iMAC: An interactive atlas to explore phenotypic differences between *in vivo*, *ex vivo* and *in vitro*-derived myeloid cells in the Stemformatics platform.

Nadia Rajab[1,2], Paul W Angel[1], Yidi Deng[1,6], Jennifer Gu[1], Vanta Jameson[3], Mariola Kurowska-Stolarska[4], Simon Milling[4], Chris M Pacheco[1], Matt Rutar[1], Andrew L Laslett[5,6], Kim Anh Le Cao[7], Jarny Choi[1] and Christine A Wells*[1].

AFFILIATIONS:

1. The Centre for Stem Cell Systems, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, 30 Royal Parade Parkville, Victoria 3010, Australia.
2. CSIRO Synthetic Biology Future Science Platform
3. Melbourne Cytometry Platform (MBC node), Department of Anatomy and Neuroscience, University of Melbourne
4. The Institute of Infection, Immunity and Inflammation, Rheumatoid Arthritis Pathogenesis Centre of Excellence (RACE), The University of Glasgow.
5. CSIRO Manufacturing, Clayton, Victoria 3168, Australia
6. Australian Regenerative Medicine Institute, Monash University, Victoria 3800, Australia.
7. Melbourne Integrative Genomics, School of Mathematics and Statistics, Faculty of Science, The University of Melbourne, 30 Royal Parade Parkville, Victoria 3010, Australia

*CORRESPONDING AUTHOR email wells.c@unimelb.edu.au

[Key words: monocyte, macrophage, dendritic cell, microglia, langerhan cell, tissue resident macrophage, hematopoietic progenitor, monocyte-derived macrophage, pluripotent stem cell-derived macrophage.]

Summary:

Understanding the myriad of myeloid subsets and identifying specific and discrete phenotypes is challenging in a human setting. Laboratory models of human myeloid biology lack tissue context, but the differences between common models and *in vivo* phenotypes have not been systematically evaluated. Here, we assemble a large transcriptional atlas of human myeloid biology representing ~1000 samples, comparing freshly isolated, cultured, and stem-cell derived myeloid cell types. Common myeloid models including pluripotent stem cell-derived macrophages fail to recapitulate developmental or tissue contexts. A shared feature of pluripotent stem cell-derived macrophages is the atypical expression of collagen, in addition to being highly efferocytotic. The resulting iMAC atlas is available at the Stemformatics.org platform, where users can review gene expression as well as upload and benchmark their own samples for comparison against a library of human tissue-resident and laboratory models of macrophage or dendritic cell biology.

Introduction

Macrophages are innate immune cells that are found resident in every tissue, with roles in tissue homeostasis, and response to infection or injury. The distinct functional roles of tissue macrophages are reflected in their transcriptional phenotypes: atlases of tissue-resident mouse macrophages, for example, have given great insight into their complexity and heterogeneity (Gautier et al., 2012; Gosselin et al., 2017). Individual transcriptome studies have played an essential role in unravelling the importance of the environment on macrophage phenotype and function (reviewed by (Huang and Wells, 2014)). These data evidence the shared molecular pathways of myeloid cells responding to pathogenic challenge, and the receptiveness of macrophages to environmental cues.

Much of our understanding of macrophage biology, including many of the molecular mechanisms of innate immune signaling, have arisen from mouse gene knock-out studies. However, cross-species comparisons of immune cells highlight differences between mouse and man. These include the glycolytic switch associated with metabolic reprogramming in activated mouse macrophages (Vijayan et al., 2019); divergent patterns of pathogen receptor expression (Vijayan et al., 2012) and transcriptional responses to innate immune stimuli (Schroder et al., 2012). Cross-species comparisons are further hampered by the absence of population level immune-activation maps, with most mouse studies in macrophage biology conducted on a limited number of inbred lines.

The need for improved molecular models of primary human cells is evident from the rising popularity of single cell transcriptomic atlases, exemplified by the human cell atlas consortium (Hay et al., 2018; Regev et al., 2017). However, unbiased profiling of cells also requires computational predictions of cell identity, raising further questions about how best to accurately identify immune cell populations resident in tissues, and discriminate these from circulating or infiltrating peripheral blood cells. The isolation, and identification of tissue-resident myeloid cells can be particularly fraught if populations are rare or hard to isolate using enzymatic or other dissociation methods. These procedures can alter myeloid transcriptomes (Gosselin et al., 2017), resulting in underrepresentation or phenotypic ambiguity of resident macrophages in single cell maps of a tissue. It might be argued that human macrophages suffer from an identity crisis, relying on equivalency to laboratory models of human macrophage biology such as ex-vivo culture of monocyte-derived macrophages, which may not be appropriate as a benchmark for specialized tissue functions.

The potential to model tissue residency, disease phenotypes and activation status of human macrophages using pluripotent stem cells (PSC) is a growing area of interest (Reviewed (Lee et al., 2018; Rajab et al., 2018)). However, the anatomical context or developmental ontogeny of these cells is still not well understood, nor their capacity to model specialized behaviors of myeloid cells including roles within a tissue niche. While databases such as BloodSpot (Bagger et al., 2016) and Haematlas (Watkins et al., 2009) provide a useful snapshot of gene expression of blood types, these lack depth with regard to tissue representation, common laboratory models or activating stimuli. Here, we describe an integrated myeloid transcriptome atlas to identify and benchmark human myeloid subpopulations from *ex vivo*, *in vivo* and *in vitro*

sources. Our motivation was to construct a reference atlas of human myeloid biology that draws on important studies already in the public domain, and that can be added to by the research community as new cell models or new profiling platforms become available. The atlas is made available as an interactive online resource https://www.stemformatics.org/atlas/imac.

A reference atlas for human myeloid biology.

We first compiled a reference transcriptional atlas (Figure 1A) from public and proprietary transcriptomic data from 44 studies and ~1000 samples representing peripheral blood monocytes, tissue-resident, *ex vivo* and *in vitro*-derived macrophages and dendritic cells. Samples were curated with respect to phenotype, source and isolation method. Datasets were processed through the Stemformatics pipeline which includes a stringent set of quality control requirements for hosting on the Stemformatics.org portal (Choi et al., 2019a). We constructed the atlas by implementing a two-step process (Angel et al., 2020). Firstly, transformation of expression values from the original studies to percentile values to facilitate the comparison of different experimental series was performed. Secondly, using a univariate estimation of their platform dependence, genes whose expression values were significantly impacted by the way that they were measured were removed from the atlas. This approach led to reproducible clustering of distinct myeloid classes on a PCA (Figure 1B and Supplementary Figure 1). Variables such as progenitor source (hematopoietic stem cell (HSC), monocyte-derived, or pluripotent stem cell (PSC)-derived), culture status or activating condition contributed to clustering of samples. The reproducibility of myeloid subsets including dendritic cells, monocytes and neutrophils was validated by projecting an independent RNAseq dataset of well annotated blood cell types from Haemopedia (Choi et al., 2019b) (Figure 1C).

We next assessed the usefulness of the atlas as an annotation reference tool for single cell data. We took two datasets describing blood monocytes and dendritic cells (Villani et al., 2017) and (Dutertre et al., 2019) which had previously not shown clear agreement on the identity of dendritic cell (DC) subsets (Figure 1D). Using the iMAC atlas as the reference, the Villani scRNA seq clusters clearly aligned to respective DC and monocyte groups, with a Capybara similarity estimate (Figure 1E) predicting that the Villani data had identified 2 distinct plasmacytoid DC (pDC) subsets, 2 distinct DC2 (CD1c+) subsets and an intermediate subset that sat between the classical monocyte and DC2 clusters. The DC5 cluster in the original study was described as an AXL+ SIGLEC6+ (AS-DC) subset that shared classical cell surface markers with pDCs. AS-DCs were suggested to be the subset contaminating traditional pDC isolation strategies with the capacity to stimulate T-cells. The Dutertre DC2 subsets aligned closely with atlas DC2 cells. Thus, the iMAC atlas provides a useful annotation tool to benchmark myeloid datasets that does not require identification of classical cell surface markers, and that can predict novel or intermediate cell clusters.

Monocytes acquire a 'Culture-Phenotype'

Monocytes are post-mitotic blood cells derived from bone marrow progenitors that are short lived in circulation and can repopulate macrophages in some tissue niches. The largest population of circulating monocytes is marked by high levels of expression of the LPS co-receptor CD14, which is typically used to isolate monocytes from blood. Intermediate and

nonclassical subsets are marked by acquisition of the type III FcR□, CD16 (Schmidl et al., 2014) and are included in the atlas. Cultured monocytes have been previously described as 'activated', but while we observe a distinct culture phenotype (Figure 2), the transcriptome of cultured cells mimics many of the features of a monocyte after extravasation into tissue. Figure 2B shows the grouping of peripheral blood monocytes, in a distinct cluster to monocytes that have been exposed to tissue culture plastic and culture media. This culture phenotype is typified here by a decrease in endothelial-adhesion proteins including the selectin SELL (Figure 2D). Regulators of RAS/RAF signaling including SPRED2 (Wakioka et al., 2001) have an elevated expression in cultured monocytes (Figure 2E), consistent with spreading and migration across tissue culture plastics. The classical monocyte-derived macrophage (MDM) requires *in vitro* differentiation of monocytes using several days exposure to growth factors such as macrophage-colony stimulating factor (M-CSF; CSF-1) or granulocyte-macrophage colony-stimulating factor (GM-CSF). These group distinctly from the cultured monocyte cluster, spreading further upwards along the culture axis (Figure 2F).

