

Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S, Jardine L, Dixon D, Stephenson E, Nilsson E, Grundberg I, McDonald D, Filby A, Li W, De Jager PL, Rozenblatt-Rosen O, Lane AA, Haniffa M, Regev A, Hacohen N. [Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors](#). *Science* 2017, 356(6335).

Copyright:

This is the author's version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in Science Journal Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 2017, 356(6335), doi: 10.1126/science.aah4573

DOI link to article:

[10.1126/science.aah4573](https://doi.org/10.1126/science.aah4573)

Date deposited:

07/07/2017

Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes and progenitors

Alexandra-Chloé Villani^{1,2,*‡}, Rahul Satija^{1,3,-4*}, Gary Reynolds⁵, Siranush Sarkizova¹, Karthik Shekhar¹, James Fletcher⁵, Morgane Griesbeck⁶, Andrew Butler³⁻⁴, Shiwei Zheng³⁻⁴, Suzan Lazo⁷, Laura Jardine⁵, David Dixon⁵, Emily Stephenson⁵, Emil Nilsson⁸, Ida Grundberg⁸, David McDonald⁵, Andrew Filby⁵, Weibo Li^{1,2}, Philip L. De Jager^{9,1}, Orit Rozenblatt-Rosen¹, Andrew A. Lane^{1,7}, Muzlifah Haniffa^{5,10,‡}, Aviv Regev^{1,11,-12‡}, Nir Hacohen^{1,2,‡}

1. Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
2. Center for Cancer Research, Massachusetts General Hospital, Department of Medicine, Boston, Massachusetts, USA.
3. New York Genome Center, New York University Center for Genomics and Systems Biology, New York, New York, USA.
4. New York University, Center for Genomics and Systems Biology, New York, New York USA.
5. Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK.
6. Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA.
7. Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA.
8. Olink Proteomics, Watertown, Massachusetts, USA.
9. Program in Translational NeuroPsychiatric Genomics, Department of Neurology, Brigham and Women's Hospital and Harvard Medical School.

10. Department of Dermatology, Royal Victoria Infirmary, Newcastle Hospitals NHS Foundation Trust, UK.
11. Department of Biology and Koch Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
12. Howard Hughes Medical Institute, Chevy Chase, Maryland, USA.

* These authors contributed equally to this work.

‡Correspondence to: cvillani@broadinstitute.org (A.C.V.), m.a.haniffa@newcastle.ac.uk (M.H.), aregev@broadinstitute.org (A.R.), and NHACOHEN@mgh.harvard.edu (N.H.).

One Sentence Summary: Single cell RNA-sequencing and functional studies were used to revise the definitions of human blood dendritic cells and monocytes.

Abstract

Dendritic cells (DC) and monocytes play a central role in pathogen sensing, phagocytosis and antigen presentation and consist of multiple specialized subtypes. However, their identities and inter-relationships are not fully understood. Using unbiased single-cell RNA sequencing of ~2400 cells, we identify 6 human DC and 4 monocyte subtypes in human blood. Our study reveals: a new DC subset that shares properties with plasmacytoid DCs (pDC) but potently activates T cells, thus re-defining pDCs; a new subdivision within the CD1C⁺ subset of DCs; the relationship between blastic plasmacytoid DC neoplasia cells and healthy DCs; and circulating progenitors of conventional DCs (cDC). Our revised taxonomy will enable more accurate functional and developmental analyses as well as immune monitoring in health and disease.

Introduction

Dendritic cells (DCs) are mononuclear phagocytes found in blood, lymphoid organs and all tissues. One of their central functions is to ingest materials such as pathogens, present processed epitopes to T cells and regulate innate and adaptive immune responses (1-3). DCs are heterogeneous and consist of multiple subtypes with unique functions that have been defined over the past decade in mice and humans. However, it is unclear how many DC subtypes exist, how they are related to each other, and how they differ from other mononuclear phagocytes.

The results of numerous studies have shown that human dendritic cells express high levels of major histocompatibility complex class II (HLA-DR), a molecule essential for antigen presentation, and lack key markers of T, B, NK, granulocyte and monocytes. In the blood, DC subtypes include CD11C⁺ conventional DCs (cDCs), consisting of either CD141⁺ or CD1C⁺ cells, and plasmacytoid DCs (pDC), consisting of CD123⁺ cells. cDCs are effective at antigen-specific stimulation of CD4⁺ and CD8⁺ T cells, while pDCs specialize in producing type I interferons in response to viruses. pDCs and cDC subtypes differ in their expression of numerous sensors, pathways and effectors, and play distinct roles in the immune response (1-3).

The different DC subtypes have historically been defined by a combination of morphology, physical properties, localization, molecular markers, functions and developmental origins, converging to the current model described above (1-3). However, the definition of DCs is still likely to be biased by the limited markers available to identify, isolate and manipulate the cells. Such biases, in turn, would alter the assignment of function and ontogeny to each DC subtype.

