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], Mariola Kurowska-Stolarska[3], Simon Milling[3], Chris M Pacheco[1], Matt Rutar[1], Jarny Choi[1] and Christine A Wells*[1,4].

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. The Institute of Infection, Immunity and Inflammation, Rheumatoid Arthritis Pathogenesis Centre of Excellence (RACE), The University of Glasgow.
4. The Walter and Eliza Hall Institute for Medical Research. 1G, Royal Parade, Parkville VIC 3052

*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:

Macrophages isolated from different tissues have distinct niche roles that maintain tissue homeostasis, as well as providing surveillance for tissue injury or infection. By assembling a large transcriptional atlas of human myeloid biology representing ~1000 samples, we address the question of how well laboratory models, including pluripotent-derived cells, represent macrophage biology, and particularly whether these can model resident tissue macrophage specialisation.

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: transcriptional 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 do 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), so 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, and therefore may not be appropriate as a benchmark for specialized tissue functions. A reference atlas of human myeloid biology that draws on current benchmarks, and that can be added to by the research community would therefore be of enormous value.

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. To address these questions, we undertook an integrated myeloid atlas to identify and benchmark human myeloid subpopulations from ex vivo,

in vivo and *in vitro* sources. The atlas is made available as an interactive online resource in the www.stemformatics.org platform.

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 >900 samples representing peripheral blood monocytes, tissue-resident, ex vivo and *in vitro*-derived macrophages and dendritic cells. Samples included were selected from datasets processed through the Stemformatics pipeline which had passed the stringent set of quality control requirements for hosting on the Stemformatics.org portal (Choi et al., 2019a). We constructed a PCA by implementing a two-step process. Firstly, transformation of expression values from the original studies to percentile values to facilitate the comparison of microarray and sequencing experiments. Secondly, using a univariate estimation of their platform dependence, genes whose expression values had low contribution from the way they were measured were included. This approach led to reproducible clustering of distinct myeloid classes, regardless of the platform that they were measured on (Figure 1B and Supplementary Figure 1). Variables such as progenitor source (HSC, monocyte-derived, or iPSC-derived), culture status or activating condition contributed to clustering of samples. Myeloid subsets including dendritic cells, monocytes and neutrophils were validated by projecting RNAseq data independently-derived from Haemopedia (Choi et al., 2019b) onto the atlas (Figure 1C).

Monocytes acquire a 'Culture-Phenotype'

Monocytes are post-mitotic blood cells derived from bone marrow progenitors that are short lived in circulation and can give rise to some tissue macrophages. The largest population of circulating monocytes is marked by high levels of expression of the LPS co-receptor CD14. Intermediate and nonclassical subsets are marked by acquisition of the type III FcR□, CD16 (Schmidl et al., 2014). Monocytes begin to roll along blood vessels in response to chemokines such as CCL2, before diapedesis into tissues. Cultured monocytes have been previously described as 'activated', but while we observe a distinct culture phenotype, the transcriptome of cultured cells mimics many of the features (Figure 2A) 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 selectin SELL (Figure 2C). Regulators of RAS/RAF signaling including SPRED2 (Wakioka et al., 2001) have an elevated expression in cultured monocytes (Figure 2D), consistent with spreading and migration across tissue culture plastics.

The culture phenotype acquired by monocytes appears to be a prelude to activation, which can be observed along an adjacent axis in Figure 2E and is exemplified by the expression of IL6. Cultured monocytes express higher levels of CCL20 and CCL2 than peripheral monocytes (Figure 2F). 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).

Furthermore, pathogen-activated phenotypes of cultured monocytes or monocyte-derived macrophages are typified by high expression of cytokines.

Primary tissue-resident cells

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 3A), including down-regulation of receptors such as CX3CR1, IL18R1 and TLR7 (Figure 3B and Supplementary Table 4). Other molecules, such as the cel-fusion protein DC-STAMP are gained in culture (Figure 3C and Supplementary Table 4). So, while providing useful models of myeloid biology, cultured or progenitor-derived dendritic cells don't adequately capture key aspects of *in vivo* myeloid biology.

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). The classical laboratory model of human macrophage biology 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.

