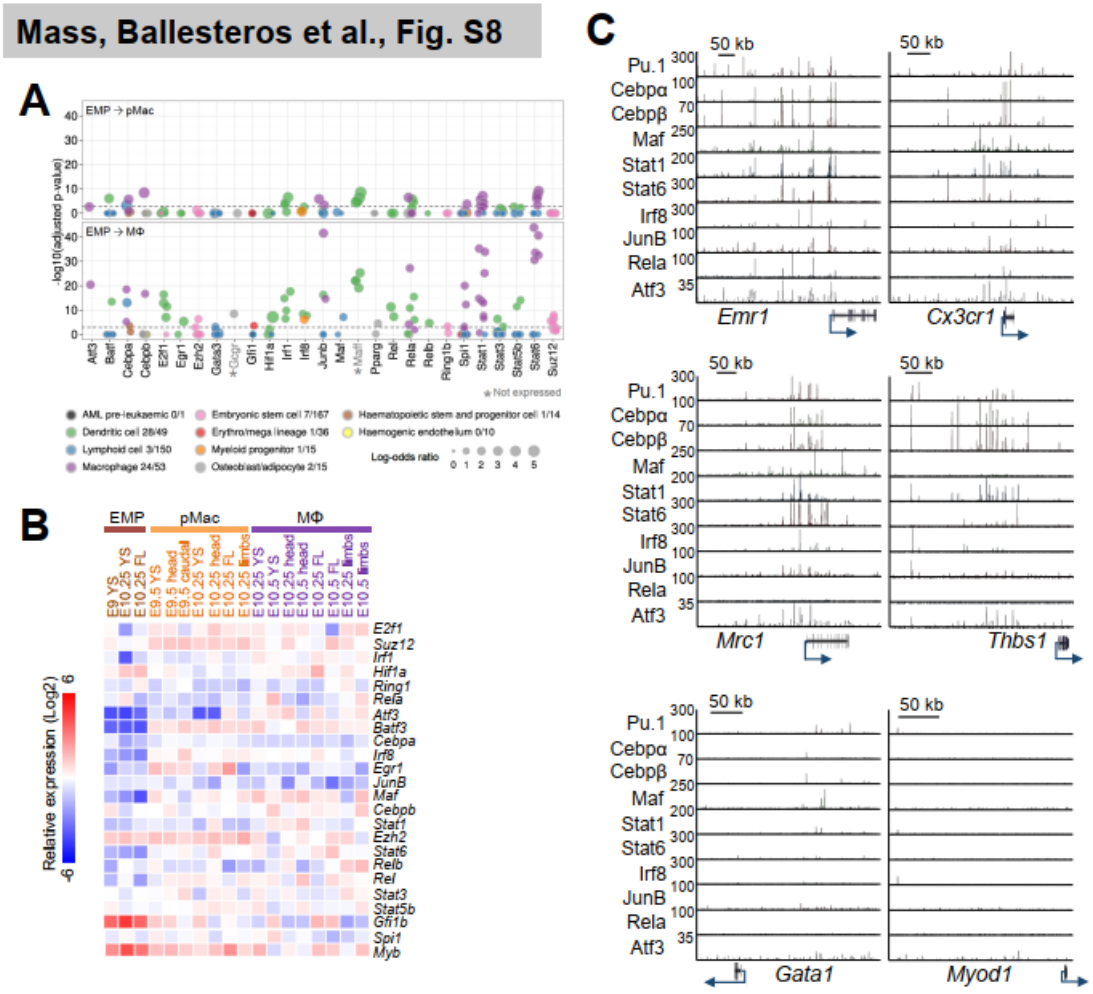
Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238