One interactive feature of the atlas is that gene expression and cell source can be explored together. The culture phenotype acquired by monocytes appears to be a prelude to activation, which can be observed along an adjacent axis in Figure 2F and is exemplified by the expression of IL6. Pathogen-activated phenotypes of cultured monocytes or MDM are typified by high expression of this and other cytokines, but this is. Cultured monocytes express higher levels of chemokines, such as CCL2, than peripheral monocytes (Figure 2G). Culture induces the expression of SLAMF1 (Supplementary Table 1), which has shown to be necessary for TLR4 activation in human macrophages (Yurchenko et al., 2018). Cultured monocytes also express higher levels of ITGB8 than circulating monocytes (Supplementary Table 1), which is necessary for activating latent TGF-β (Kelly et al., 2018).

Primary tissue-resident cells segregate into 'DC-like' and 'monocyte-like' communities.

Circulating and tissue resident dendritic cells occupied a distinct transcriptional niche but share the activation axis seen in monocytes and macrophages. *In vitro* differentiated dendritic cells, expanded from cord blood isolated hematopoietic progenitor cells and monocytes, are considerably removed from the *in vivo* cells (Figure 1 D and Figure 3A), including down-regulation of receptors such as CX3CR1, IL18R1 and TLR7 (Figure 3B and Supplementary Table 4). Other molecules, such as the cell-fusion protein DC-STAMP are gained in culture (Figure 3C and Supplementary Table 4). In terms of proximity to other atlas cell types, these appear high on the activation axis and are closely associated with monocyte-derived macrophages. So, while providing useful models of myeloid biology, cord blood-derived dendritic cells don't adequately capture key aspects of *in vivo* myeloid biology.

Isolation of primary tissue-resident macrophages is particularly difficult as this can result in alterations in phenotype (Gosselin et al., 2017). The difficulty of isolating tissue-resident populations from healthy human tissue is evident from the spread of tissue resident macrophages in comparison to tissue-resident dendritic cells (Figure 3D), noting that several of the macrophage datasets were obtained through surgical biopsies from patients with inflammatory disease, and indeed mapped across the inflammatory axis. Tissue resident

macrophages, including Kupffer cells, microglia, alveolar macrophages, gut and synovial macrophages occupy a broad niche on the atlas between dendritic cells, peripheral blood monocytes, and cultured monocytes (Figure 3D). Although there was no discrete partitioning of macrophages from individual tissues, distinct classes of tissue resident macrophages were observed in the atlas –(the alveolar, colon and macrophages isolated from tumour ascites (TAM) (Figure 3D and Supplementary Figure 2A) grouped together between cultured monocytes and CD1c+ dendritic cells, and a second spread containing synovial macrophages, microglia and Kupffer cells aligned with the iPSC-derived cells.

Primary microglia included in the iMAC atlas include both *in vivo* isolated fetal and cultured *ex vivo* fetal and adult microglia (Supplementary Figure 2A). The profiles of *in vivo* isolated fetal microglia cluster apart from the spread of *ex vivo* cultured adult and fetal microglia. Cultured tissue macrophages including cultured primary microglia or Kupffer cells shared a broad transcriptional signature with monocyte-derived macrophages, and pluripotent-stem cell derived myeloid cells.

PSC-derived macrophages share features with tissue-resident macrophages despite poor maturation

Macrophages derived from human PSCs offer new opportunities to model *in vivo* macrophage biology. When reviewing the studies contributing to this atlas, we noted that PSC-derived macrophages are typically benchmarked against monocyte-derived macrophages (MDMs), or cultured primary cells, using a suite of phenotyping techniques. Each experiment includes a small number of samples for transcriptional profiling, with a few notable exceptions (Alasoo et al., 2018). We argue that, given the spectrum of possible resident tissue macrophage phenotypes, it would be more useful to compare PSC-derived cells against an atlas of possible macrophage phenotypes. Whilst several groups reuse publicly available tissue macrophage data, the opportunity to carry out large-scale comparisons to different primary myeloid cells has been limited by the availability of relevant data on a compatible platform.

Microglia represent just over a third of PSC-directed myeloid differentiation studies in the atlas. These do not resolve into a unique cluster but share transcriptional phenotypes with PSC-derived and tissue-resident macrophages, which sit deep in the broad culture axis associated with the expression of lipid-scavenging genes required for efferocytosis. The exception are ‘cytokine-matured’ PSC-derived microglia samples from (Abud et al., 2017). These are closest to the *in vivo* fetal microglia samples of the PSC-microglia but are also closely associated with other primary tissue resident macrophages from lung, joint and gut. The atlas does provide an opportunity to review the expression of markers thought to distinguish microglia from other primary macrophages. TMEM119, for example, is largely restricted to primary or PSC-derived microglia, although the (Abud et al., 2017) samples have low expression of this marker (Supplementary Figure 2B). P2RY12 is variably expressed across all microglial samples, but its expression is also evident in different tissue-resident samples including those derived from gut and synovial tissues (Figure 3D).

Dendritic cells and macrophages are known professional antigen-presenting cells. The initiation of adaptive immune responses requires presentation of antigen through major histocompatibility complex I or II. The majority of *in vitro*-derived macrophages have low expression of HLA relevant genes (Supplementary Table 2) including CIITA, a known master regulator of MHCII gene expression, which suggests poor maturation in regard to their antigen presentation. Nevertheless, some *in vitro*-derived macrophages cultured with stimulating factors (Supplementary Figure 2C) do show inducible CIITA expression, demonstrating that they have the capacity to express antigen-presenting machinery. It is also worth noting that there are *in vitro*-derived macrophages, and microglia, (Abud et al., 2017; Honda-Ozaki et al., 2018) that do appear to have high CIITA expression without stimulating factors (Supplementary Figure 2C). For microglia samples (Abud et al., 2017), this may be an impact of long term culture conditions in addition to supplementation of CX3CL1 and CD200, the latter confirming the authors' observations. In regards to macrophage samples (Honda-Ozaki et al., 2018), this may be an impact of immortalization of progenitor cells.

PSC-derived macrophages display transcriptional hallmarks of efferocytosis

There has been growing interest in the importance of metabolic reprogramming in macrophage responses, so we asked whether media supplementation could explain the spread of PSC-derived macrophages on the iMAC atlas. All PSC-derivation protocols supplement media with fatty- or amino-acids, including L-Glutamine, non-essential amino acids (NEAA), Linoleic and Linolenic acids. Some methods add fetal bovine or calf serum, but there was no obvious correlation between serum addition and without. Overall, factors are so ubiquitously used that supplementation alone could not explain the differences between PSC and cultured primary macrophages.

Lipid homeostasis is an important role for resident tissue macrophages. A high proportion of genes differentially expressed between *in vitro*-derived macrophages/microglia/kupffer cells and tissue resident cells are involved in lipid transport, catabolism and in buffering the cells from concomitant stresses associated with lipid turn-over. For example, reduced expression of ABCA6 is consistent with high efflux of cholesterol from these macrophages (Supplementary Figure 3A). Higher levels of mitochondrial acyl-CoA dehydrogenase ACADM (Supplementary Figure 3B) and phosphatidate phosphatase LPIN3 (Supplementary Figure 3C) suggests high lipid turnover.

Efferocytosis, or apoptotic cell clearance, has broad immunomodulatory effects (Reviewed by (Elliott et al., 2017)) and active engulfment and clearance of cells by PSC-macrophages is clearly observed in the absence of any inflammatory activation (Supplementary Video). Efferocytosis modulates macrophages from a pro-inflammatory phenotype to one with resolving qualities (Yamaguchi et al., 2014), consistent with the patterns of gene expression observed in cultured macrophages. A study (Cao et al., 2019) demonstrated higher lipid uptake in PSC-macrophages compared to peripheral blood MDMs, concordant with higher expression of efferocytosis-related genes including S1PR1 and MERTK. We confirm that MERTK is generally highly expressed in PSC-derived macrophages, but that there is also a tissue-resident distribution of MERTK expression, with very low levels observed in primary alveolar

macrophages, and highest levels observed in human fetal microglia (Supplementary Figure 2). Tissue-resident macrophages are known first-responders to tissue damage and are key in orchestrating inflammation and its subsequent resolution. This appears to be a phenotype that is selected for in cultured macrophages and may be an inevitable consequence of derivation that strives for high cell yield.

PSC-derived macrophages do not recapitulate a developmental hematopoietic ontogeny.

Many PSC-derived systems recapitulate fetal, rather than adult phenotypes, so it is no surprise that others have argued that PSC-derivation protocols mimic primitive rather than definitive myeloid biology. This is largely based on discriminating MYB expression in progenitor cells, which is associated with definitive hematopoiesis and has high expression in hematopoietic progenitor cells. It is clear that MYB is not required for PSC-derived myelopoiesis as macrophages are derived in MYB-KO embryonic stem cells (Buchrieser et al., 2017). Nevertheless, MYB is highly and ubiquitously expressed in PSC-myeloid progenitors, including common myeloid progenitors and hemogenic endothelium, and is retained at high levels in some PSC-derived microglia (Supplementary Figure 3D).