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. 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. We also observed some tissue partitioning of macrophages, with alveolar macrophages sitting closely with cultured monocytes. Colon-derived macrophages grouped together with ascites-infiltrate, as well as tumour-associated macrophages (TAM) (Figure 3D and Supplementary Figure 2A) which sat between cultured monocytes and CD1c+ dendritic cells.

PSC-derived specialised macrophages

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, 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 and microglia display transcriptional hallmarks of efferocytosis and collagen production

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 recent study by (Cao et al., 2019) demonstrated higher lipid uptake in PSC-macrophages compared to peripheral blood monocyte-derived macrophages, 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.

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, 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. It is somewhat surprising to see high levels of extracellular matrix expressed by macrophages. This either suggests that *in vitro*-derived cells are incompletely patterned to macrophages, adopting a hybrid cell type that shares features with other fibroblastic cell types, or properly patterned but can be transformed into collagen-producing cells in culture. Although ex vivo cells were not included in this analysis, culture impact on collagen production can be observed in ex vivo microglia (Figure 4C).

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 which 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 myeloid progenitors, including common myeloid progenitors and hemogenic endothelium, and is retained at high levels in some PSC-derived microglia (Figure 4D).

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 3D).

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).

Concluding Remarks
PSC-Macrophages provides exciting opportunities for modelling tissue residency and specialization. Understanding the key drivers underpinning phenotypic differences between macrophage subsets will help the stem cell field to develop specialized populations such as tissue resident populations from brain, lung or liver.

By benchmarking 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, so perhaps the ontogeny question is less interesting than understanding the molecular networks that can be co-opted to deliver specific phenotypic properties outside of the constraints of development.

Acknowledgements
The authors thank Tyrone Chen for assistance with data processing and Isha Nagpal for website development to support the iMAC interactive viewer. 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

NR (review of literature, annotation of samples, network analysis, critical evaluation of analysis, writing, image design, editing), PWA (Method development, analysis), MKS (Sample provider), SM(Sample provider), CMP (Sample annotation, project discussions), MR (Sample annotation, project discussions), JC (implementation for reviewers/readers), CAW (Conception, annotation of samples, network analysis, critical evaluation of analysis, writing, editing).

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 .

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.

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 *356*, 1248–1259.

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.

Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics *17*, 483.

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). The ground state of innate immune responsiveness is determined at the interface of genetic, epigenetic, and environmental influences. J. Immunol. *193*, 13–19.

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.

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.

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* pii: e27041.

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 *362*,

eaar2971.

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.

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.

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.

Yamaguchi, H., Maruyama, T., Urade, Y., and Nagata, S. (2014). Immunosuppression via adenosine receptor activation by adenosine monophosphate released from apoptotic cells. Elife *3* e02172.

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.

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.

Figure 2: Monocytes acquire a culture phenotype
(A) STRING-DB network of top-ranked genes differentially expressed between peripheral blood (*in vivo*) and cultured monocytes indicating upregulation of cytoskeletal proteins and down regulation of endothelial-adhesion proteins (B) Cultured monocytes form a distinct cluster away from *in vivo* monocytes along PC3 (C and D) Ranked expression (Y-axis) of genes 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 (C) L-Selectin (D) Regulator of RAS/RAF signaling (E) 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) (F) Ranked expression (Y-axis) of Chemokines CCL2 and CCL20

genes 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 (A) receptor CX3CR1 and (B) 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) Ranked expression (Y-axis) of Collagen genes *in vivo* (n=10), *ex vivo* (n=21) and *in vitro*-derived microglia (n=43). Grey stripe indicates variance attributable to platform. P-value: Mann-Whitney-Wilcoxon rank sum test (D) iMAC atlas coloured by ranked expression of MYB (Scale bar: high ranked expression (dark red) to low ranked expression (grey)).

<u>Tables with Titles and Legends</u>

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

Methods

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.

Method Details

Platform Effect Analysis and Gene Selection for PCA

This method assesses each gene independently for a dependence upon experimental platform. Then genes with low platform dependence compared to biological variation and selected and used to run the PCA.