**Mass, Ballesteros et al., Fig. S6**



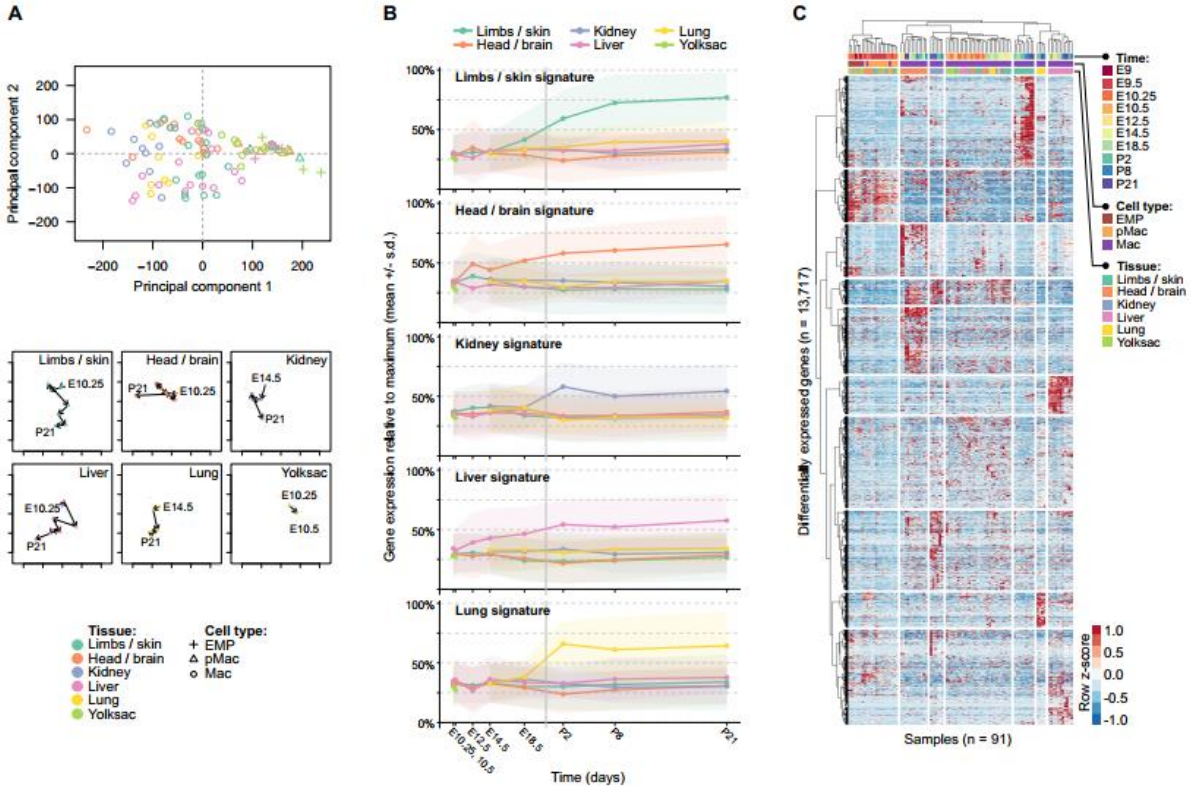
**Mass, Ballesteros et al., Fig. S7**



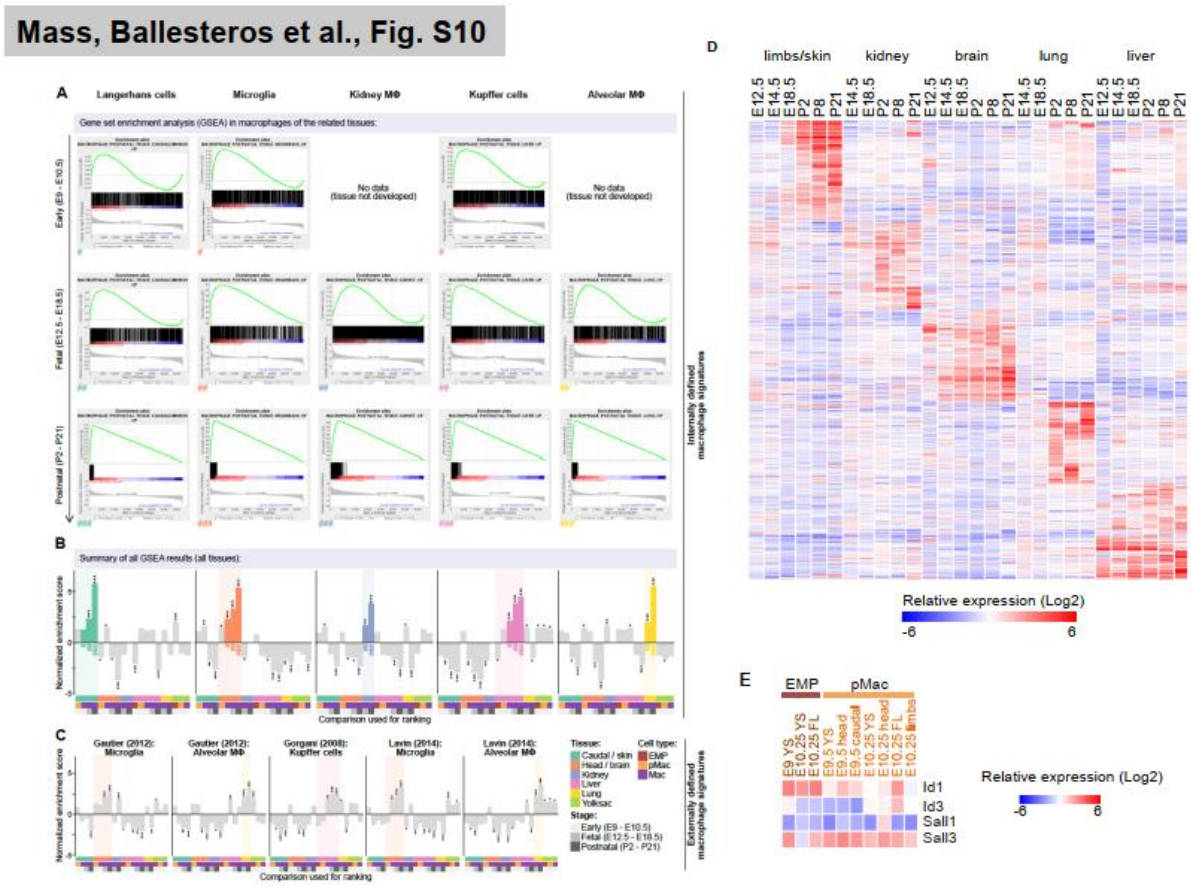

Mass, Ballesteros et al., Fig. S8
A
EMP → pMac
EMP → MΦ
-log10(adjusted p-value)
Atf3 Batf Cebpa Cebpb E2f1 Egr1 Ezh2 Gata3 *Gcgr Gfi1 Hif1a Irf1 Irf8 JunB Maf *Maf Pparg Rel Rela Relb Ring1b Spi1 Stat1 Stat3 Stat5b Stat5 Suz12
* Not expressed
AML pre-leukaemic 0/1
Dendritic cell 28/49
Lymphoid cell 3/150
Macrophage 24/53
Embryonic stem cell 7/167
Erythro/mega lineage 1/36
Myeloid progenitor 1/15
Osteoblast/adipocyte 2/15
Haematopoietic stem and progenitor cell 1/14
Haemogenic endothelium 0/10
Log-odds ratio
B
EMP
pMac
MΦ
E9 YS
E10.25 YS
E10.25 FL
E9.5 YS
E9.5 head
E9.5 caudal
E10.25 YS
E10.25 head
E10.25 FL
E10.25 limbs
E10.5 YS
E10.5 head
E10.25 FL
E10.5 FL
E10.25 limbs
E10.5 limbs
Relative expression (Log2)
E2f1
Suz12
Irf1
Hif1a
Ring1
Rela
Atf3
Batf3
Cebpa
Irf8
Egr1
JunB
Maf
Cebpb
Stat1
Ezh2
Stat6
Relb
Rel
Stat3
Stat5b
Gfi1b
Spi1
Myb
C
50 kb
Pu.1
Cebpα
Cebpβ
Maf
Stat1
Stat6
Irf8
JunB
Rela
Atf3
Emr1
Cx3cr1
Mrc1
Thbs1
Gata1
Myod1

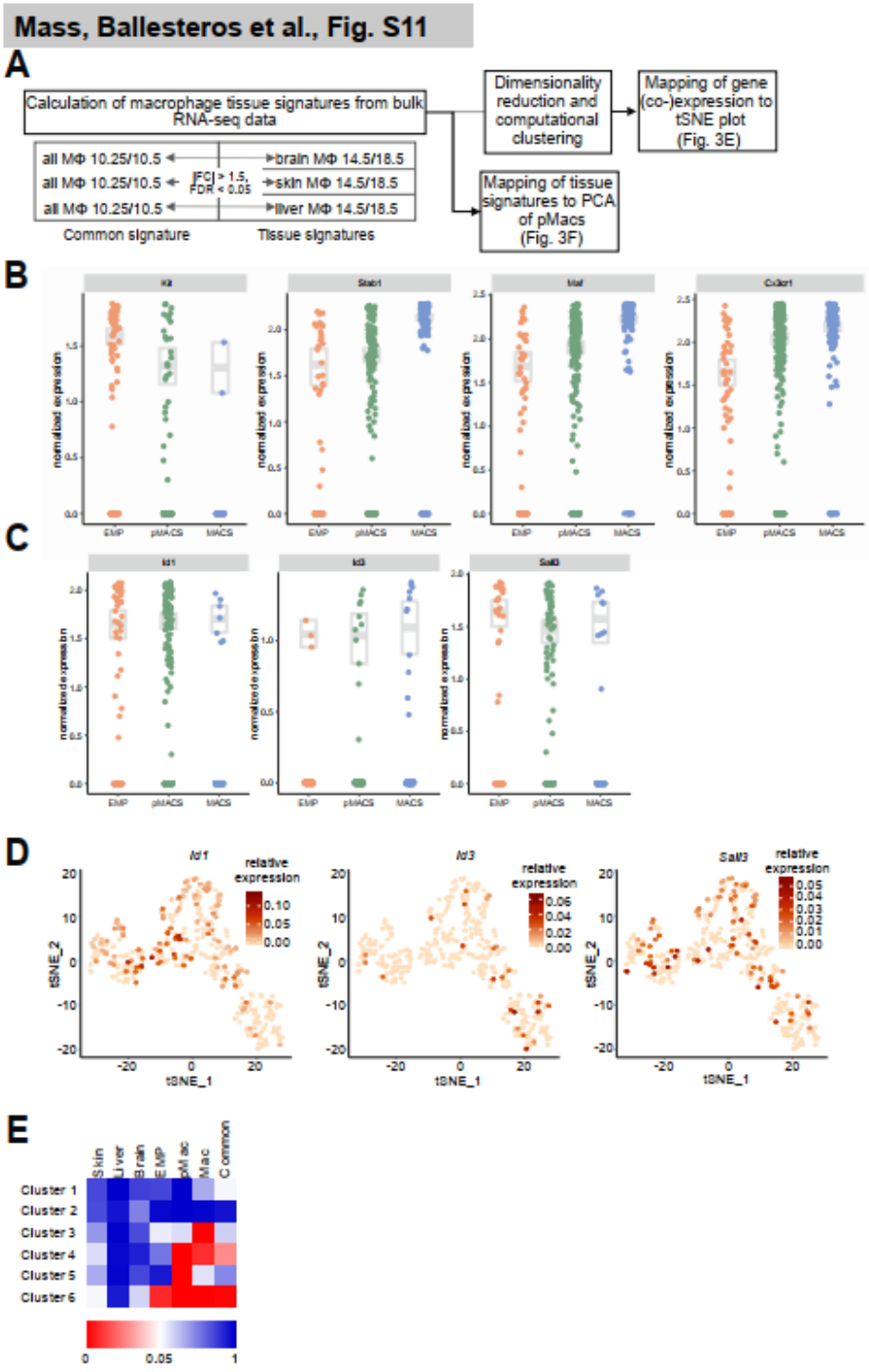
Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238