In comparison to primary cell datasets, PSC-macrophage studies that were overlaid onto the iMAC atlas grouped broadly with cultured monocytes and tissue-resident macrophages. Despite arguments on recapitulation of fetal origin, PSC-macrophages, including PSC-microglia and PSC-Kupffer cells, form an extended group associated with high expression of the human homologue of the F4/80 antigen, ADGRE1, as well as high expression of lipid-scavenging receptors such as SCARB1 (Supplementary Table 2, 3 and 5). Furthermore, MAF expression is indistinguishable in macrophages of different origin (Supplementary Figure 3E). Interestingly, LIN28B, appears to be highly expressed in *in vitro*-derived macrophages compared to *in vivo* and *ex vivo* cells (Supplementary Figure 3F), which may point to incomplete silencing of the Let7 microRNA pathway and maintenance of a fetal state(Grant Rowe et al., 2016; Zhou et al., 2017), but is also characteristic of myeloid leukemias.

Some other phenotypes previously attributed to ontogeny in PSC-derived cells may rather reflect a more general culture context. For example, ADGRE1 (F4/80) expression has been attributed to yolk-sac derived myeloid cells in mouse (Schulz et al., 2012). While high on PSC-derived cells, ADGRE1 is also clearly upregulated in culture. This is exemplified through primary human microglia that have low expression of ADGRE1 in comparison to *ex vivo* culture or PSC-derived cells (Supplementary Table 5).

PSC-Macrophages express high levels of collagen

Collagen production and deposition alongside extracellular matrix remodeling are processes involved in wound healing and scarring. Macrophages are instrumental in instructing tissue repair, particularly through the production of growth factors such as TGF-β, IGF1 and PDGF (Shook et al., 2018). Secreted growth factors drive fibroblasts and endothelial cells to produce extracellular matrix components, promoting keloid formation as well as angiogenesis. This model has macrophages influencing collagen deposition by neighboring stromal cells, however,

in some instances are capable of contributing to collagen deposition as demonstrated in mouse and zebrafish injury models (Simões et al., 2020). Gene-set enrichment analysis of the genes that are most correlated with *in vitro*-derived macrophages moving away (Figure 4A) from the tissue resident populations revealed that the most significant pathways in these cells involved collagen synthesis and production (Table 1). A STRING protein-protein interaction network (Figure 4B) shows that this phenotype is significantly enriched for highly connected matrix remodeling, collagen deposition and cadherin-mediated cell-cell and cell-matrix interactions.

Perhaps pluripotent stem cell-derived cells are being driven to adopt a pro-fibrotic phenotype through the derivation or culture contents. Initial observations on analysis of myeloid-, pluripotent stem cell- and hematopoietic progenitor-derived cells, highlight higher expression of collagen genes in pluripotent stem cell-derived cells (Figure 4C). To investigate whether this phenotype is due to culture impact, we isolated peripheral blood monocytes and derived human PSC-progenitors and cultured these cells with the same culture media in the presence of M-CSF for 5 days to drive macrophage differentiation. On day 5, cells were either stimulated with LPS for 2 hours before extraction or extracted as control samples for sequencing analysis. Analysis of sequencing samples showed the same enriched collagen and cadherin networks with high expression of these genes observed in PSC-macrophages, with little or no expression in monocyte-derived macrophages regardless of whether cells were stimulated or not (Figure 4D). To determine this at a protein level, we carried out intracellular flow cytometry analysis of type I collagen in PSC- and monocyte-derived cells (Figure 4E). Compared with staining controls, we observed higher levels of type I collagen in PSC-macrophages. Collectively, our results suggest that although culture media can impact cell phenotype, it is not the main driver for this pro-fibrotic phenotype in PSC-macrophages in the final stages of differentiation.

Concluding Remarks

Human macrophage biology is integral to the development homeostasis and disease mechanisms in every tissue in the body, but our understanding of human myeloid biology is limited by the quality by the models available to us. Here we describe an integrated myeloid transcriptome atlas as a novel resource to identify myeloid cells in single cell datasets and to benchmark *in vitro* models of *in vivo* biology. Implementation in Stemformatics enables users to upload their own data to benchmark cell types against the atlas for rapid and intuitive cell-classification.

By benchmarking curated public data of PSC-macrophages and their precursor cells against the atlas, it is apparent that these represent neither definitive nor primitive myelopoiesis, or rather, that they imperfectly recapitulate aspects of both. PSC-conditions clearly do not mimic the developmental time-frame nor tissue niche of yolk-sac, fetal liver or bone marrow. In our hands PSC-macrophages display transcriptional hallmarks of efferocytosis and surprisingly collagen production, which may suggest that the derivation process of these cells are driving a pro-fibrotic phenotype. Perhaps the ontogeny question is less interesting than understanding the molecular networks that can be co-opted in stem cell derived myeloid models, to deliver specific phenotypic properties outside of the constraints of development.

Acknowledgements

The authors thank Tyrone Chen and Othmar Korn for assistance with data processing and Isha Nagpal for website development to support the iMAC interactive viewer. The authors thank Ramaciotti Centre for Genomics (University of New South Wales; Sydney) for mRNA sequencing, and the Melbourne Cytometry Platform for flow cytometry assistance.This work was funded by Stem Cells Australia, an Australian Research Council Special Research Initiative [SRI110001002]; Wellcome Trust funding and acknowledgement of UoG facilities; NR is funded by the Centre for Stem Cell Systems and the CSIRO Synthetic Biology Future Science Platform. CAW is funded by a Future Fellowship from the Australian Research Council [FT150100330]; JC is funded by the JEM Research Foundation to the Stem Cell Atlas.

Author Contributions

Conception NR, JC, CAW; Experimental Investigation and Interpretation NR, VJ, JG, CAW; Experimental Resources ALL, CAW; Methodology PWA, JC, YD; Data provider NR, MKS, SM; Curation NR, MR, CMP, CAW; Statistical analysis YD, KALC, PWA; Writing – original draft NR, CAW; Writing - review and editing NR, PWA, SM, ALL, KALC, JC, CAW; Supervision CAW, ALL, KAC; Project Funding - CAW.

Declaration of Interests

None

References

Abud, E.M., Ramirez, R.N., Martinez, E.S., Healy, L.M., Nguyen, C.H.H., Newman, S.A., Yeromin, A. V., Scarfone, V.M., Marsh, S.E., Fimbres, C., et al. (2017). iPSC-Derived Human Microglia-like Cells to Study Neurological Diseases. Neuron *94*, 278-293.e9.

Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., Hale, C., Dougan, G., and Gaffney, D.J. (2018). Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat. Genet. *50*, 424–431.

Angel, P.W., Rajab, N., Deng, Y., Pacheco, C.M., Chen, T., Le Cao, K.-A., Choi, J., and Wells, C. (2020). A simple, scalable approach to building a cross-platform transcriptome atlas. BioRxiv 2020.03.09.984468.

Bagger, F.O., Sasivarevic, D., Sohi, S.H., Laursen, L.G., Pundhir, S., Sønderby, C.K., Winther, O., Rapin, N., and Porse, B.T. (2016). BloodSpot: A database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. Nucleic Acids Res. *44*, D917–D924.

Buchrieser, J., James, W., and Moore, M.D. (2017). Human Induced Pluripotent Stem Cell-Derived Macrophages Share Ontogeny with MYB-Independent Tissue-Resident Macrophages. Stem Cell Reports *8*, 334–345.

Cao, X., Yakala, G.K., van den Hil, F.E., Cochrane, A., Mummery, C.L., and Orlova, V. V. (2019). Differentiation and Functional Comparison of Monocytes and Macrophages from hiPSCs with Peripheral Blood Derivatives. Stem Cell Reports *12*, 1282–1297.

Choi, J., Pacheco, C.M., Mosbergen, R., Korn, O., Chen, T., Nagpal, I., Englart, S., Angel, P.W., and Wells, C.A. (2019a). Stemformatics: visualize and download curated stem cell data. Nucleic Acids Res. *47*, D841–D846.

Choi, J., Baldwin, T.M., Wong, M., Bolden, J.E., Fairfax, K.A., Lucas, E.C., Cole, R., Biben, C., Morgan, C., Ramsay, K.A., et al. (2019b). Haemopedia RNA-seq: a database of gene expression during haematopoiesis in mice and humans. Nucleic Acids Res. *47*, D780–D785.

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 Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. Immunity *51*, 573-589.e8.

Elliott, M.R., Koster, K.M., and Murphy, P.S. (2017). Efferocytosis Signaling in the Regulation of Macrophage Inflammatory Responses. J. Immunol. *198*, 1387–1394.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res. *46*, D649–D655.

Gautier, E.L., Shay, T., Miller, J., Greter, M., Jakubzick, C., Ivanov, S., Helft, J., Chow, A., Elpek, K.G., Gordonov, S., et al. (2012). Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. Nat. Immunol. *13*, 1118–1128.

Gosselin, D., Skola, D., Coufal, N.G., Holtman, I.R., Schlachetzki, J.C.M., Sajti, E., Jaeger, B.N., O'Connor, C., Fitzpatrick, C., Pasillas, M.P., et al. (2017). An environment-dependent transcriptional network specifies human microglia identity. Science (80-.). *356*, 1248–1259.

Grant Rowe, R., Wang, L.D., Coma, S., Han, A., Mathieu, R., Pearson, D.S., Ross, S., Sousa, P., Nguyen, P.T., Rodriguez, A., et al. (2016). Developmental regulation of myeloerythroid progenitor function by the Lin28b-let-7-Hmga2 axis. J. Exp. Med. *213*, 1497–1512.