To overcome some of these limitations, we used single-cell RNA sequencing (scRNA-seq) (4-5) to better assess the diversity of blood DCs and monocytes, leading us to identify new subtypes of DCs and monocytes, refine their existing classification, and pinpoint a precursor of cDCs in the blood. Using discriminative markers associated with the newly defined DC subtypes, we also assessed the functions of some of the DC subtypes. Overall, our analysis provides a relatively unbiased and comprehensive map of human blood DCs and monocytes.

RESULTS

Strategy for discovery and validation of DC and monocyte subtypes

To determine the subtypes of DCs and monocytes in the human blood, we developed an experimental and computational strategy to: (i) perform single-cell RNA-sequencing on DCs and monocytes derived from a single healthy individual; (ii) identify clusters of cells that are similar to each other; (iii) find discriminative markers per cluster; (iv) prospectively isolate cells corresponding to key clusters using newly identified surface markers; (v) validate the identity of the sorted cells using scRNA-seq; (vi) confirm the existence of these cell types in up to 10 independent healthy individuals; (v) perform functional analyses for selected cell types.

Single cell profiling of blood DCs and monocytes

We analyzed blood DC and monocyte populations from ficoll-purified blood cells that were FACS-sorted (**Fig. 1A**) and excluded for B, T and NK markers (6). For DCs, we sampled LIN⁻ HLA-DR⁺CD14⁻ cells across the CD11C⁺ (to enrich for CD141⁺ and CD1C⁺ cDCs) and CD11C⁻ (to enrich for CD123⁺ pDCs) fractions (**Fig. 1B**). For monocytes, we sampled LIN⁻CD14^{lo/++} cells (including classical CD14⁺⁺CD16⁻, intermediate CD14⁺⁺CD16⁺, and non-classical CD14⁺CD16⁺⁺). We used additional markers (DCs: CD123, CD141, CD1C; monocytes: CD14,

CD16) to create overlapping gates that comprehensively and evenly sample DCs and monocytes (6).

To define subpopulations and identify useful markers for further isolation, we performed deep scRNA-seq using a modified Smart-Seq2 protocol (6), followed by sequencing of ~1-2 million paired-end reads per cell (7-8). Of 768 DCs and 372 monocytes initially profiled in the selected individual for discovering subsets, we focused on 742 DCs and 339 monocytes that passed quality control (QC) filters (6) with an average of 5,326 unique genes detected per cell. In subsequent validation and characterization phases, we additionally profiled ~1200 cells.

Unbiased classification of LIN⁻HLA-DR⁺CD14⁻ subsets

We defined six cell clusters within the LIN⁻HLA-DR⁺CD14⁻ population using unsupervised analysis that did not rely on any marker gene expression. Briefly, we identified 595 genes exhibiting high variability across single cells, reduced the dimensionality of this data with principal components analysis (PCA), and identified five significant PCs using a previously described permutation test (6,9). We used these PC loadings as input to t-distributed stochastic neighbor embedding (t-SNE) (10) for visualization, and clustered cells using a graph-based approach similar to one recently developed for CyTOF data (6, 11). We observed 6 clusters: (**a**) two clusters mapping closely to the well-established DC subsets, with cluster DC1 mapping to CD141⁺ DCs, and cluster DC6 to pDCs, based on the *post-hoc* overlap of transcript and surface marker expression; (**b**) two clusters containing the CD1C⁺ cDCs, thus splitting them into 2 subsets: CD1C_A (cluster DC2) and CD1C_B (cluster DC3); (**c**) a cluster corresponding to the poorly characterized CD141⁻CD1C⁻ population (cluster DC4); (**d**) and one cluster that does not

correspond to any of the known blood DC subtypes (cluster DC5; **Fig. 1C**; **Fig. S1**).

We identified 242 genes ($AUC \geq 0.85$) that best classified cells into these 6 putative cell populations (**Fig. 1D**, **Fig. S2A**; **Tables S1-2** for a list of markers, including surface markers). While cluster DC1 mapped most closely to $CD141^+$ DCs, this commonly used $CD141$ (*THBD*/BDCA-3) marker was a poor discriminator for this cluster, being also expressed by cells captured in clusters DC5 and DC6 (pDC) (**Fig. S2B**). As *CLEC9A* appeared to be a perfect discriminative surface marker for the DC1 cluster, we refer to this subset henceforth as $CLEC9A^+$ DCs. Clusters DC2 and DC3 mapped to $CD1C^+$ DCs. $CD1C$ was the best and sole marker uniquely shared by both clusters. The DC4 cluster mapped to the $CD141^-CD1C^-$ population and was accurately delineated by *FCGR3A*/ $CD16$. The DC5 cluster was best defined by the surface markers *AXL* and *SIGLEC6*. Finally, while the DC6 cluster mapped to pDCs, several markers commonly used to identify these cells (e.g. *IL3RA*/ $CD123$, *CLEC4C*/ $CD303$) were also expressed in the population defined by the DC5 cluster, leading us to define a new combination of markers that distinguish pDCs from the DC5 population. Altogether, we identified sets of discriminative markers that can be used in combination to isolate cell populations corresponding to known DC subsets (but with higher purity) as well as to previously uncharacterized subsets.