## Mass, Ballesteros et al., Fig. S9

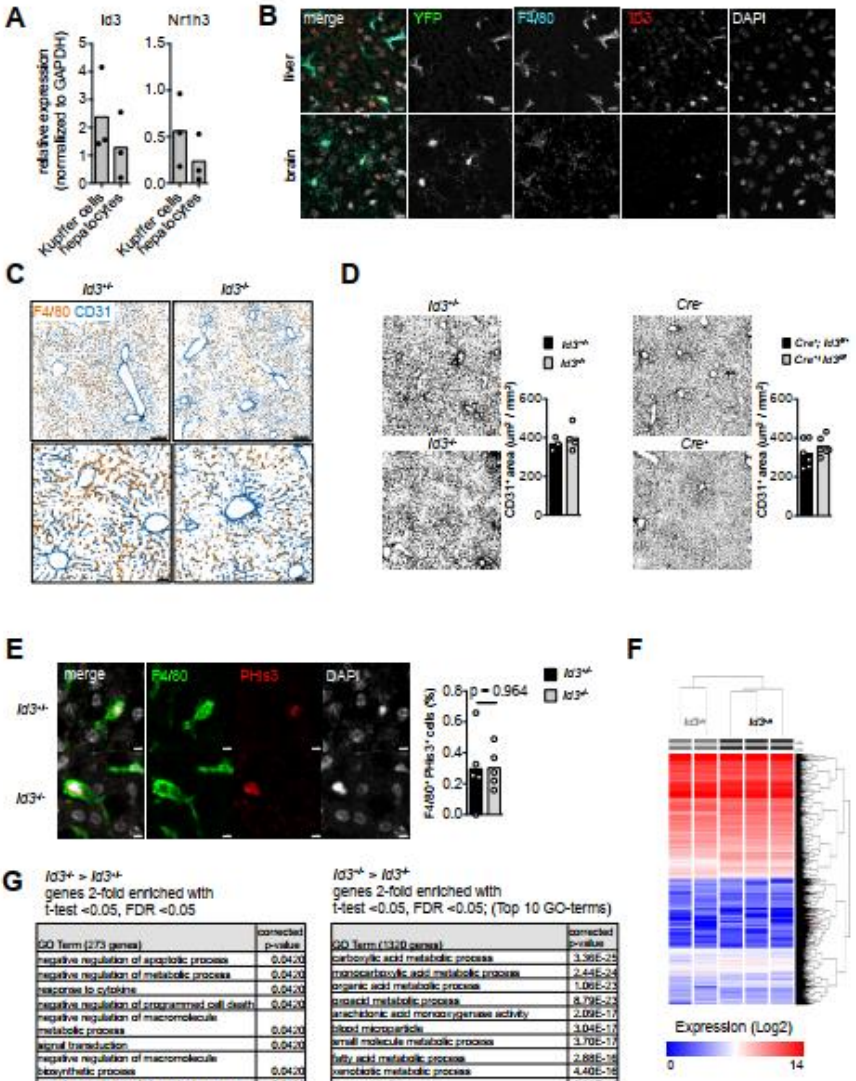


Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238




Mass, Ballesteros et al., Fig. S11
A
Calculation of macrophage tissue signatures from bulk RNA-seq data
Dimensionality reduction and computational clustering
Mapping of gene (co-)expression to tSNE plot (Fig. 3E)
Mapping of tissue signatures to PCA of pMacs (Fig. 3F)
all MΦ 10.25/10.5
all MΦ 10.25/10.5
all MΦ 10.25/10.5
FC| > 1.5, FDR < 0.05
brain MΦ 14.5/18.5
skin MΦ 14.5/18.5
liver MΦ 14.5/18.5
Common signature
Tissue signatures
B
C
D
Id1
Id3
Sall3
relative expression
tSNE_1
tSNE_2
E
Skin
Liver
Brain
EMP
pMac
Mac
Common
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
0
0.05
1

**Mass, Ballesteros et al., Fig. S12**




**G**

*Id3+/+ > Id3+/+*
genes 2-fold enriched with
t-test <0.05, FDR <0.05

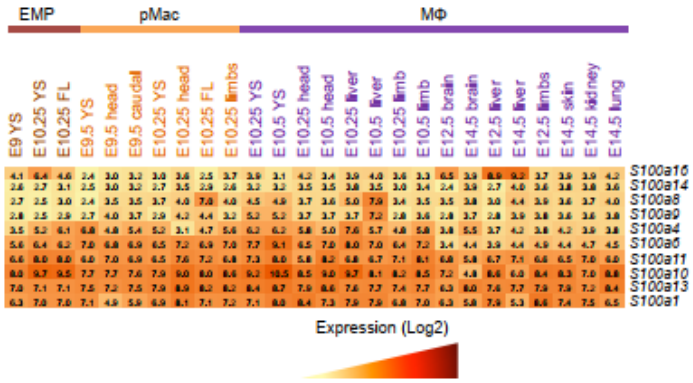
| GO Term (273 genes) | corrected p-value |
|---|---|
| negative regulation of apoptotic process | 0.0420 |
| negative regulation of metabolic process | 0.0420 |
| response to cytokine | 0.0420 |
| negative regulation of programmed cell death | 0.0420 |
| negative regulation of macromolecule metabolic process | 0.0420 |
| signal transduction | 0.0420 |
| negative regulation of macromolecule biosynthetic process | 0.0420 |
| negative regulation of biosynthetic process | 0.0420 |
| negative regulation of protein kinase activity | 0.0495 |

*Id3-/- > Id3+/+*
genes 2-fold enriched with
t-test <0.05, FDR <0.05; (Top 10 GO-terms)

| GO Term (1320 genes) | corrected p-value |
|---|---|
| carboxylic acid metabolic process | 3.39E-25 |
| monocarboxylic acid metabolic process | 2.44E-24 |
| organic acid metabolic process | 1.06E-23 |
| oxoacid metabolic process | 8.79E-23 |
| arachidonic acid monooxygenase activity | 2.09E-17 |
| blood microparticle | 3.04E-17 |
| small molecule metabolic process | 3.70E-17 |
| fatty acid metabolic process | 2.88E-16 |
| xenobiotic metabolic process | 4.40E-16 |
| drug catabolic process | 6.32E-16 |

Accepted Version: Mass et al., Science 2016, Vol. 353, Issue 6304, doi: 10.1126/science.aaf4238

**Mass, Ballesteros et al., Fig. S13**



**Supplemental figure legends**

**Figure S1: Quality control and analysis of bulk RNA-seq. (A)** Number of reliably detected transcripts (covered with a least 25 reads) in each library. The colored circles underneath indicate the batch (sequencing flow cell), time, cell type, and tissue of each sample with the color code as given below panel B. **(B)** Hierarchical clustering with Euclidean distance and complete linkage of all samples based on their RNA-seq gene expression profiles. The colored circles underneath the cluster dendrogram indicate the batch (sequencing flow cell), time, cell type, and tissue of each sample with the color code as given below. **(C)** Scorecard analysis of differentially up-regulated genes (DESeq2 Wald test, adjusted p-value<0.05, BH-correction) in early macrophages (E10.25, E10.5) in comparison to EMPs. The table shows the relative enrichment of differentially upregulated genes in macrophages across cell types and tissues (y-axis) and developmental time points (x-axis, from E9 to P21). See Methods for details of the scorecard. **(D)** Principal component analysis (PCA) plot of EMPs (red, E9-E10.25), pMacs (yellow, E9.5-E10.25) and macrophages (purple, E10.25-E10.5) from the head, caudal, fetal liver (FL) and yolk sac (YS). The shape of each dot indicates the tissue the sample was taken from**.** The first and second principal component explain 18.9% and 11.1% of the entire variation in the data, respectively.

**Figure S2: Quality control and analysis of single-cell RNA-seq. (A)** Workflow of the MARS-seq single cell data analysis. **(B)** Mean-variability plot shows average expression and dispersion for each gene. This analysis was used to determine highly variable genes (labeled by gene symbol). These 138 highly variable genes were used to perform a dimensionality reduction of the single-cell data by a principal component analysis. **(C)** The highest gene loadings in the first and second principal component from the PCA of 408 high quality cells, colored by batch association, showed even distribution of cells among the PCA plot based on the 138 most highly variable genes. **(D)** Heatmap of 138 highly variable genes among single-cell clusters as defined by DBScan clustering. **(E)** Optimal cluster number was identified by calculation of diverse indices for determining the best clustering scheme using the NbClust R package. **(F)** PCA plot of 408 single cells colored by cluster association. Clusters were defined by PCA + DBScan clustering. **(G)** Kinetic diagram shows the pseudotemporal ordering of single cells as determined by Monocle 2. Dots indicate individual cells and are colored according to the cluster association as in (F). Black line indicates the progression of single cells over developmental pseudotime.

**Figure S3 Expression of surface markers on EMP-derived cells during development. (A)** Flow cytometry analysis of E10.25 *Csf1r*$^{MeriCreMer}$; *Rosa26*$^{LSL\text{-}YFP}$ (OH-TAM at E8.5) tissues showing expression of Il4ra, Il13ra1, Tnfr2, Ifngr, CD16.2, CD64, Tim4, and CD206 on YFP$^{+}$ Kit$^{+}$ progenitors (gray), pMacs (blue) and macrophages (orange). Histograms represent the fluorescence intensity for each antibody in each cell subset. Data are representative of n=4 independent experiments with 4-6 embryos per marker. **(B,C)** Flow cytometry analysis of *Csf1r*$^{MeriCreMer}$; *Rosa26*$^{LSL\text{-}YFP}$ (OH-TAM at E8.5) liver and brain F4/80$^{+}$ YFP$^{+}$ cells from E14.5 embryos and adult mice (>8 week old) showing expression of Il4ra, Il13ra1, Tnfr2, Ifngr, Dectin-1, CD64, Tim4, and CD206 (purple). Gray histograms show the fluorescence intensity of the FMO controls.

**Figure S4 Expression of the 'core' macrophage program on EMP-derived cells. (A)** Immunostaining on cryosections from E10.25 *Csf1r*$^{MeriCreMer}$; *Rosa26*$^{LSL\text{-}YFP}$ embryos, pulse-labeled with OH-TAM at E8.5 with antibodies against YFP (green), Iba1 (red/cyan), and CD206 (red), Ifngr (red), Tnfr2 (red), Dectin-1 (red), Trem2 (red), CD16/32 (red), Granulin (Grn, red), or F4/80 (cyan). Scale bars represent 10 µm. Data are representative of n=3 embryos for each marker.