Hay, S.B., Ferchen, K., Chetal, K., Grimes, H.L., and Salomonis, N. (2018). The Human Cell Atlas bone marrow single-cell interactive web portal. Exp. Hematol. *68*, 51–61.

Honda-Ozaki, F., Terashima, M., Niwa, A., Saiki, N., Kawasaki, Y., Ito, H., Hotta, A., Nagahashi, A., Igura, K., Asaka, I., et al. (2018). Pluripotent Stem Cell Model of Nakajo-Nishimura Syndrome Untangles Proinflammatory Pathways Mediated by Oxidative Stress. Stem Cell Reports *10*, 1835–1850.

Huang, E., and Wells, C.A. (2014). Environmental Influences Interface of Genetic, Epigenetic, and Responsiveness Is Determined at the The Ground State of Innate Immune Downloaded from. J Immunol J. Immunol. J. Immunol. *193*, 13–19.

Jones, J.C., Sabatini, K., Liao, X., Tran, H.T., Lynch, C.L., Morey, R.E., Glenn-Pratola, V., Boscolo, F.S., Yang, Q., Parast, M.M., et al. (2013). Melanocytes derived from transgene-free human induced pluripotent stem cells. J. Invest. Dermatol. *133*, 2104–2108.

Joshi, K., Elso, C., Motazedian, A., Labonne, T., Schiesser, J. V., Cameron, F., Mannering, S.I., Elefanty, A.G., and Stanley, E.G. (2019). Induced pluripotent stem cell macrophages present antigen to proinsulin-specific T cell receptors from donor-matched islet-infiltrating T cells in type 1 diabetes. Diabetologia *62*, 2245–2251.

Kelly, A., Gunaltay, S., McEntee, C.P., Shuttleworth, E.E., Smedley, C., Houston, S.A., Fenton, T.M., Levison, S., Mann, E.R., and Travis, M.A. (2018). Human monocytes and macrophages regulate immune tolerance via integrin αvβ8-mediated TGFβ activation. J. Exp. Med. *215*, 2725–2736.

Kong, W., Fu, Y.C., and Morris, S.A. (2020). Capybara: A computational tool to measure cell identity and fate transitions. BioRxiv 2020.02.17.947390.

Lee, C.Z.W., Kozaki, T., and Ginhoux, F. (2018). Studying tissue macrophages in vitro: are iPSC-derived cells the answer? Nat. Rev. Immunol. *18*, 716–725.

Ng, E.S., Davis, R., Stanley, E.G., and Elefanty, A.G. (2008). A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. Nat. Protoc. *3*, 768–776.

Rajab, N., Rutar, M., Laslett, A.L., and Wells, C.A. (2018). Designer macrophages: Pitfalls and opportunities for modelling macrophage phenotypes from pluripotent stem cells. Differentiation *104*, 42–49.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell Atlas. Elife *6*.

Schmidl, C., Renner, K., Peter, K., Eder, R., Lassmann, T., Balwierz, P.J., Itoh, M., Nagao-Sato, S., Kawaji, H., Carninci, P., et al. (2014). Transcription and enhancer profiling in human monocyte subsets. Blood *123*, e90–e99.

Schroder, K., Irvine, K.M., Taylor, M.S., Bokil, N.J., Le Cao, K.-A., Masterman, K.-A., Labzin, L.I., Semple, C.A., Kapetanovic, R., Fairbairn, L., et al. (2012). Conservation and divergence in Toll-like receptor 4-regulated gene expression in primary human versus mouse macrophages. Proc. Natl. Acad. Sci. U. S. A. *109*, E944-53.

Shook, B.A., Wasko, R.R., Rivera-Gonzalez, G.C., Salazar-Gatzimas, E., López-Giráldez, F., Dash, B.C., Muñoz-Rojas, A.R., Aultman, K.D., Zwick, R.K., Lei, V., et al. (2018). Myofibroblast proliferation and heterogeneity are supported by macrophages during skin repair. Science (80-.). *362*, eaar2971.

Simões, F.C., Cahill, T.J., Kenyon, A., Gavriouchkina, D., Vieira, J.M., Sun, X., Pezzolla, D., Ravaud, C., Masmanian, E., Weinberger, M., et al. (2020). Macrophages directly contribute collagen to scar formation during zebrafish heart regeneration and mouse heart repair. Nat. Commun. *11*, 1–17.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

Turlach, B.A., and Weingessel, A. (2019). quadprog: Functions to Solve Quadratic Programming Problems. R package version 1.5-8.

Vijayan, D., Radford, K.J., Beckhouse, A.G., Ashman, R.B., and Wells, C.A. (2012). Mincle polarizes human monocyte and neutrophil responses to *Candida albicans*. Immunol. Cell Biol. *90*, 889–895.

Vijayan, V., Pradhan, P., Braud, L., Fuchs, H.R., Gueler, F., Motterlini, R., Foresti, R., and Immenschuh, S. (2019). Human and murine macrophages exhibit differential metabolic responses to lipopolysaccharide - A divergent role for glycolysis. Redox Biol. *22*, 101147.

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 (80-.). *356*.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2019). SciPy 1.0--Fundamental Algorithms for Scientific Computing in Python. Nat. Methods *17*, 261–272.

Vlahos, K., Sourris, K., Mayberry, R., McDonald, P., Bruveris, F.F., Schiesser, J. V., Bozaoglu,

K., Lockhart, P.J., Stanley, E.G., and Elefanty, A.G. (2019). Generation of iPSC lines from peripheral blood mononuclear cells from 5 healthy adults. Stem Cell Res. *34*, 101380.

Wakioka, T., Sasaki, A., Kato, R., Shouda, T., Matsumoto, A., Miyoshi, K., Tsuneoka, M., Komiya, S., Baron, R., and Yoshimura, A. (2001). Spred is a Sprouty-related suppressor of Ras signalling. Nature *412*, 647–651.

Watkins, N.A., Gusnanto, A., de Bono, B., De, S., Miranda-Saavedra, D., Hardie, D.L., Angenent, W.G.J., Attwood, A.P., Ellis, P.D., Erber, W., et al. (2009). A HaemAtlas: characterizing gene expression in differentiated human blood cells. Blood *113*, e1-9.

Yamaguchi, H., Maruyama, T., Urade, Y., and Nagata, S. (2014). Immunosuppression via adenosine receptor activation by adenosine monophosphate released from apoptotic cells. Elife *3*.

Yurchenko, M., Skjesol, A., Ryan, L., Richard, G.M., Kandasamy, R.K., Wang, N., Terhorst, C., Husebye, H., and Espevik, T. (2018). SLAMF1 is required for TLR4-mediated TRAM-TRIF–dependent signaling in human macrophages. J. Cell Biol. *217*, 1411–1429.

Zhou, J., Chan, Z.L., Bi, C., Lu, X., Chong, P.S.Y., Chooi, J.Y., Cheong, L.L., Liu, S.C., Ching, Y.Q., Zhou, Y., et al. (2017). LIN28B activation by PRL-3 promotes leukemogenesis and a stem cell-like transcriptional program in AML. Mol. Cancer Res. *15*, 294–303.

Figure Titles and Legends

Figure 1: A reference atlas for human myeloid biology

(A) iMAC atlas with samples coloured by cell type. Navy blue - monocytes, blue - macrophages, aqua -Dendritic cells, dark green - CD141+ DC, light green - CD1c+ DC, yellow - pDC, brown – granulocytes, pink stem and progenitor cells, hemogenic endothelium. (B) iMAC atlas coloured by platform: red various microarray platforms, black RNAseq platforms. (C) Validation with Haemopedia RNAseq myeloid samples: diamond shape – monocytes, circle granulocytes, cross DC. (D) Single cell projections of Villani *et al.* (2017) and Dutertre *et al.* (2019) samples onto iMAC atlas (F) Heatmap derived from Capybara analysis of Villani *et al.* (2017) and Dutertre *et al.* (2019) samples compared to iMAC cell types. Colour gradients reflect similarity of single cell clusters to iMAC atlas cell types (dark least similar, to light most similar).

Figure 2: Monocytes acquire a culture phenotype

(A)Schematic of rolling monocytes, highlighting cultured cells mimic many of the features of a monocyte after extravasation (B) Cultured monocytes form a distinct cluster away from *in vivo* monocytes along PC3 (C) STRING-DB network of top-ranked genes differentially expressed between peripheral blood (*in vivo*, n=107) and cultured monocytes (*ex vivo*, n=171) indicating upregulation of cytoskeletal proteins and down regulation of endothelial-adhesion proteins (D) Ranked expression (Y-axis) of gene involved in endothelial adhesion, SELL, comparing cultured monocytes (n=171) with monocytes directly profiled from blood (*in vivo*, n=107). Grey stripe indicates variance attributable to platform. P-value: Mann-Whitney-Wilcoxon rank sum test (E)Ranked expression (Y-axis) of gene involved in the regulation of RAS/RAF signaling comparing cultured monocytes (n=171) with monocytes directly profiled from blood (*in vivo*, n=107). Grey stripe indicates variance attributable to platform. P-value: Mann-Whitney-Wilcoxon rank sum test (F) iMAC atlas coloured by ranked expression of IL-6 (Scale bar: high ranked

expression (dark red) to low ranked expression (grey) indicating axis of activation) (G) Ranked expression (Y-axis) of Chemokine CCL2 comparing cultured monocytes (n=171) with monocytes directly profiled from blood (*in vivo*, n=107). Grey stripe indicates variance attributable to platform. P-value: Mann-Whitney-Wilcoxon rank sum test.