The problem is similar to the standard batch correction problem, but with an important difference - in this case the effect of batch (experimental platform) is present across multiple datasets, and we can assume it is consistent and measurable across these datasets. This approach is made possible by the large collection of datasets in Stemformatics, which collectively sample the different platforms and cell types. Note, only genes measurable in *all* the above platforms are used in the analysis, which is approximately 13000 genes.

The initial step is to transform expression values from RNA Sequencing and Microarray onto the same scale. Gene expression for each sample is transformed into percentile values - the highest expression gene is given the value 1.0, while the lowest receives the value 0. Values in between are uniformly spaced according to the rank of the gene's expression. Ties receive the same values.

The second step uses a univariate estimate of gene platform dependence and then selects genes with very little platform dependence. We achieve this by fitting univariate linear model on each gene, with independent variable as platform,

$$y = X_p B_p + \epsilon \qquad \text{(eq 1)}$$

$$\epsilon \sim N(0, \sigma_\epsilon) \qquad \text{(eq 2)}$$

where y is the expression of a single gene across all samples, X_p indicates membership of the platform, with coefficient β_p. The variance due to platform is σ_p and is calculated as

$$\sigma_p = var(X_p B_p) \qquad \text{(eq 3)}$$

The total variance, σ^2_{Total} is

$$\sigma^2_{Total} = \sigma^2_p + \sigma^2_\epsilon \qquad \text{(eq 4)}$$

and fraction of variance due to platform is

$$f = \sigma^2_p / \sigma^2_{Total} \qquad \text{(eq 5)}$$

In practice we implement this using the variance partitioning package (Hoffman and Schadt, 2016) with a single fixed effect, platform. This model is a fixed effect analysis of variation (ANOVA). We then filter genes that have a significant percentage of the variance due to platform.

Genes with a f ≤ 0.2 are selected and used to run the PCA (approximately 3600 genes pass this cut). This threshold value is selected by using the Kruskal Wallis H Test to assess the difference in platform medians. The value of 0.2 is found empirically to remove the platform effect from the first three principal components

Quantification and Statistical Analysis

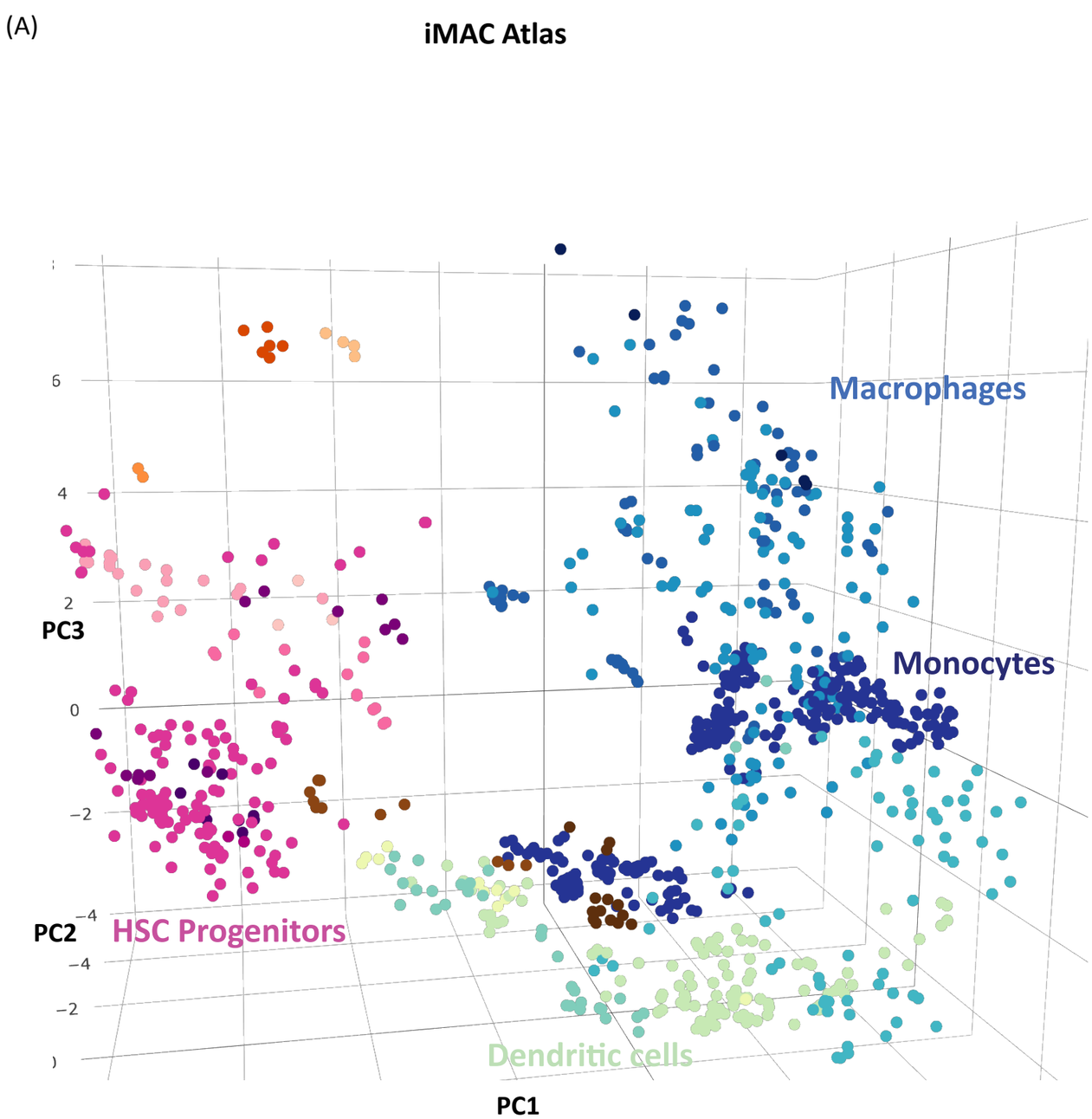
P-values were re-calculated using the Mann-Whitney-Wilcoxon rank-sum test (two-sided). This is implemented via the python scipy package. Multiple testing over the set of genes is accounted for with Bonferroni correction implemented in the statsmodels package (Seabold and Perktold, 2010)