**(B)** Whole mount immunostaining of E9.5 *Csf1r*$^{iCre}$*; Rosa26*$^{LSL\text{-}YFP}$ embryo labeled with antibodies against YFP (green), Iba1 (red), F4/80 (cyan) and DAPI (white). Scale bars represent 10 µm. Data are representative of n=3 embryos. **(C)** Immunostaining on cryosections from E10.25 *Csf1r*$^{MeriCreMer}$*; Rosa26*$^{LSL\text{-}YFP}$ embryo liver, pulse-labeled with OH-TAM at E8.5 with antibodies against YFP (green), Dectin-1 (red) and Iba1 (cyan) (upper panel) or YFP (green), Kit (red) and F4/80 (cyan) (lower panel) Scale bars represent 15 µm. **(D, E)** Immunostaining on cryosection from E14.5 (D) and E18.5 (E) *Csf1r*$^{MeriCreMer}$*; Rosa26*$^{LSL\text{-}YFP}$ mouse embryos stained with antibodies against YFP (green), Iba1 (red), and F4/80 (cyan). **(F)** Immunostaining on cryosection from E14.5 *Csf1r*$^{MeriCreMer}$*; Rosa26*$^{LSL\text{-}YFP}$ mouse embryo stained with antibodies against YFP (green), Granulin (Grn, red) and F4/80 (cyan). Scale bars represent 10 µm.

**Figure S5 Analysis of *Tnfrsf11a*$^{Cre}$; *Rosa26*$^{LSL\text{-}YFP}$ mice. (A)** Gating strategy for *Tnfrsf11a*$^{Cre}$; *Rosa26*$^{LSL\text{-}YFP}$ embryos in E10.25 YS pMacs (Kit$^-$ CD45$^+$ F4/80$^-$ CD11b$^{low}$ Gr1$^-$ Ter119$^-$; green) and macrophages (CD45$^+$F4/80$^+$CD11b$^{lo}$; blue) (upper panel), and in E14.5 fetal liver LT-HSCs (Lin$^-$Kit$^+$Sca1$^+$CD150$^+$CD48$^-$; orange), ST-HSCs (Lin$^-$ Kit$^+$Sca1$^+$CD150$^-$CD48$^-$; blue) and MPPs (Lin$^-$Kit$^+$Sca1$^+$CD150$^-$CD48$^+$; purple) (lower panel). Histograms represent YFP expression in *Tnfrsf11a*$^{Cre-}$; *Rosa26*$^{LSL\text{-}YFP}$ (grey) and *Tnfrsf11a*$^{Cre+}$; *Rosa26*$^{LSL\text{-}YFP}$ (color for cell type indicated in gating strategy). **(B)** Immunostaining on cryosection from E14.5 *Tnfrsf11a*$^{Cre}$*; Rosa26*$^{LSL\text{-}YFP}$ embryo, with antibodies against YFP (green), Iba1 (red) and F4/80 (cyan). Scale bars represent 10 µm.

**Figure S6 Analysis of *Cx3cr1*-deficient mice. (A)** Immunostaining on cryosection from E10.25 *Cx3cr1*$^{gfp/+}$ embryo with antibodies against GFP (green), Iba1 (red) and F4/80 (cyan). **(B, C, D)** Flow cytometry analysis in *Cx3cr1*$^{+/-}$ and *Cx3cr1*$^{-/-}$ of F4/80$^+$ macrophages from liver, brain, and limbs at E12.5 (B), liver, brain, limbs, and lung from at E14.5 (C) and liver, brain, limbs, lung, and epidermis from 12 week-old mice (D). Circles represent individual mice. Data are representative of n=7 independent experiments.

**Figure S7 Gene set enrichment analysis (GSEA) of cell type-specific expression patterns. (A)** GSEA plots illustrating the relative enrichment of genes that we found statistically significantly upregulated (see Methods) in the comparison of macrophage vs. pMac (top row), macrophage vs. EMP (middle row), or pMac vs. EMP (bottom row). The black bars in the middle of each plot indicate the transcripts which are in each respective lists. The order in which transcripts were input into the analysis was defined by the relative expression change (logarithmic fold change) in macrophage vs. pMac (left), macrophage vs. EMP (center), or pMac vs. EMP (right), respectively. **(B)** The GSEA results from panel A are summarized here by the normalized enrichment score (NES) of each analysis. The colored squares indicate the two cell types compared with the color code as defined at the bottom right of this figure. Asterisks adjacent to bars indicate significance (FDR-corrected p-value); *: q≤0.05, **: q≤0.01, ***: q≤0.001. **(C)** Comparison of transcripts upregulated in macrophage vs. pMac, macrophage vs. EMP, and pMac vs. EMP, and the core macrophage signature extracted from Gautier *et al.* (2012). After translating gene denominators to Ensembl transcript IDs, Gautier's list matches 179 transcripts, only 27 of which occur in one of the lists of upregulated genes in this study. **(D)** GSEA results of Gautier's core macrophage signature genes in the comparisons of macrophage vs. pMac, macrophage vs. EMP, or pMac vs. EMP from this study (same gene rankings used in panels A and B). In all three comparisons, the core macrophage signature is enriched significantly (q≤0.01).

**Figure S8 LOLA enrichment analysis and enhancer regions. (A)** Results of a LOLA (*1*) enrichment analysis of promoter-adjacentl regions (TSS +/- 20kb) of upregulated genes during transition from EMPs to pMacs (upper panel) or EMPs to early

macrophages (E10.25-E10.5; lower panel). Each dot represents one single ChIP-seq experiment with the size relative to the quantity of enrichment (log odds ratio) and colors indicating the cell type used in the respective experiment. The key below the plots denominates the color coding. The numbers ("x/y") given behind the cell type specify the number of enriched (x) out of all available datasets (y) from the respective cell type. Shown are all transcription factors of the genomic binding locations, which are significantly enriched (adjusted p-value <0.001, Benjamini Yekutieli correction) in at least one dataset in either comparison. All transcription factors shown are expressed in at least one sample examined, with the exception of *Gcgr* and *Maff* (greyed out and marked with an asterisk). **(B)** Heat map representation of the expression of the transcription factors shown in (A) in EMPs, pMacs and early macrophages (E10.25-E10.5) **(C)** Genome browser tracks showing ChIP-seq signals for the indicated transcription factors at the *Emr1*, *Cx3cr1*, *Mrc1*, *Thbs1, Gata1* and *Myod1* loci. The tracks display ChIP-seq data from macrophage populations, except for Maf (T cells), and Irf8 (dendritic cells).

**Figure S9 Coordinated changes of gene expression identified during the specification of tissue-resident macrophages. (A)** Top panel shows an unsupervised, low-dimensional projection via principal component analysis of the RNA-seq data in this study. The color of each dot indicates the tissue the sample was taken from. Bottom panels show only the macrophage datasets from one tissue at a time (gray dots). The mean of the coordinates at each time point was calculated and indicated as a colored dot (with the same color code as above). Consecutive time points were then connected with arrows to visualize the differentiation trajectory described by the transcriptional profiles of these samples. **(B)** Line plots illustrating average expression of tissue-specific gene signatures in macrophages over time. Each panel corresponds to one list of differentially upregulated tissue-specific genes (Table S3) and each line to the average of all macrophage samples of one tissue at the given time. Shaded areas indicate +/- standard deviation. **(C)** Heatmap showing the expression levels of all RNA-seq datasets across all transcript differentially upregulated in any tissue in either stage (early embryonic, fetal, postnatal). Expression values have been scaled as z-scores per row, resulting in a color scheme in which red values represent highly expressed and blue values represent lowly expressed transcripts. Hierarchical Ward clustering with Euclidean distance was used to arrange the samples and genes and the resulting dendrogram was cut at equal height into seven (column-wise) or ten (row-wise) clusters to highlight the strongest clusters of genes and samples more clearly. The color bars on top of the heatmap indicate the time, cell type, and tissue of each sample.

**Figure S10 Gene set enrichment analysis (GSEA) of tissue-specific expression patterns in adult macrophages. (A)** GSEA plots illustrating the relative enrichment of genes that were found to be statistically significantly upregulated (see Methods) in the comparison to Langerhans cells, microglia, kidney macrophages, Kupffer cells, and alveolar macrophages (from left to right) in comparison to the respective other adult macrophage populations. The black bars in the middle of each plot indicate the transcripts, which are in each respective lists. The order in which transcripts were input into the analysis was defined by the relative expression change (logarithmic fold change) in macrophages in the corresponding tissue (e.g. head/brain for microglia) compared to other macrophages at the same developmental stage (from top to bottom: early, fetal, or postnatal). The colored hashes (#) beneath each plot link to the bars in panel B. **(B)** The GSEA results from (A) are summarized here by the normalized enrichment score (NES) of each analysis. The colored squares underneath the bar plots indicate the tissue, cell type, and developmental stage of the sample for which enrichment is shown, following the color code as used in panel C. The bars highlighted in color correspond to the plots shown in panel A. Additionally, these and other bars

from samples from a tissue related to the signature at hand are indicated by shading. Asterisks adjacent to bars indicate significance (FDR-corrected p-value); *: q≤0.05, **: q≤0.01, ***: q≤0.001. **(C)** GSEA results against externally defined gene signatures extracted from Gautier *et al.* (2012), Gorgani *et al.* (2008), and Lavin *et al.* (2014) (*1-3*). The same color code, shading, and annotations are used as in panel B. **(D)** Hierarchical clustering of differentially up-regulated genes (2-fold change, adj. P-value < 0.05, BH-correction) in post-natal (P2-P21) brain, liver, kidney, epidermis or lung macrophages comparing one population vs. the others (see also Table S3). Each sample represents the mean of at least two biological replicates and two technical replicates, except for E14.5 liver macrophages and P8 and P21 lung macrophages, which consist of one biological replicate and two technical replicates. **(E)** Heatmap representation of the expression of *Id1, Id3, Sall1,* and *Sall3* in EMPs and pMacs from bulk RNA-seq data.