Figure 3: Cultured and *in vitro*-derived dendritic cells do not capture aspects *in vivo* myeloid biology

(A) iMAC atlas coloured by cell source to highlight origin of dendritic cells contributes to clustering. Dark green samples are profiled directly from blood or tissue (*in vivo*); mid-green are cultured primary cells (*ex vivo*); light green are differentiated *in vitro* from cord blood or iPSC progenitors (*in vitro*). (B and C) Ranked expression (Y-axis) of (B) receptor CX3CR1 and (C) cel-fusion protein DC-STAMP *in vivo* dendritic cells (n=145), *ex vivo* dendritic cells (n=17) and *in vitro*-derived dendritic cells (n=57). Grey stripe indicates variance attributable to platform. P-value: Mann-Whitney-Wilcoxon rank sum test. (D) iMAC atlas coloured by ranked expression of P2RY12. (Scale bar: high ranked expression (dark red) to low ranked expression (grey)).

Figure 4: *in vitro*-derived macrophages do not capture developmental ontogeny

(A) iMAC atlas coloured by cell source to highlight *in vitro*-derived macrophages used for regression testing (B) STRING_DB Protein-Protein network of *in vitro*-derived macrophages highlights enrichment of collagen, growth factor and cadherin networks. Line color indicates the type of interaction evidence. Light blue solid lines indicate known interactions from curated databases, solid pink line indicates known interactions that have been experimentally determined, bright green lines indicate gene neighborhood predicted interactions, red lines indicate gene fusions predicted interactions, dark blue lines indicate gene co-occurrence predicted interactions, yellow/green lines indicate textmining, black lines indicate co-expression, and light purple lines indicate protein homology. (C) Violin plots of myeloid- (n=584), pluripotent stem cell- (n=116) and hematopoietic progenitor- (n=201) derived cells comparing expression of collagen genes (COL1A1 and COL4A2)(D) mRNA-seq gene expression from human peripheral blood monocyte-derived macrophages (HMDM) (n=3) and human pluripotent stem cell derived macrophages (PSCM) (n=2) samples (C= control, S = stimulated) (E) Intracellular flow cytometry analysis of human peripheral blood monocyte-derived macrophages (HMDM) and human pluripotent stem cell derived macrophages (PSCM), representative of 2 experimental repeats (n=2 HMDM, n=2 PSCM). Red = no primary antibody control, Black = isotype control, Purple = Type I Collagen stain.

Supplementary Figure Titles and Legends

Supplementary Figure 1: iMAC as a reference atlas and annotation tool

(A) Pre-gene filtering (left) and Post-gene filtering (right). Different platforms highlighted by different colours. (B) Kolmogorov–Smirnov (KS) statistics (y-axis) to assess the difference in gene expression distribution between pseudo-bulk single cells DC6 and bulk sample plasmacytoid dendritic cells from the atlas, with respect to the number of single cells that are aggregated (x-axis). Each line indicates one of thirty random sub-samplings with replacement

trial. KS statistics are calculated on each gene and averaged across all genes. A minimum KS statistic is obtained when aggregating 8 cells. (C) iMAC atlas cell types before single cell projection (D) Single cell projection of (Dutertre et al., 2019) and (Villani et al., 2017) samples onto the iMAC atlas where an 8 cells are aggregated based on (B) cells.

Supplementary Figure 2: Tissue Macrophages

iMAC atlas coloured by ranked expression of (A) MERTK (B) TMEM119 (Scale bar: high ranked expression (dark red) to low ranked expression (grey) (C) Ranked expression (Y-axis) of Class II transactivator (CIITA) in vivo versus in vitro-derived macrophages (gut, synovial, kupffer, microglia, macrophage). Red – activated, yellow – (Abud et al., 2017) microglia samples, khaki-Honda-Ozaki et al. (2018) macrophage samples.

Supplementary Figure 3: *in vitro*-derived macrophages do not capture developmental ontogeny

Ranked expression (Y-axis) of genes comparing *in vivo*, *ex vivo* and *in vitro*-derived macrophages (gut, synovial, kupffer, microglia, macrophage) (A, B, C) Lipid turnover (A) cholesterol efflux (B) mitochondrial acyl-CoA dehydrogenase (C) phosphatidate phosphatase. (D) iMAC atlas coloured by ranked expression of MYB (Scale bar: high ranked expression (dark red) to low ranked expression (grey). (E) Comparison of MAF expression from *in vivo* (n=61), *ex vivo* (n=26) and *in vitro*- (n=96) derived macrophages (gut, synovial, kupffer, microglia, macrophage). Grey stripe indicates variance attributable to platform. (E) Comparison of LIN28B expression from *in vivo* (n=61), *ex vivo* (n=26) and *in vitro*- (n=96) derived macrophages (gut, synovial, kupffer, microglia, macrophage). Grey stripe indicates variance attributable to platform. P-value MannWhitney-Wilcoxon rank sum test.

Tables with Titles and Legends

Table1: Gene-Set Enrichment Analysis

Table of the top 10 Reactome pathways enriched in genes highly correlated with *in vitro*-derived macrophage spread. Enrichment: number of genes in the list/number of genes in that pathway (False Discovery Rate-value). Genes: multiple entries assigned to the same gene indicated by underlining of gene symbol with UniProt accessions in brackets.

Table 1		
Pathway	**Enrichment**	**Genes**
Collagen biosynthesis and modifying enzymes	11/76 (3.76e-09)	ADAMTS3, COL1A2, COL4A2, SERPINH1, COL14A1, COL3A1, COL4A5, COL1A1, COL4A1 *(P02462, Q03692)*, COL5A1
Extracellular matrix organization	18/329 (4.48e-09)	ADAMTS1, COL1A1, COL4A1*(P02462, Q03692)*, COL5A1, LTBP1, PTPRS, ADAMTS3, COL1A2, COL4A2, KDR, LUM, SERPINH1, COL14A1, COL3A1, COL4A5, LAMB1, MFAP4
Collagen chain trimerization	9/44 (4.81e09)	COL14A1, COL3A1, COL4A5, COL1A1, COL4A1*(P02462, Q03692)*, COL5A1, COL1A2, COL4A2
Collagen formation	11/104 (2.39e-08)	ADAMTS3, COL1A2, COL4A2, SERPINH1, COL14A1, COL3A1, COL4A5, COL1A1, COL4A1*(P02462, Q03692)*, COL5A1
ECM proteoglycans	10/79 (2.39e-08)	COL1A1, COL4A1, COL5A1, PTPRS, COL1A2, COL4A2, LAMB1, COL3A1, COL4A5, LUM
Non-integrin membrane-ECM interactions	9/61 (3.83e-08)	COL1A1, COL4A1*(P02462, Q03692)*, COL5A1, COL1A2, COL4A2, LAMB1, COL3A1, COL4A5
Integrin cell surface interactions	10/86 (3.83e-08)	COL1A1, COL4A1*(P02462, Q03692)*, COL5A1, COL1A2, COL4A2, KDR, COL3A1, COL4A5, LUM
Assembly of collagen fibrils and other multimeric structures	9/67 (6.95e-08)	COL14A1, COL3A1, COL4A5, COL1A1, COL4A1*(P02462, Q03692)*, COL5A1, COL1A2, COL4A2
Collagen degradation	9/69 (7.93e-08)	COL14A1, COL3A1, COL4A5, COL1A1, COL4A1*(P02462, Q03692)*, COL5A1, COL1A2, COL4A2
Degradation of the extracellular matrix	11/148 (3.73e-07)	ADAMTS1, COL1A2, COL4A2, LAMB1, COL14A1, COL3A1, COL4A5, COL1A1, COL4A1*(P02462, Q03692)*, COL5A1

Supplementary Tables and Legends

Supplementary Table 1: *in vivo* vs. *ex vivo* monocytes

Comparison of gene expression of *ex vivo* (n=171) and *in vivo* (n=107) monocytes. Columns refer to gene symbols, P-values re-calculated by Mann-Whitney-Wilcoxon rank-sum test, mean and standard deviation.

Supplementary Table 2: *in vivo* vs. *in vitro* macrophages
Comparison of gene expression of *in vivo* (n=61) and *in vitro*- (n=96) derived macrophages (gut, synovial, kupffer, microglia, macrophage). Columns refer to gene symbols, P-values recalculated by Mann-Whitney-Wilcoxon rank-sum test, mean and standard deviation.

Supplementary Table 3: *in vivo* vs. *ex vivo* vs. *in vitro* macrophages
Comparison of gene expression of *in vivo* (n=61), *ex vivo* (n=26) and *in vitro*- (n=96) derived macrophages (gut, synovial, kupffer, microglia, macrophage). Columns refer to gene symbols, P-values re-calculated by Mann-Whitney-Wilcoxon rank-sum test, mean and standard deviation.

Supplementary Table 4: dendritic cells
Comparison of gene expression of *in vivo* (n=145), *ex vivo* (n=17) and *in vitro*- (n=57) derived dendritic cells. Columns refer to gene symbols, P-values re-calculated by Mann-WhitneyWilcoxon rank-sum test, mean and standard deviation.