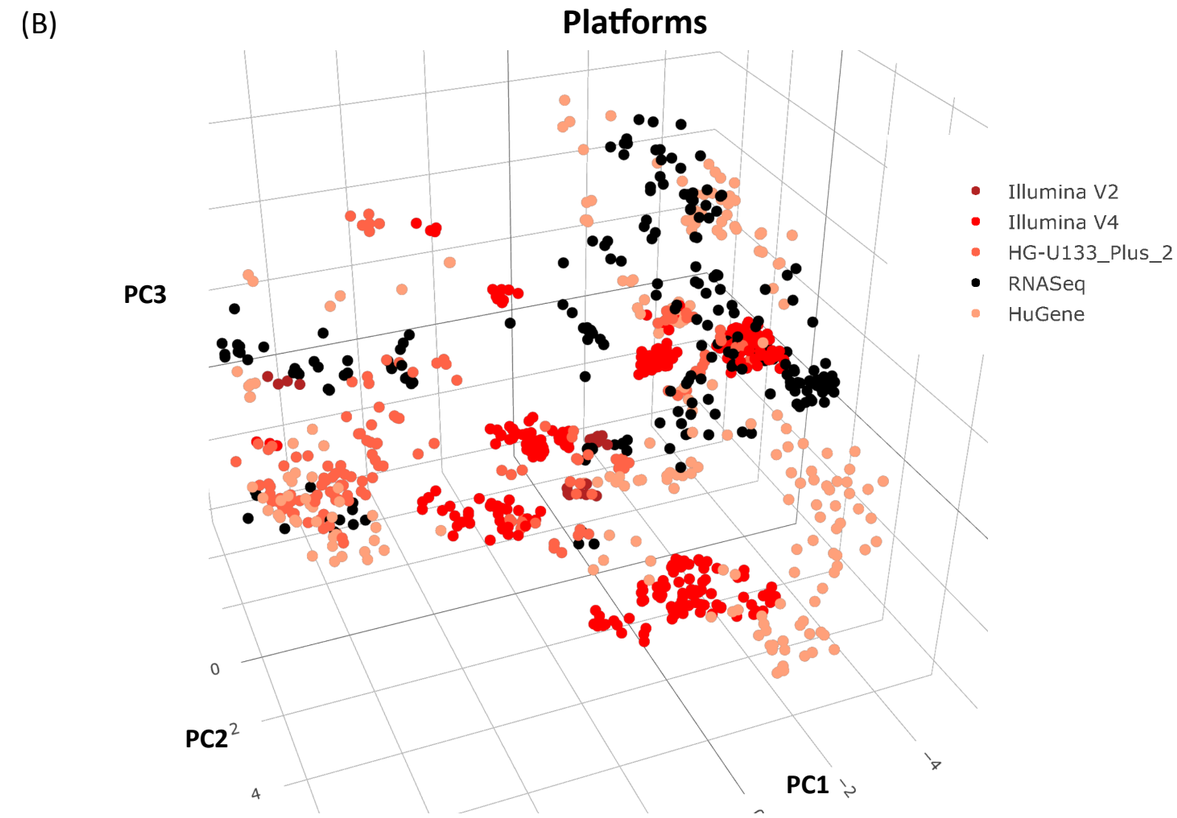
Enrichment analysis and Protein-Protein Network