**Figure S11 scRNA-seq analysis of specification of tissue-resident macrophages**. **(A)** Workflow for overlaying transcription factor co-expression and tissue macrophage-specific signatures onto the tSNE plots. **(B)** Normalized expression of *Kit, Stab1, Maf*, and *Cx3cr1* within EMPs, pMacs, and macrophages. **(C)** Normalized expression of *Id1, Id3*, and *Sall3* within EMPs, pMacs, and macrophages. **(D)** tSNE plots showing expression of *Id1, Id3,* and *Sall3*. **(E)** Heatmap depicting enrichment of tissue macrophage-specific signatures as in Fig. 3F or differentiation signatures as in Fig. 1E in pMacs on gene clusters defined by multimodal gene expression with subsequent hierarchical clustering of scRNA-seq data.

**Figure S12 Role of Id3 in Kupffer cell development (A)** Relative expression of *Id3* and *Nr3h1* transcript by qRT-PCR (normalized to GAPDH) in sorted Kupffer cells and hepatocytes of C57BL/6 mice n=3. **(B)** Immunostaining with antibodies against YFP (green), ID3 (red) and F4/80 (cyan) on cryosection from adult (4 week-old) liver (upper panel) or brain (lower panel) from a pulse labeled $Csf1r^{MeriCreMer}$; $Rosa26^{LSL\text{-}YFP}$ mouse (OH-TAM at E8.5). Nuclei were counterstained with DAPI (white). Scale bar represents 10 µm. **(C)** Immunostaining with antibodies against CD31 and F4/80 on liver cryosections from 4 week-old $Id3^{-/-}$ and $Id3^{+/-}$ mice. The figure displays isovolume-rendered images. Scale bars represent 150 µm for the overview of the adult tissue and 50 µm for insets. Data are representative of 5 adult mice. **(D)** $CD31^+$ area quantification on liver sections from $Id3^{+/-}$ (n=3) and $Id3^{-/-}$ (n=4) 4 week-old mice (left panels) or from $Tnfrsf11a^{Cre+}$; $Id3^{+/f}$ (n=6) and $Tnfrsf11a^{Cre+}$; $Id3^{f/f}$ (n=6) 2 week-old mice (right panels). **(E)** Immunostaining with antibodies against F4/80 (green) phospho-histone 3 (PHis3, red) and DAPI (gray) on liver cryosections from 4 week-old $Id3^{-/-}$ and $Id3^{+/-}$ mice. Graph shows the $PHis3^+$ cells percentage of total $F4/80^+$ cells. Scale bar represents 10 µm. Data are representative of 5 mice per genotype. Each dot represents the mean of $PHis3^+$ cells (in %) of total $F4/80^+$ found in 5 sampling areas (830 $\mu m^2$) for each individual liver. **(F)** Unsupervised hierarchical clustering on whole transcriptome from $Id3^{+/-}$ and $Id3^{-/-}$ Kupffer cells (Distance metric: Euclidian, linkage rule: Ward's, number of genes: 18882) **(G)** Significantly enriched Gene Ontology (GO) terms identified for 2-fold up-regulated (left table) or down-regulated (right table) genes in $Id3^{-/-}$ vs. $Id3^{+/-}$ Kupffer cells enriched by a t-test (P<0.05; FDR<0.05). GO analysis was performed using GeneSpring. GO terms are depicted and raked by corrected P-value (FDR false discovery rate corrected for multiple testing).

**Figure S13 Heat map representation of the normalized counts (log2) reads of S100a mRNA found in our dataset**.

**Supplemental tables**

Supplemental Table S1: Genes differentially regulated in pMacs and macrophages (E10.25-E10.5) in bulk RNA-seq experiments

Supplemental Table S2: Gene signatures derived from bulk RNA-seq experiments used for single cell RNA-seq analysis

Supplemental Table S3: Genes differentially regulated in postnatal macrophages (P2-P21) in bulk RNA-seq experiments

Supplemental Table S4: antibodies used for Flow cytometry

Supplemental Table S5: complete bulk RNA-seq data http://www.biomedical-sequencing.at/bocklab/fhalbrit/macro/data/EC7D8Y7EL5QDMYIOL7GW2SECRWZUN1CWPQCUSJ47FFRO1WHD.zip

Supplemental Table S6: RNA-seq analysis on Kupffer cells from Id3KO and control littermates

**Supplemental Materials and Methods**

**Animals**. *Csf1r*$^{MeriCreMer}$, *Csf1r*$^{iCre}$, *Rosa26*$^{YFP\text{-}LSL}$ reporter mice (*4*) and *Cx*$_3$*cr1*$^{gfp/+}$ *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ (*5*) have been previously described. *Tnfrsf11a*$^{Cre}$ mice were kindly provided by Dr Y. Kobayashi (*6*), *Id3*$^{-/-}$; *Id1*$^{fl/fl}$ and *Id3*$^{f/f}$ mice were kindly provided by R. Benezra (*7-9*). Embryonic development was estimated as previously (*4*) considering the day of vaginal plug formation as 0.5 days post-coitum (dpc), and staged by developmental criteria. E9: 20-25sp, E9.5: 26-29sp, E10.25: 30-35sp, E10.5: 36-44sp. All animal procedures were performed in adherence to our project licence issued by the United Kingdom Home Office under the Animals (Scientific Procedures) Act 1986 and by the Institutional Review Board (IACUC 15-04-006) from MSKCC. All mice were maintained under SPF conditions.

**Genotyping.** PCR genotyping of *Csf1r*$^{iCre}$ (*10*) *Csf1r*$^{MeriCreMer}$ (*5*), *Cx*$_3$*cr1*$^{gfp/+}$ *Id3*$^{-/-}$; *Id1*$^{fl/fl}$ (*7*) and *Tnfrsf11a*$^{Cre}$ (*6*) mice was performed according to protocols described previously. *Cx*$_3$*cr1*$^{gfp/+}$ genotyping was performed with following primers: Cx$_3$cr1 F 5'-CCC AGA CAC TCG TTG TCC TT-3', Cx$_3$cr1 R 5'-GTC TTC ACG TTC GGT CTG GT and Cx$_3$cr1 R mut 5'CTC CCC CTG AAC CTG AAA C-3'

**Pulse labelling of *Csf1r*$^{MeriCreMer}$; *Rosa26*$^{LSL\text{-}YFP}$**. For genetic cell labelling to perform immunofluorescence we crossed tamoxifen-inducible *Csf1r*$^{MeriCreMer}$ transgenic mouse strains with *Rosa26*$^{LSL\text{-}YFP/LSL\text{-}YFP}$ reporter mice. In *Csf1r*$^{MeriCreMer}$; *Rosa26*$^{LSL\text{-}YFP}$ embryos recombination was induced by single injection at E8.5 of 75 mg per g (body weight) of 4-hydroxytamoxifen (OH-TAM, Sigma) into pregnant females. The OH-TAM was supplemented with 37.5 mg per g (body weight) progesterone (Sigma) to counteract the mixed oestrogen agonist effects of tamoxifen, which can result in fetal abortions.