Supplementary Table 5: *in vivo* vs. *ex vivo* vs. *in vitro* microglia
Comparison of gene expression of *in vivo* (n=10), *ex vivo* (n=21) and *in vitro*- (n=43) derived microglia. Columns refer to gene symbols, P-values re-calculated by Mann-Whitney-Wilcoxon rank-sum test, mean and standard deviation.

Supplementary Table 6: Datasets and samples to compile iMAC atlas and single cell projection
Tissue resident macrophages and dendritic cells from peripheral blood, spleen, thymus, joint, lung, gut, brain and liver. Samples also included monocytes from peripheral and cord blood, as well as *in vitro* differentiated DCs from cord blood progenitors or monocyte-derived macrophages. Columns include dataset accession ID, platform, Stemformatics Dataset ID, number of samples, tier categorization, cell type and relevant tissue/organism part.

Additional Supplementary Material

Supplementary Video
Active engulfment and clearance of cells by pluripotent stem cell-derived macrophages.

Methods
Atlas formation was developed as described in (Angel et al., 2020) and is comprised of 44 datasets, 901 samples and 3757 genes. Mapping, and analysis of microarray and RNA sequencing datasets were conducted in the Stemformatics platform. Scripts are available for download from the Stemformatics BitBucket (Choi et al., 2019a)). All datasets and relevant samples (Supplementary Table 6) passed stringent quality control checks required for hosting on the stemformatics platform. Quality control checks include evaluation of library quality, and inclusion of replicates associated with experimental design. These datasets were either already hosted on stemformatics, or were downloaded from public depositories and processed through the stemformatics pipeline for inclusion.

Platform Effect Analysis and Gene Selection for PCA

This method assesses each gene independently for a dependence upon experimental platform. The initial step is to transform expression values from RNA Sequencing and Microarray into percentile values. The second step uses a univariate estimate of gene platform dependence and then selects genes with a small ratio of platform dependent variance to total variance. These genes are used to run the PCA (3757 genes passed this cut). The threshold for gene selection is empirically determined by using the Kruskal Wallis H Test to assess the difference in platform expression distribution for each principal component. A platform variance fraction of 0.2 is found to remove the platform effect for the first three principal component. For further details, please refer to (Angel et al., 2020).

Quantification and Statistical Analysis

P-values were re-calculated using the Mann-Whitney-Wilcoxon rank-sum test (two-sided). This was implemented via the python (version 3.7.5) SciPy package (version 1.3.1)(Virtanen et al., 2019). Multiple testing over the set of genes was accounted for with Bonferroni correction implemented in the statsmodels package (Seabold and Perktold, 2010)

Pseudo-bulk samples from Villani's Single Cell Data

Single cells were aggregated to form pseudo-bulk samples to mitigate library size differences between single cell and bulk data, and to project samples onto the atlas. Each group of (known) cell type in the single cell data was randomly sampled for *k* single cells with replacement. Aggregation consisted in summing up their expression profiles. k was determined by the number of sampling (i.e. how many pseud-bulk samples for each cell type) as half of the cell type's population size. The optimum aggregation size *k* was investigated by evaluating the similarity in distribution between the atlas' plasmacytoid dendritic cells and the aggregated pseudo-bulk DC6 samples using Kolmogorov–Smirnov (KS) statistic $\boldsymbol{D}$ averaged across all genes. KS statistic measured the difference between the empirical cumulative distribution functions of two groups of samples; the smaller the value, the closer the aggregated DC6 resembled the reference transcriptional profiles from the pDC in the bulk atlas. A minimum $\boldsymbol{D}$ value was obtained for *k = 8* across 30 iterations (Supplementary Figure 1). Similar results were obtained for other cell types. Thus, every pseudo-bulk samples were aggregated from 8 single cells.

Capybara Cell Score

Capybara (Kong et al., 2020) cell scores was used to measure cell identities continuum of the pseudo-bulk samples using iMac as the reference. Capybara cell scores were calculated by performing restricted linear regression of reference iMac samples on each of the pseudo-bulk samples' expression profiles. Denote $\boldsymbol{y_i}$ the expression profile of the $\boldsymbol{i^{th}}$ pseudo-bulk sample of length $\boldsymbol{G}$, where $\boldsymbol{G}$ represents the total number of genes in the data, and $\boldsymbol{X}$ a ($\boldsymbol{G}$X$\boldsymbol{T}$) the reference iMac matrix, where $\boldsymbol{T}$ represents the number of known cell types of interest. We considered 5 cell types: Dendritic cell, Monocyte, CD141+ dendritic cell, CD1c+ dendritic cell and Plasmacytoid dendritic cells. $\boldsymbol{X}$ is obtained by averaging the expression profiles of the iMac samples according to their cell types. Capybara solves the optimization problem

$$argmin_{\beta} \ (y_i - X\beta_i)^T (y_i - X\beta_i)$$

under the constraint that

$$\beta_{it} > 0 \quad \forall t \in \{1, 2, \ldots, T\}, \sum_{t}^{T} \beta_{it} < 1$$

where $\boldsymbol{\beta_{it}}$ is a regression coefficient, or cell score, for each pseudo-bulk sample i and each iMac cell type $\boldsymbol{t}$. The cell score is obtained using quadratic programming implemented with R (version 3.6.2) package *quadprog (version 1.5-8)*(Turlach and Weingessel, 2019).

Enrichment analysis and Protein-Protein Network

An enrichment analysis was conducted on the top 92 genes ranked by Pearson correlation (≥0.7) along the upward axis including invitro-derived cells. Enriched pathways were identified using these genes at Reactome (Fabregat et al., 2018) and significance ranked by p-value/false discovery rate. A protein-protein network was generated using the top 92 genes on STRING-DB (Szklarczyk et al., 2019). Disconnected nodes are not shown.

Pluripotent Stem Cell Differentiation

Stem cell work was carried out in accordance with The University of Melbourne ethics committee HREC (approval 1851831). Stem cell lines used were PB001.1 (Vlahos et al., 2019), a kind gift from the Stem Cell Core Facility at the Murdoch Children's Research Institute, and HDF51(Jones et al., 2013) was kindly provided to ALL by Prof. Jeanne Loring (The Scripps Research Institute, CA, USA). Human pluripotent stem cells were differentiated into macrophages based on protocol described by (Joshi et al., 2019; Ng et al., 2008), with modifications. Modifications were as follows: embryoid bodies were kept in rotational cultures without transference to matrigel plates for adherence, and the collection of progenitors from week 2 were immediately re-suspended in RPMI-1640 containing L-Glutamine (Life Technologies) and 10% Fetal Bovine Serum for macrophage differentiation (see macrophage differentiation).

Monocyte isolation from peripheral blood

Buffy Coat was obtained from the Australian Red Cross Blood Service in accordance with The University of Melbourne ethics committee HREC (approval 1646608). The blood was diluted with PBS at a 1:3 dilution and underlayed with Ficoll-Hypaque. The underlayed blood samples were centrifuged at 350g for 30 minutes at 24°C with no brake. Peripheral blood mononuclear cells were isolated from the interphase and washed twice by using MACs buffer (DPBS, 0.5% heat inactivated Fetal Bovine Serum, 2mM EDTA) and centrifuging at 400g for 5 minutes at 4°C. Cells were centrifuged at 400g for 5 minutes at 4°C and resuspended in 40µl MACs buffer per 107 cells. Monocytes were positively selected by a magnetic field using Human CD14 MicroBeads (MACS Miltenyi Biotec) and LS Columns (Miltenybiotec). These cells were plated for macrophage differentiation (see Macrophage differentiation

Macrophage differentiation

Monocytes/progenitor cells were cultured in tissue-culture treated 6 well plates. Cells were cultured in RPMI-1640 medium containing L-Glutamine (Life Technologies) with 10% Fetal Bovine Serum and 100ng/ml recombinant Human M-CSF (R&D Systems; 216-MC) for 5 days. Media changes were carried out on day 4.

Flow Cytometry

HMDM and PSCM were collected and centrifuged at 400g for 5 minutes. Supernatant was aspirated and 5µl mouse serum was added to 'dry' pellets for 5 minutes on ice. Cells were

resuspended in FACS Buffer (Hanks Balanced Salt Solution, 0.5% Human Serum Albumin) and stained with CD14 or matched isotype control antibodies on ice for 20 minutes then washed twice (3ml FACS Buffer, spun at 400g, 5 minutes). Resuspended cells were fixed with 4% paraformaldehyde (PFA) for 15 minutes at room temperature. PFA was washed out and cells rinsed twice in FACS buffer before resuspending in PBS and stored overnight at 4°C. Fixed cells were permeabilized with 0.1% Triton X-100 (in 1XPBS) for 10 minutes at room temperature and washed twice. Blocking buffer (0.3M glycine buffer, 10% Goat Serum, 1XPBS) was added to 'dry' pellets on ice for 1 hour. Cells were stained with antibodies to Type I Collagen or matched isotype control on ice for 20 minutes. Cells were washed twice then incubated with secondary antibody for 20 minutes on ice in the dark. Cells were then washed twice and resuspended for analysis on a CytoFLEX S flow cytometer (Beckman Coulter, Brea, CA) using CytExpert acquisition software. Post-acquisition analysis was performed with FCS Express 7 flow cytometry software.