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.

Data and Code Availability

All public accessions are listed in Supplementary Table 6. All code is available from the Stemformatics bitbucket.





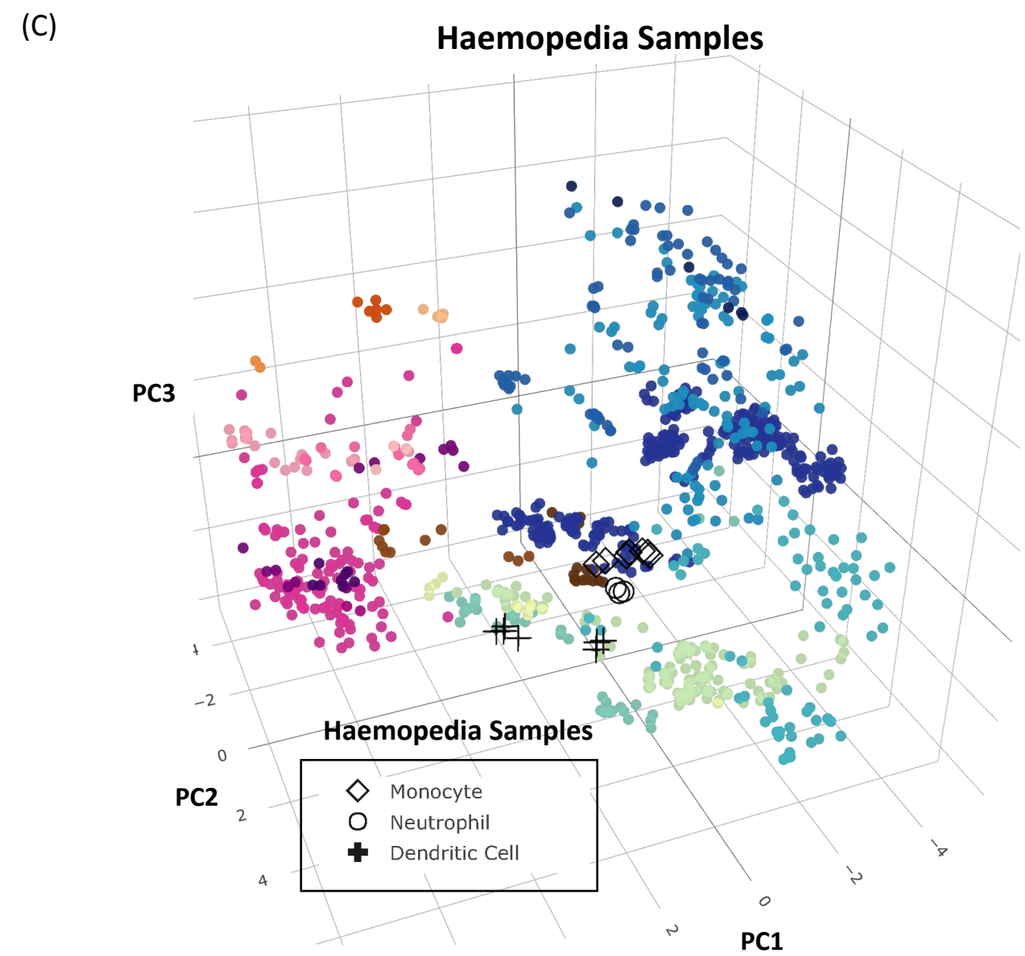


Figure 1: A reference atlas for human myeloid biology

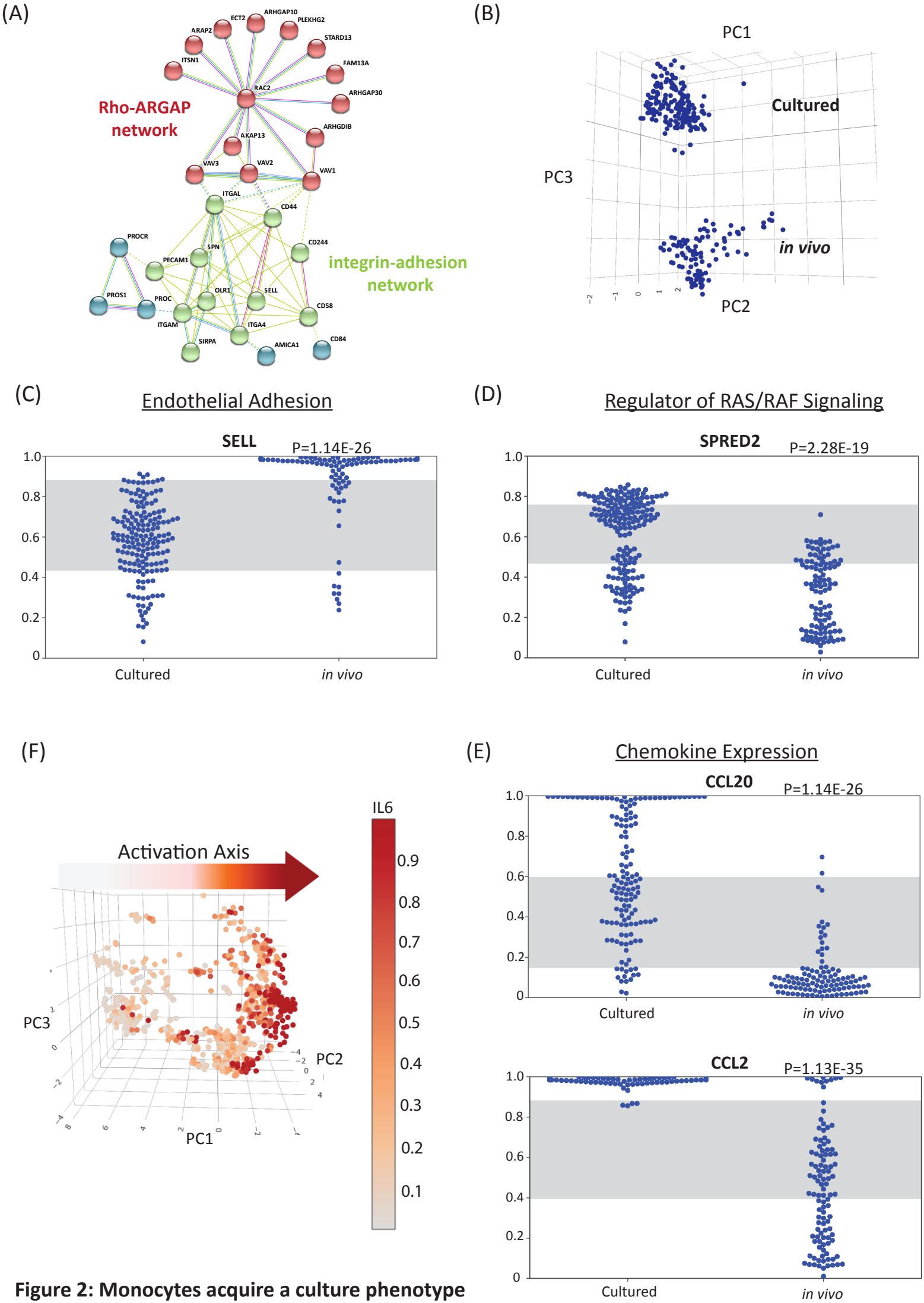