**Processing of tissues for flow cytometry and cell sorting**. Pregnant females were killed by cervical dislocation or by exposure to $CO_2$. Embryos ranging from embryonic day (E) 9 to E18.5 were removed from the uterus, washed in 4ºC phosphate-buffered saline (PBS, Invitrogen) and dissected under a Leica M80 microscope. Yolk sacs (YS) were harvested from embryos between E9 and E10.5. To obtain single-cell suspensions for FACS sorting, tissues were included in cold PBS and mechanically disrupted under a 100um filter. Postnatal tissues were collected following the same

procedure. For collection of Langerhans cells, epidermal sheets (from E18.5 to P21) were separated from the dermis after incubation for 45 min at RT in 4.8mg/ml of dispase (Invitrogen), 3% fetal calf serum (FCS, Invitrogen) and 1uM of flavopiridol (Sigma). The epidermis was further digested for 30min at RT in PBS containing 2 mg/ml of collagenase D (Roche), 200U/ml DNase I (Sigma), 4.8 mg/ml of dispase (Invitrogen), 3% FCS (Invitrogen) and 1uM of flavopiridol (Sigma) followed by mechanical disruption under a 100um filter to obtain a single cell suspension.
For blood phenotyping, mice were anaesthetized and blood was collected by cardiac puncture. For flow cytometry experiments, organs were incubated in PBS containing 1mg/ml collagenase D (Roche), 100U/ml DNase I (Sigma), 2.4mg/ml of dispase (Invitrogen) and 3% FCS (Invitrogen) at 37ºC for 30 min prior to mechanical disruption. Epidermal sheet were obtained as previously described (*4*). For embryonic tissue incubation time at 37ºC was reduced to 20 min. Cell suspensions were centrifuged at 320g for 7 min, resuspended in FACS buffer (PBS, 0.5% BSA and 2 mm EDTA) containing purified anti-CD16/32 (FccRIII/II) (1:100 dilution) and incubated for 15min at 4ºC. Samples were immunostained with antibodies mixes for 30 min at 4ºC. The full list of antibodies used can be found in Table S4.

**Gating strategy and cell sorting for bulk sequencing.** Cell sorting was performed using an Aria II BD cell sorter. Single live cells were gated on the basis of DAPI exclusion and identified using side (SSC-A) and forward scatter (FSC-A) gating followed by doublet exclusion using forward scatter width (FSC-W) against FSC-A. EMPs were identified after gating on $Kit^{+}CD45^{lo}$ cells based on AA4.1 expression. pMacs were identified after gating on $Kit^{-}CD45^{+}$ based on CD11b expression and no expression of F4/80. Additionally, $Gr1^{+}$ or $Ter119^{+}$ cells were excluded from the $F4/80^{-}$ $CD11b^{+}$ gate. Macrophage populations were identified after gating on CD45 based on expression of F4/80 and CD11b. 100 cells for each sample were directly sorted into a 96 well plate (Biorad) in 4ul of $H_2O$ containing 0.2% of triton TXT (Sigma) and 0.8U/ul of RNAse inhibitor (Clontech), and processed as indicated below (*Generation of 'bulk' transcriptomes from candidate EMP, pMac, and macrophage populations from E9 to P21*).

**Gating strategy and cell sorting for single cell sequencing.** Cell sorting was performed as above, but after doublet exclusion all $CD45^{low/+}$ single cells from a E10.25 (30-34sp) embryo proper were sorted into 384-well plates filled with 2 µl lysis buffer (Triton-X 0.2% (Sigma) in molecular biology grade $H_2O$ (Sigma) supplemented with 0.4 U/µl protein-based RNase inhibitor (Takara) and barcoded poly-T primers, and processed as described below (*Generation of single cell transcriptomes*).

**Cytospin**. E10.25 embryos were staged by somite counting. E10.25 yolk sac, head, liver and limbs and E12.5 head, liver and limbs were dissected. Processing and labelling of cells was performed as described above. Cells were collected into FCS. Cytospin preparations were stained with May-Grunwald-Giemsa method for morphological inspection. Cytospin preparations were performed using a Cytospin 3 (Thermo Shandon) by centrifuging sorted cells at 800 rpm for 10 min (low acceleration). Slides were air-dried for at least 30 min, and fixed for 5 min in methanol. Methanol-fixed cytospin preparations were manually stained in 50 % May-Grunwald solution for 5 min, 14 % Giemsa for 15 min, washed with Sorensons buffered distilled water (pH 6.8) for 5 min and rinsed with Sorensons buffered distilled water (pH 6.8). After air-drying, slides were mounted with Entellan New (Merck) and representative pictures were taken using an Axio Lab.A1 microscope (Zeiss) under a N-Achroplan 100x/01.25 objective.

**Immunofluorescence, imaging, analysis and illustrations.** For whole mount immunofluorescence E9.5 embryos were processed as described elsewhere (*11*). For

cryosections E10.25 Embryos were fixed for 4 hours in 4% formaldehyde (Sigma) under agitation and >E10.25 embryos were fixed overnight. After fixation, embryos were incubated overnight in 30% sucrose and embedded in OCT compound (Sakura Finetek). Cryoblocks were cut at a thickness of 10 -12µm and then blocked with PBS containing 5% normal goat serum (Invitrogen); 1 % BSA (w/v); 0.3% Triton X-100 for 1 hour at room temperature. Samples were incubated overnight at 4˚C with rat anti-mouse F4/80 (1:200, Biorad), rabbit anti-mouse Iba1 (1:200; Wako), chicken anti-GFP for YFP detection (1:500, Invitrogen), rat anti-mouse Dectin-1 (1:200, Biorad), anti-Granulin (1:200, abcam), rat anti-mouse CD206 (1:200, Biorad), armenian hamster anti-mouse PECAM-1 (1:300, Thermo Scientific), rabbit anti-mouse/human ID3 (1:500, Biocheck), rabbit anti-mouse ID1 (1:500, Biocheck), rat anti-mouse CD16/CD32 (BD Biosciences), goat anti-mouse Trem2 (1:200, Abcam), armenian hamster anti-mouse CD119 (1:200, Clone 2E2, eBioscience), armenian hamster anti-mouse CD120b (1:200, clone 55R-286, Biolegend). Secondary antibodies used were anti-rabbit Cy3 (1:500; Invitrogen), anti-chicken Alexa Fluor 488 (1:500; Invitrogen), anti-rat Alexa Fluor 555 (1:500; Invitrogen), anti-rat Alexa Fluor 488 (1:500; Invitrogen), anti-rat Alexa Fluor 647 (1:500; Invitrogen), anti-goat Alexa Fluor 568 and anti-armenian hamster Dylight 649 (Jackson ImmunoResearch Laboratories). Samples were then mounted with Fluoromount mounting medium with DAPI (eBiosciences) and visualized using a LSM880 Zeiss microscope with 20x/0.5 (dry), performing a tile scan and Z-stack on whole embryos or tissues. Image analysis and cell quantification was performed using Imaris (Bitplane) software. To determine the $CD31^+$ area in *Id3*$^{+/-}$ and *Id3*$^{-/-}$ liver sections, maximum intensity Z-projections pictures were converted into binary images and the $CD31^+$ area was measured using Image J (NIH, Bethesda, MD, USA) (*12*). Results were normalized per $mm^2$ of tissue. Illustrations were created by adapting templates from Servier Medical Arts (http://www.servier.com/Powerpoint-image-bank, licensed under a Creative Commons Attribution 3.0 Unported License).

**qRT-PCR on Kupffer cells and hepatocytes.** Cells were sorted as described above. Hepatocytes were enriched by centrifugation of the whole liver cell suspension at 50 g for 3 min (Sorvall Legend XTR centrifuge). Supernatant was taken for further staining of macrophages ($CD45^+$, $CD11b^{lo}$, $F4/80^+$). Hepatocytes were sorted using the FSC-A and SSC-A gate with subsequent exclusion of doublets and $CD45^+$ cells. Cells were sorted into RNA lysis buffer and RNA extraction was performed as per manufacturers protocol (Macherey-Nagel). cDNA was synthesized using the QuantiTect Reverse Transcription Kit (Qiagen) as per manufacturers protocol. qRT-PCR was performed on a QuantStudio 6 Flex using TaqMan Fast Advanced Master Mix (applied biosystems) and TaqMan probes for Id3 (Mm00492575_m1), Nr1h3 (Mm00443451_m1), and GAPDH (Mm99999915_g1) (Life Technology).

**Statistics analysis (apart from RNA-seq experiments)**. Unpaired students t-tests were used to assess statistical differences between measurements where populations were distributed normally. Where populations were not normally distributed, Mann-Whitney tests were used to assess statistical differences. Normal distribution was assessed using D'Agostino and Pearson omnibus normality test. All statistics were performed on GraphPad Prism 6 (GraphPad Software).