Flow Cytometry antibodies	
CD14	Brilliant Violet 421™ anti-Human CD14 (BioLegend: Cat. No. 325628)
CD14 Isotype	Brilliant Violet 421™ Mouse IgG1 (BioLegend: Cat. No. 400157)
Type I Collagen Primary antibody	Rabbit Anti-Collagen I antibody (Abcam: Cat. No. ab264074)
Secondary Antibody	Goat Anti-Rabbit IgG H&L (Alexa Fluor®488) (Abcam: Cat. No. ab150077)
Type I Collagen Isotype	Rabbit IgG monoclonal Isotype Control (Abcam: Cat. No. 172730)

Stimulation Assay

On day 5 of differentiation, one well containing peripheral blood monocyte- or human pluripotent stem cell-derived macrophages were stimulated with 10ng/ml Lipopolysaccharide (LPS) (Sigma-Aldrich; *Salmonella enterica* serotype minnesota) for 2 hours. After stimulation period, media was aspirated, and the wells were washed twice with PBS (Ca2+Mg2+ free). before cell lysis using 2-mercaptoethanol (Sigma-Aldrich) and RNeasy Plus Lysis Buffer (Qiagen). Samples were placed into Eppendorf's and stored at -80°C before RNA extraction.

RNA extraction

Total RNA was isolated using the RNeasy® Plus Mini Kit (Qiagen) according to manufacturer's instructions. In summary: for the removal of genomic DNA, samples were placed into gDNA columns and centrifuged for 30 seconds at 8000g. Ethanol (70%) was mixed with the flow through and samples were transferred to RNeasy spin columns. The columns were centrifuged for 15 seconds at 8000g. Buffer RW1 was then added to the columns and columns were centrifuged for 15 seconds at 8000g. Buffer RPE was added to the columns and columns were centrifuged for 15 seconds. Buffer RPE was again added to the columns with centrifugation at 8000g for 2 minutes. Columns were then placed into new collection tubes and centrifuged at full speed for 1 minute to dry the membrane. RNase-free water was then added directly onto the column membrane and columns placed into Eppendorf's and centrifuged at 8000g for 1 minute to collect RNA. RNA quality and quantity were determined using a Tapestation (Agilent Technologies 2200). Samples were stored at -80°C. Zymo Research RNA Clean &

Concentrator-25 Kit was used to pool replicates (from the same donor) together and elute in elute into smaller volume with maximum concentration.

RNA sequencing

RNA samples were processed by the Ramaciotti Centre for Genomics (University of New South Wales; Sydney). Illumina Novaseq_6000 was used for mRNA-sequencing.

Data and Code Availability

mRNA-sequencing data is available through accession GSE150893. All public accessions are listed in Supplementary Table 6. Stemformatics code is publicly available at bitbucket.org/stemformatics. Atlas code is available at bitbucket.org/stemformatics/s4m_pyramid/src/master/scripts/atlas.py.

Graphing software and Illustration

Graphs for mRNA-seq gene expression were generated using Graphpad Prism. Violin plots were generated through the www.stemformatics.org platform. Schematic Figure illustration created with BioRender.com