Figure 2: Monocytes acquire a culture phenotype

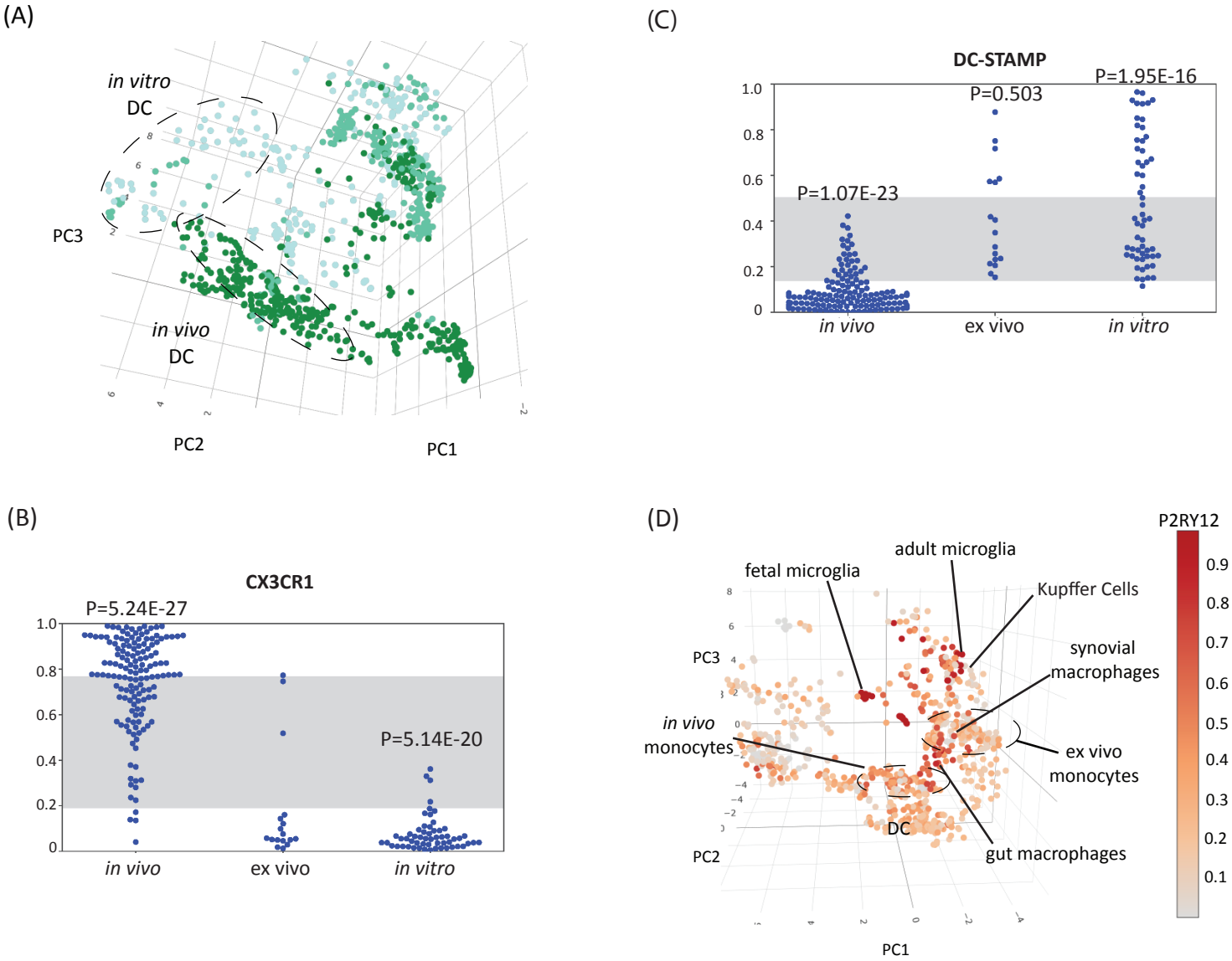


Figure 3: Cultured and *in vitro*-derived dendritic cells do not capture aspects of *in vivo* myeloid biology