**RNA sequencing of *Id3*$^{+/-}$ and *Id3*$^{-/-}$ Kupffer cells.** Kupffer cells were sorted into Trizol as described above. RNA from sorted cells was extracted using RNeasy mini kit (Qiagen) according to instructions provided by the manufacturer. After ribogreen quantification and quality control of Agilent BioAnalyzer, 400pg of total RNA underwent amplification (12 cycles) using the SMART-seq V4 (Clonetech) ultra low input RNA kit for sequencing. 10 ng of amplified cDNA was used to prepare Illumina hiseq libraries with the Kapa DNA library preparation chemistry (Kapa Biosystems) using 8 cycles of

PCR. Samples were barcoded and run on a Hiseq 2500 1T in a 50bp/50bp Paired end run, using the TruSeq SBS Kit v3 (Illumina). An average of 54 million paired reads were generated per sample and the percent of mRNA bases was closed to 77% on average.The output data (FASTQ files) were mapped to the target genome using the rnaStar aligner (*13*) that maps reads genomically and resolves reads across splice junctions. 2 pass mapping method was used, outlined in Engstrom et al.(*14*) in which the reads are mapped twice. The first mapping pass uses a list of known annotated junctions from Ensemble. Novel junctions found in the first pass are then added to the known junctions and a second mapping pass is done (on the second pass the RemoveNoncanoncial flag is used). After mapping we post process the output SAM files using the PICARD tools to: add read groups, AddOrReplaceReadGroups which in additional sorts the file and coverts it to the compressed BAM format. We then compute the expression count matrix from the mapped reads using HTSeq and one of several possible gene model databases. The raw count matrix generated by HTSeq are then processed using the R/Bioconductor package DESeq, which was used to both normalize the full dataset and analyze differential expression between sample groups. Gene Ontology analysis was performed using the GO analysis function in GeneSpring GX 13.0 (Agilent), with the p-value calculated using a hypergeometric test with Benjamini–Yekutieli correction. For that, genes with a fold change difference of ± 2 between *Id3*$^{+/-}$ and *Id3*$^{-/-}$ Kupffer cells were selected. Significantly regulated genes (t-test $p < 0.05$; FDR <0.05) from this selection were grouped into gene ontology (GO) terms.

**Generation of 'bulk' transcriptomes from candidate EMP, pMac, and macrophage populations from E9 to P21**

*RNA Isolation and Library Construction.* cDNA synthesis and enrichment was performed following the Smart-seq2 protocol as described (*15, 16*). ERCC spike-in RNA (Ambion) was added to the lysis buffer in a final dilution of 1:1,000,000. After the cDNA was synthesized and amplified from single cells using 18 cycles, quantitative PCR was performed with GoTaq-PCR master mix (Promega) on a C1000 Touch Thermal Cycler qPCR instrument (Bio-Rad) to test for house keeping gene expression. Library preparation was conducted on 1ng of cDNA using the Nextera XT library preparation kit (Illumina) as described (*16*). Sequencing was performed by the Biomedical Sequencing Facility at CeMM using a 50bp single-read setup on the Illumina HiSeq 2500 platform.

*RNA-seq Analysis.* We first trimmed off sequencing adapter from the reads generated, and then aligned the reads using Bowtie v 1.1.1, (*17*) parameters: -q -p 6 -a -m 100 --minins 0 --maxins 5000 --fr --sam --chunkmbs 200) to the cDNA reference transcriptome (mm10 cDNA sequences from Ensembl). For genome browser track visualization, we generated a second alignment with Tophat2 (*18*) (v 2.0.13, parameters: --b2-L 15 --library-type fr-unstranded --mate-inner-dist 150 --max-multihits 100 --no-coverage-search --GTF) against the reference genome (mm10). Next, we removed duplicate reads before quantifying transcript levels with BitSeq (*19*) (v 1.12.0). The raw transcript counts were loaded into R and processed further. At this stage, we removed samples with a substandard quality, that is, all samples that had less than 2 million reads, less than 33% of reads aligned, or less than 5,000 transcripts detected (with >= 25 reads). 20 out of 178 datasets failed these criteria (11.2%). Of the remaining datasets, we merged technical replicates creating 93 unique biological samples. We took forward only transcripts that were detected reliably (>=50 reads) in at least 4 samples (n = 37,521). For statistical analysis, we used raw read counts as input for DESeq2 (*20*), factoring in the flowcell identifier as a covariate to reduce the effect of technical variation. To identify genes of particular interest to the development of tissue-resident macrophages over time and in different tissues, we performed two

types of comparisons: (a) Cell type-specific: Pairwise comparisons (Wald test) between EMPs, pre-macrophages, and macrophages (up to E10.5) independent of their tissue of origin (treating time and tissue as covariates). (b) Tissue-specific: Pairwise comparisons (Wald test) between macrophages from all tissues stratified and stage (post-natal). We considered genes with an FDR-corrected p-value <= 0.05 as differentially expressed. For visualization and illustration purposes (PCA, supplementary tables, heatmaps), we used values adjusted by variance stabilizing transformation from DESeq2 in which batch-effects had been corrected for with ComBat (*21*).

*Evaluation of lists of differentially expressed genes from the RNAseq data***.** We bioinformatically investigated the differentially expressed transcripts (Table S5) from our statistical comparisons in several ways: To identify transcription factors specifically regulating each set of transcripts, we use LOLA(*22*) (v 0.99.4) together with its core database of ChIP-seq binding peaks from CODEX (*23*) to identify enrichment of experimentally-derived transcription factor binding locations in a window around the promoter regions (TSS +/- 20kb) of differentially expressed transcripts. We corrected for multiple testing using the Benjamini & Yekutieli method. To visualize and summarize the expression patterns of many genes (lists of differentially expressed genes) in many different conditions (different tissues at different time points) and across replicates, we sought to use an adaptation of lineage scorecards(*24*). Briefly, we considered each list of differentially upregulated genes as a set of marker genes and determined the relative enrichment of these marker sets in each individual condition (tissue by cell type by time) in comparison to all other datasets using a modified version of parametric gene set enrichment analysis in R (package: *PGSEA*). We also used gene set enrichment analysis (GSEA) to test for the relative overrepresentation of gene signatures in sorted gene lists. To this end, we extracted lists of genes sorted by logarithmic fold change between the mean expression levels in any one tissue Mac, pMac, or EMP sample stratified by stage (early = E9 - E10.5, fetal = E12.5 - E18.5, postnatal = P2 - P21) compared to all other samples at the same stage. Additionally, we extracted lists of all genes we found differentially upregulated in any tissue-specific signature, cell type-specific signature, or in lists of genes extracted from the publications of Gorgani *et al.* (2008), Gautier *et al.* (2012), and Lavin *et al.* (2014) (*1-3*). External gene lists were translated to Ensembl Transcript IDs using g:profiler (*25*). Both sets of data were loaded into and analyzed using the GSEA tool (*26*) and the results read and summarized using the *metaGSEA* R library. A subset of genes was investigated manually in the UCSC genome browser from our track hub (http://genome.ucsc.edu/cgi-bin/hgTracks?db=mm10&position=chr2:91092057-91120496&hubUrl=http://www.biomedical-sequencing.at/bocklab/papers/mass2016/tracks/hub.txt).. We also incorporated ChIP-seq binding profiles (bigWig) for factors identified in the LOLA analysis from the CODEX database. Heatmaps were generated using GeneSpring GX 13.0 (Agilent). All other analyses and plotting was performed in R.

**Generation of single cell transcriptomes**

For single-cell RNA sequencing the MARS-Seq approach described by Jaitin et al. (*27*) was applied using the Biomek FXP system (Beckman Coulter). In brief, single cells were sorted into each well of 384-well plates filled with 2 µl lysis buffer (Triton-X 0.2% (Sigma) in molecular biology grade $H_2O$ (Sigma) supplemented with 0.4 U/µl protein-based RNase inhibitor (Takara)) and barcoded poly-T primers (400 nM (IDT); for barcode details see Jaitin et al., 2014). Samples were pre-incubated (3 min at 80°), reverse transcriptase mix (10 mM DTT (Invitrogen), 4 mM dNTPs (NEB), 2.5 U/µl Superscript III RT enzyme (Invitrogen) in 50 mM Tris-HCl (pH 8.3; Sigma), 75 mM KCl (Sigma), 3 mM $MgCl_2$ (Sigma), ERCC RNA Spike-In mix (Life Technologies) at

1:80x$10^7$ dilution per cell) was added to each well and mRNA was reverse transcribed to cDNA (2 min at 42°C, 50 min at 50°C, 5 min at 85°C). Excess primers were digested (Exo I (NEB); 37°C for 30 min then 10 min at 80°C) followed by a 1.2 x SPRI bead (Beckman Coulter) cleanup, samples were pooled and second strands synthesized (second strand synthesis kit (NEB); 2.5 h at 16°C) followed by a 1.4 x SPRI bead (Beckman Coulter) cleanup. Samples were linearly amplified by T7-promoter guided *in vitro* transcription (T7 High Yield RNA polymerase IVT kit (NEB); 37°C for 12h). DNA templates were digested (Turbo DNase I (Ambion); 15 min at 37°C) followed by a 1.2 x SPRI bead (Beckman Coulter) cleanup and the RNA was fragmented ($Zn^{2+}$ RNA fragmentation solution (Ambion); 1.5 min at 70°C) followed by a 2 x SPRI bead (Beckman Coulter) cleanup. Barcoded ssDNA adapters (IDT; for barcode details see Jaitin et al., 2014, (*27*)) were ligated to the fragmented RNA (9.5% DMSO (Sigma), 1 mM ATP, 20% PEG8000 and 1 U/µl T4 RNA ligase I (NEB) in 50 mM Tris HCl pH7.5 (Sigma), 10 mM $MgCl_2$ and 1mM DTT; 22°C for 2 h) and a second RT reaction (Affinity Script RT buffer, 10 mM DTT, 4 mM dNTP, 2.5 U/µl Affinity Script RT enzyme (Agilent); 2 min at 42°C, 45 min at 50°C, 5 min at 85°C) was performed followed by a 1.5 x SPRI bead (Beckman Coulter) cleanup. Final libraries were generated by subsequent nested PCR reaction (0.5 µM of each Illumina primer (IDT; for primers details see Jaitin et al., 2014) and KAPA HiFi HotStart ready mix (Kapa Biosystems) for 15 cycles according to manufacturer's protocol followed by a 0.7 x SPRI bead (Beckman Coulter) cleanup. Library quantity and quality were assessed using the Agilent 2200 Tapestation system and libraries were subjected to next generation sequencing using an Illumina HiSeq1500 instrument (PE with no index; read 1: 61 bases (3 bases random nucleotides, 4 bases pool barcode, 53 bases specific sequence), read 2: 13 bases (6 bases cell barcode, 6 bases unique molecular identifier)).