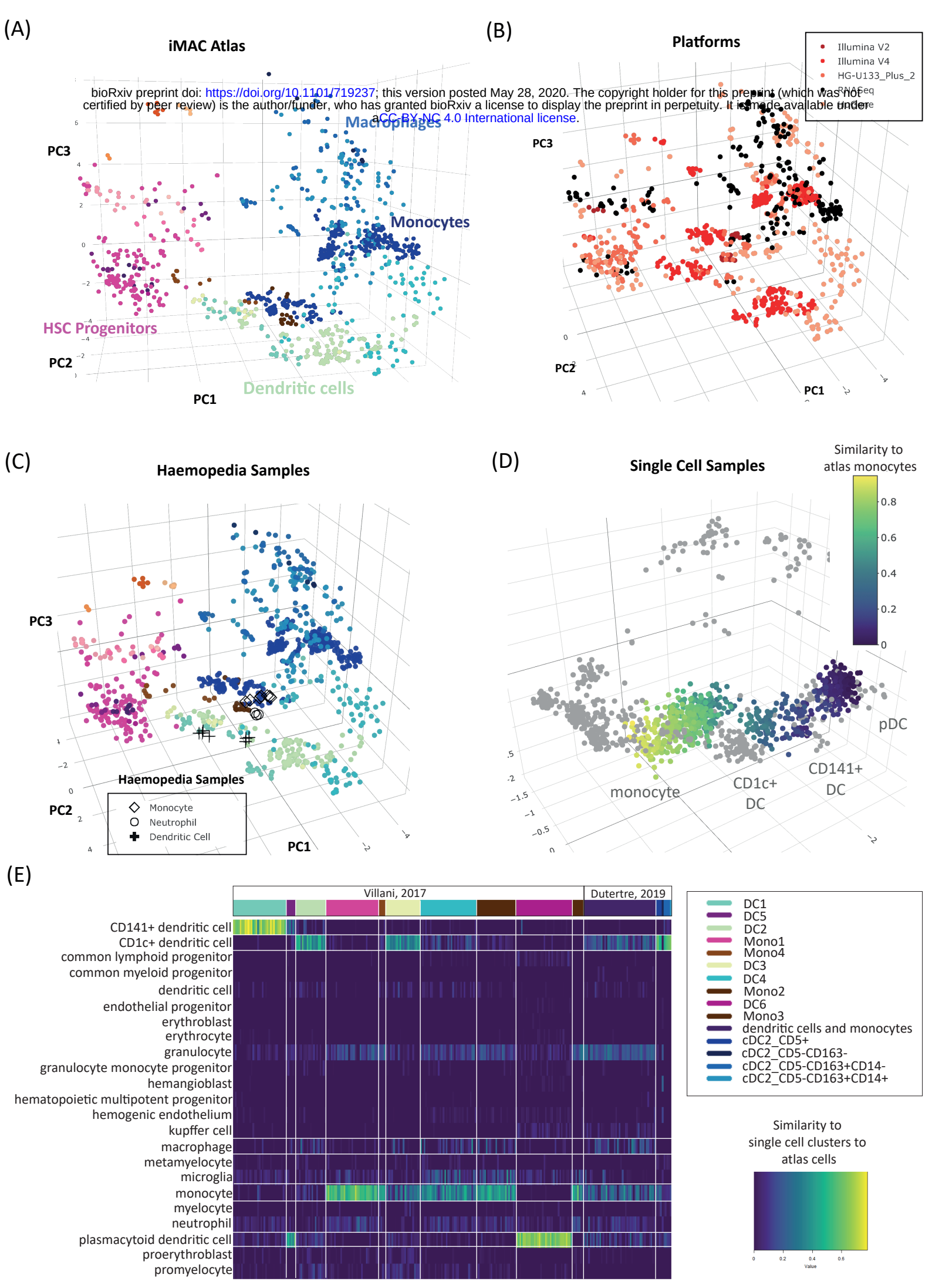

Figure 1: A reference atlas and annotation tool for human myeloid biology

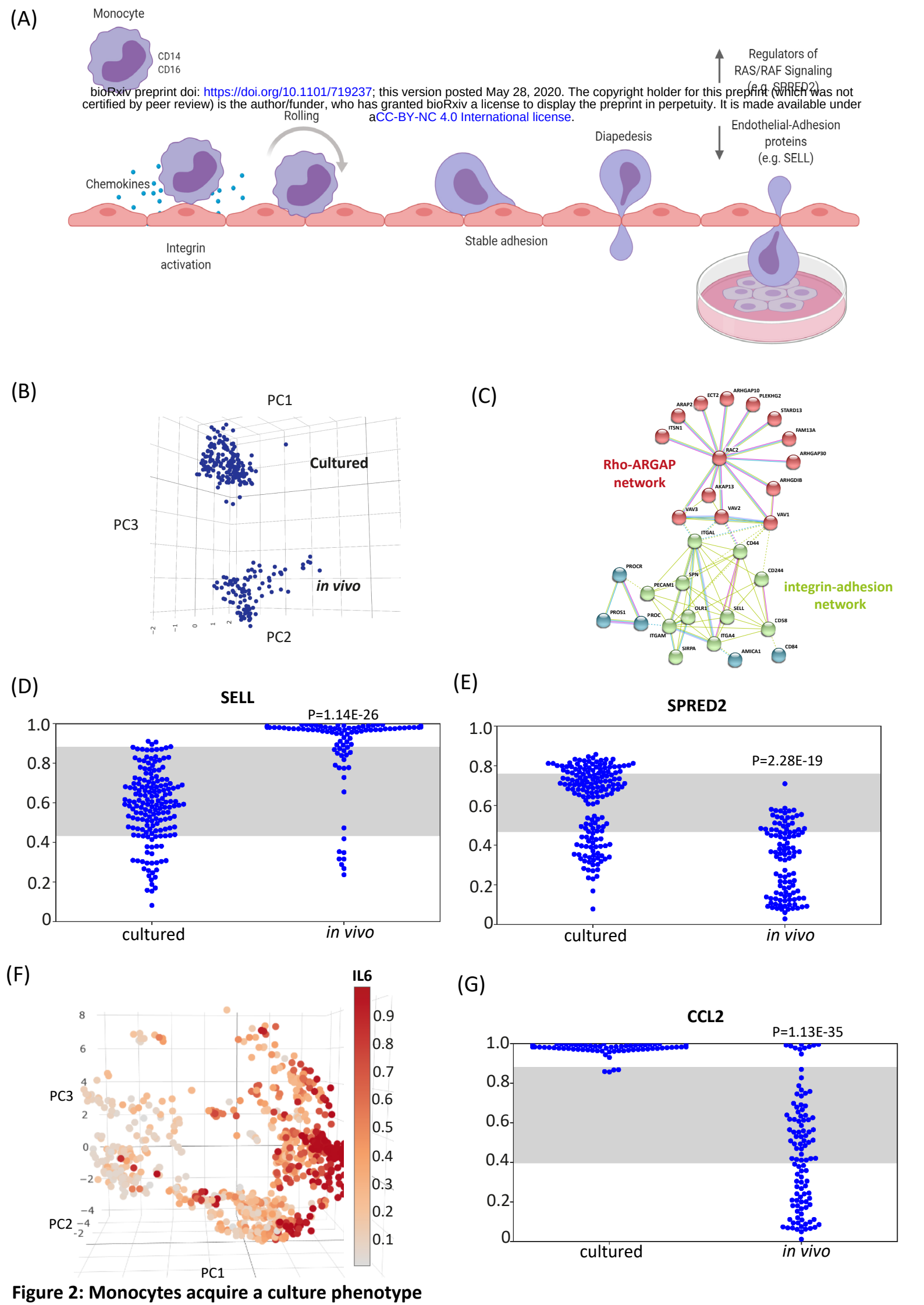

Figure 2: Monocytes acquire a culture phenotype

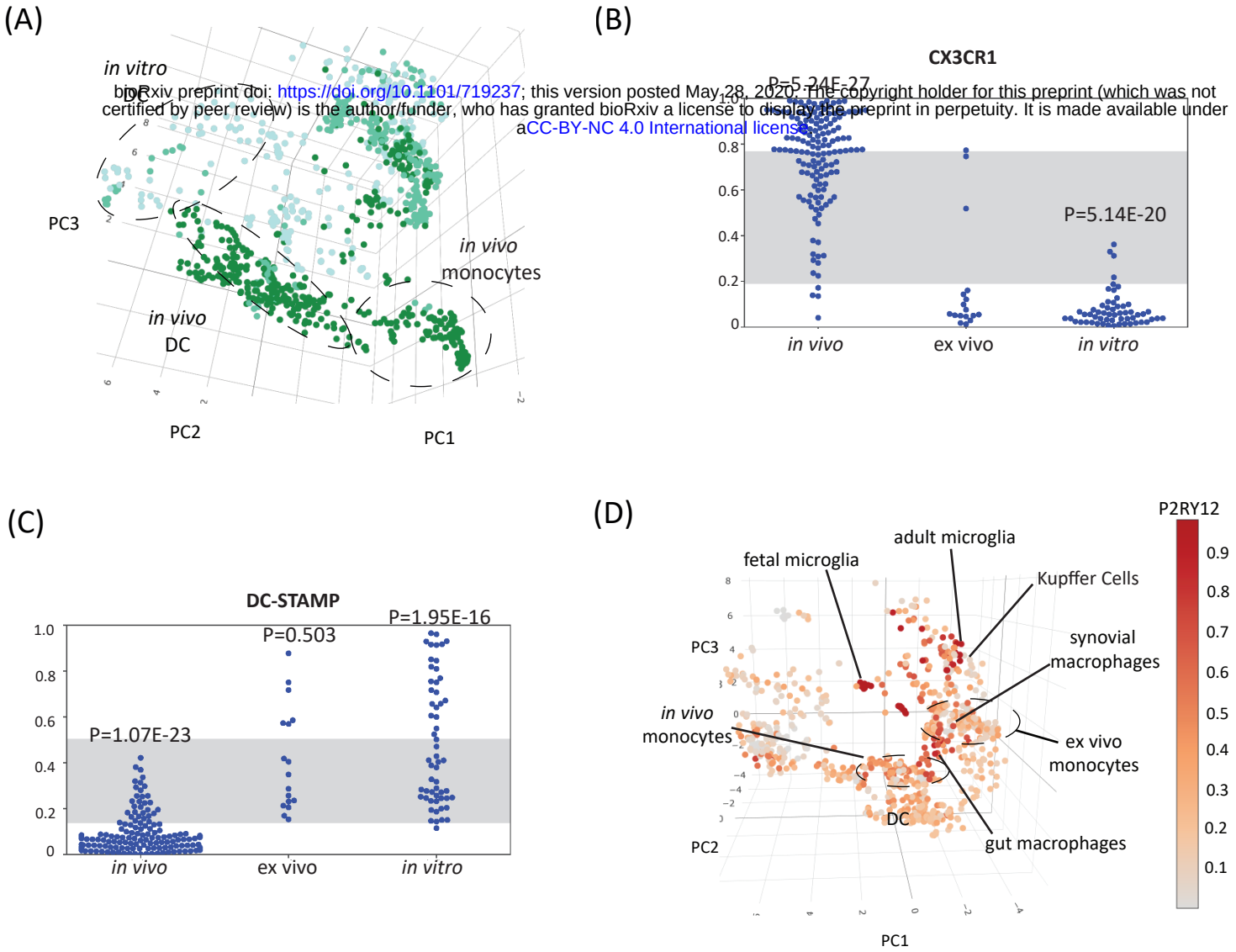

Figure 3: Cultured and *in vitro*-derived dendritic cells do not capture aspects of *in vivo* myeloid biology

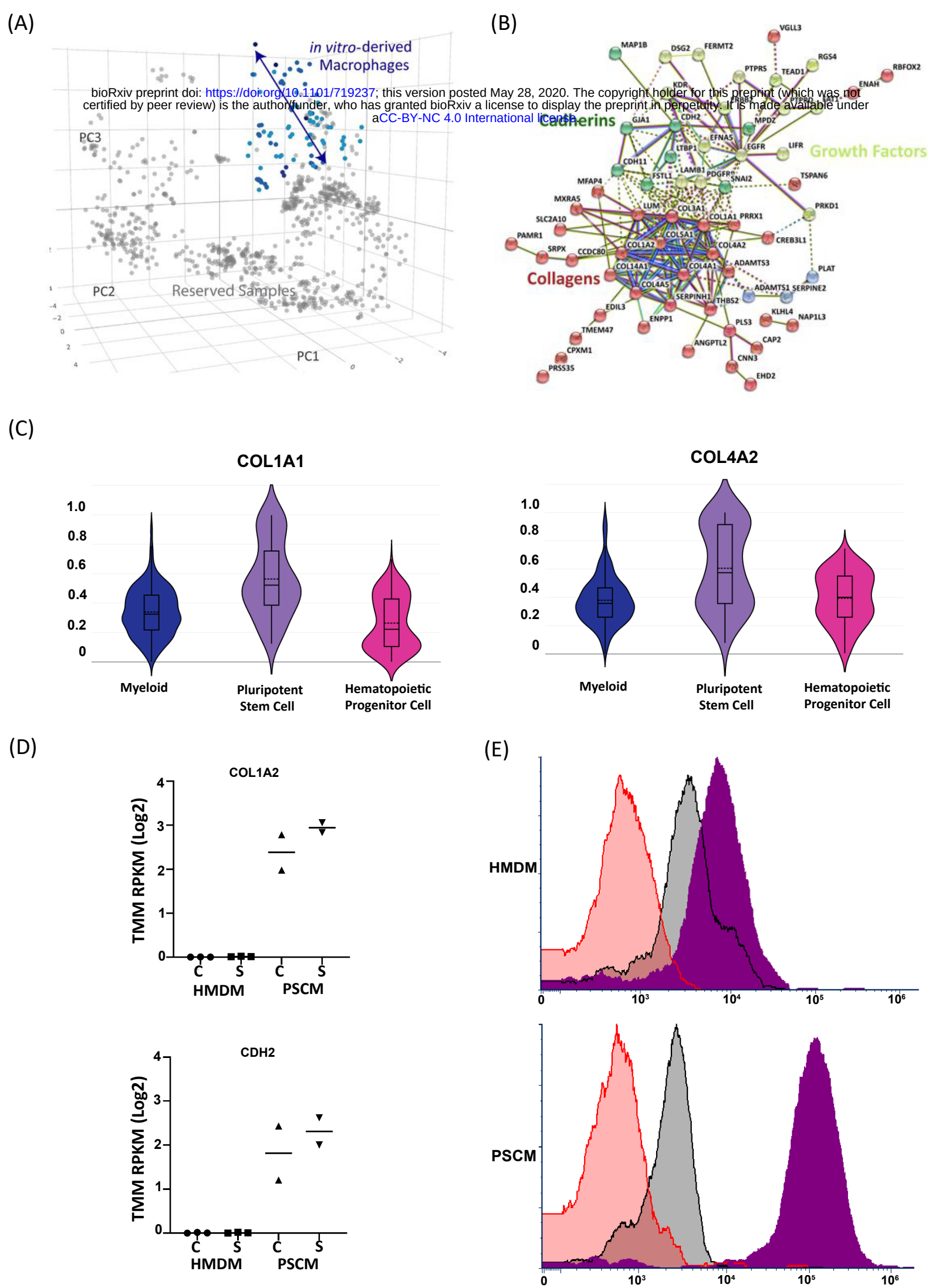

Figure 4: PSC-Macrophages display phenotype enriched for collagen production