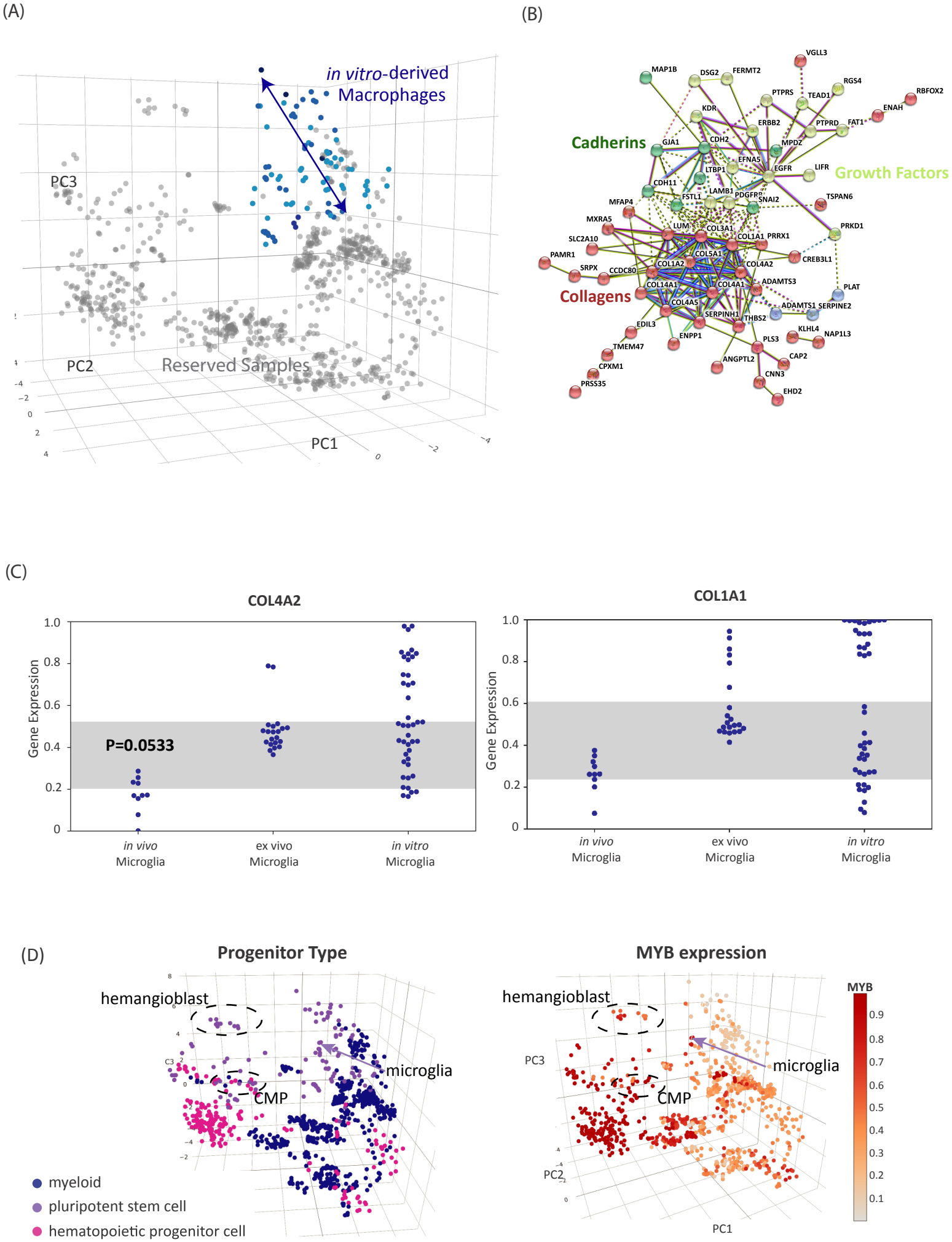


Figure 4: *in vitro*-derived macrophages do not capture developmental ontogeny