*Pre-processing, quality assessment and control of single cell transcriptome data*. From sequenced data, pool barcodes, cell specific tags and Unique molecular identifiers (UMI) were extracted (576 cells sequenced). Subsequently, sequencing reads with ambiguous plate/cell-specific tags or UMI sequence with low quality (Phred < 27) and reads which map to *E. coli* were eliminated using Bowtie with parameters "-M 1 -t --best --chunkmbs 64 –strata". Next, fastq files were demultiplexed using the fastx_barcode_splitter from the fastx_toolkit and R1 reads (after trimming of pool barcode sequences) were mapped to the mouse mm10 & ERCC pseudo genome assembly using Bowtie "-m 1 -t --best --chunkmbs 64 –strata".
Valid reads were then counted using unique molecular identifiers if they mapped to the exon based gene model derived from Ensembl's biomart, mm10. Following this, a gene expression matrix was generated containing the number of unique UMIs associated with valid reads for every cell and every gene. Additionally, UMI sequencing errors were corrected for and filtered as described in (*27*).

*Filtering single cell transcriptome data*. In order to avoid biases introduced by low quality data we performed the following filtering of single cell data. Removal of cells with a ratio of mitochondrial versus endogenous genes exceeding 0.15 and cells with less than 320 molecule counts or less than 150 unique genes were removed from the analysis. Prior to analysis expression tables were filtered for mitochondrial, ribosomal and predicted genes to reduce noise.

*Analysis of MARS-seq single cell transcriptome data*. Analysis of the normalized and filtered single cell gene expression data (8657 genes across 408 single cell transcriptomes used in the final expression table) was done using several functions of the SEURAT single cell analysis package (*28*) and Monocle 2 (*29*). First, highly variable genes were determined as genes exceeding the dispersion threshold of 0.75. To infer the structure of the gene expression data a PCA was performed on the basis of highly variable genes. Following t-distributed stochastic neighbor embedding (tSNE)

DBScan clustering was performed to identify clusters of cells. The optimal number of clusters was identified by calculating several cluster indices by the NbClust R package(*30*). Relative expression for a cell was calculated as gene expression of a gene/gene set in relation to the total molecule counts in this cell. To identify clusters within the MARS-seq data the relative gene expression profiles of cell type specific gene signatures were overlaid to the tSNE plots. Pseudotime analysis was performed by the Monocle 2 algorithm by genes exceeding the average expression cutoff of 1 while having an empirical dispersion higher than 1. To analyze expression of single genes, relative expression was overlaid onto the tSNA plots or visualized as dot plots.

*Generation of cell signatures for analysis of single-cell data*. In order to unambiguously identify cell state specific genes for EMPs, pMacs and early macrophages we generated exclusive gene signatures. Here, differentially expressed (DE) genes were identified by a 1-way ANOVA model (|FC| > 1.4, FDR-adjusted p-value < 0.05 (*31*)) between EMPs (E10.25 yolk sac and fetal liver), pMacs (E10.25 yolk sac, fetal liver, head and limbs) and early macrophages (E10.25 and E10.5 yolk sac, fetal liver, head and limbs). The non-overlapping DE genes between these three contrasts were chosen as exclusive gene signatures for further analysis.

*Generation of tissue signatures for analysis of single-cell data*. To identify genes that are upregulated in tissue macrophages in relation to early macrophages a 1-way ANOVA model (|FC| > 1.5, FDR-adjusted p-value < 0.05 (*31*)) was calculated. Upregulated non-overlapping DE genes between early macrophages (E10.25 and E10.5, fetal liver, head and limbs/skin) vs brain macrophages (E14.5 and E18.5 brain), liver macrophages (E14.5 and E18.5 liver) or limb/skin macrophages (E14.5 and E18.5 limb/skin) were used as tissue macrophage specific gene signatures. Furthermore, a common early macrophage signature was defined as being upregulated in early macrophages (E10.25 and E10.5 fetal liver, head and limbs) vs. all other late tissue macrophages (E14.5 and E18.5 liver, head and limbs/skin). We assessed enrichment of these signatures in scRNA-seq data from pMacs by calculating an relative enrichment score for each signature in pMacs (molecule count of signature genes/(total molecule count * number of signature genes).

*Enrichment of tissue signatures in scRNA-seq data*. To assess enrichment of tissue macrophage-specific signatures or differentiation signatures in scRNA-seq data from pMacs, we identified genes with multimodal expression using Hartigan's Dip test statistic for unimodality. Next, multimodal genes were grouped by hierarchical clustering and enrichment of signatures within the clusters tested using hypergeometric testing with FDR correction (Benjamini-Hochberg).

**Data access and availability**. Sequencing datasets described in this work have been submitted to the Gene Expression Omnibus (GEO) repository (GEO accession number for bulk RNA-seq: GSE81686, accession numbers for scRNA-seq and *Id3*$^{+/-}$ and *Id3*$^{-/-}$ littermates pending). Additionally, a genome browser track hub, as well as additional supplementary data are available at the following URL: http://macrophage-development.computational-epigenetics.org/.

1. E. L. Gautier *et al.*, Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nature immunology* **13**, 1118-1128 (2012).
2. N. N. Gorgani, Y. Ma, H. F. Clark, Gene signatures reflect the marked heterogeneity of tissue-resident macrophages. *Immunology and cell biology* **86**, 246-254 (2008).
3. Y. Lavin *et al.*, Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312-1326 (2014).
4. E. Gomez Perdiguero *et al.*, Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature* **518**, 547-551 (2015).
5. C. Schulz *et al.*, A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science* **336**, 86-90 (2012).
6. K. Maeda *et al.*, Wnt5a-Ror2 signaling between osteoblast-lineage cells and osteoclast precursors enhances osteoclastogenesis. *Nature medicine* **18**, 405-412 (2012).
7. Q. Zhao *et al.*, Developmental ablation of Id1 and Id3 genes in the vasculature leads to postnatal cardiac phenotypes. *Developmental biology* **349**, 53-64 (2011).
8. D. Lyden *et al.*, Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature* **401**, 670-677 (1999).
9. Z. Guo *et al.*, Modeling Sjogren's syndrome with Id3 conditional knockout mice. *Immunol Lett* **135**, 34-42 (2011).
10. L. Deng *et al.*, A novel mouse model of inflammatory bowel disease links mammalian target of rapamycin-dependent hyperproliferation of colonic epithelium to inflammation-associated tumorigenesis. *The American journal of pathology* **176**, 952-967 (2010).
11. T. Yokomizo *et al.*, Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos. *Nature protocols* **7**, 421-431 (2012).
12. V. Sreeramkumar *et al.*, Neutrophils scan for activated platelets to initiate inflammation. *Science* **346**, 1234-1238 (2014).
13. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
14. P. G. Engstrom *et al.*, Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature methods* **10**, 1185-1191 (2013).
15. J. Li *et al.*, Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep*, (2015).
16. S. Picelli *et al.*, Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols* **9**, 171-181 (2014).
17. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25 (2009).
18. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, R36 (2013).

19. P. Glaus, A. Honkela, M. Rattray, Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics* **28**, 1721-1728 (2012).
20. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
21. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
22. N. C. Sheffield, C. Bock, LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*, (2015).
23. M. Sanchez-Castillo *et al.*, CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic acids research* **43**, D1117-1123 (2015).
24. C. Bock *et al.*, Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439-452 (2011).
25. J. Reimand, T. Arak, J. Vilo, g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic acids research* **39**, W307-315 (2011).
26. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550 (2005).
27. D. A. Jaitin *et al.*, Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776-779 (2014).
28. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* **33**, 495-502 (2015).
29. C. Trapnell *et al.*, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology* **32**, 381-386 (2014).
30. M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, Nbclust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J Stat Softw* **61**, 1-36 (2014).
31. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289-300 (1995).