Two subpopulations within $CD1C^+$ DCs

The $CD1C^+$ DCs were distributed across two clusters with similar numbers of cells, which we termed $CD1C_A$ (cluster DC2) and $CD1C_B$ (cluster DC3). Comparing the two clusters, the $CD1C_B$ cells were distinguished through their expression of a strong unique signature that

includes acute and chronic inflammatory genes (*12-14*), such as *CD14*, *S100A9* and *S100A8*, while CD1C_A cells were marked only by slightly higher levels of MHC class II genes (**Fig. 2A; Table S3**).

We validated the presence of the two populations by combining prospective isolation with a new sorting panel followed by scRNA-seq. To isolate these cells by flow sorting, we developed a panel incorporating surface markers derived from the set of uniquely expressed genes: *FCGR2B* (CD32B) for CD1C_A, and *CD163* and *CD36* for CD1C_B subsets (**Fig. 2B**). scRNA-seq of prospectively isolated cells from each subset recapitulated the original split observed in CD1C⁺ DCs (**Fig. 2C**). Unlike monocytes and pDCs, both CD1C_A and CD1C_B subsets (isolated with the newly identified markers) were potent stimulators of naïve T cell proliferation ($p < 0.05$, paired t-test), consistent with the functional characteristics of cDCs (**Fig. 2D**). Subsequent activation of both CD1C subsets with LPS, R848 and poly (I:C) highlighted functional differences between these subsets (**Fig. S3; Table S4**), with CD1C_A secreting higher levels of the immune mediators CCL19, IL-10, IL-12B, IL-18. Thus, scRNA-seq revealed unappreciated heterogeneity in this particular subset, leading to new hypotheses about the functions of CD1C⁺ DCs.

Discovering monocyte subsets and their relationships to DC subsets

Some key genes that were known to be associated with monocytes were also expressed by CD1C_B (cluster DC3) and CD141⁺CD1C⁺ (cluster DC4) cells (*e.g.*, *CD14* and *FCGR3A/CD16*, respectively). To analyze the relationships between monocytes and CD1C_B cells, we profiled 372 single blood monocytes (**Fig. 1A, Fig. 3A**). Based on 339 monocytes that passed QC, we

identified four clusters (**Fig. 3B, Fig. S4A**) distinguished by 102 classifier genes ($AUC \geq 0.85$; **Fig. 3C, Fig. S4B, Table S5**). The two largest clusters, Mono1 and Mono2, contained the $CD14^{++}CD16^{-}$ ('classical') and $CD14^{+}CD16^{++}$ ('non-classical'), respectively. However, Mono1 and Mono2 also included 88 of the 124 cells derived from the 'intermediate' monocyte gate ($CD14^{++}CD16^{+}$) (**Fig. S4A**), demonstrating that the 'intermediate' monocytes do not form a homogenous population. The two smaller clusters, Mono3 and Mono4, contained 40 of the 124 'intermediate' cells and expressed many of the Mono1 ('classical' monocyte) signature genes. Mono3 also uniquely expressed genes involved in cell cycle arrest and cell differentiation inhibition (*e.g.*, *G0S2*, *MXD1*), as well as trafficking (*e.g.*, *CXCR1*, *CXCR2*, *VNN2*), while Mono4 distinctively expressed a cytotoxic gene signature (*e.g.* *PRF1*, *GNLY*, *CTSW*) resembling previously reported 'natural killer dendritic cells' (15-17) (**Fig. 3C, Fig. S4B**). We conclude that the previously defined 'classical' and 'non-classical' subtypes are contained in 2 distinct clusters (Mono1 and Mono2, respectively), but that the 'intermediate' monocytes are far more heterogeneous than previously appreciated, being distributed across 2 known and 2 new clusters (**Fig. S4A**).

All monocyte subtypes shared a signature that distinguished monocytes collectively from $CD1C^{+}$ DC (cluster DC2 and DC3), $CLEC9A^{+}$ DC (cluster DC1), and pDC (cluster DC6) populations (*e.g.*, *ITGAM/CD11B*, *ITGB2*, *TLR2*, and *CLEC7A*) (**Fig. 3B, Fig. 3C, Fig. S4B**). Importantly, despite co-expressing genes such as *CD14* and *S100A8*, Mono1 and $CD1C_B/DC3$ cells did not form a single cluster (**Fig. 3B; Fig. 3C**). $CD1C^{+}$ DCs (DC2 and DC3) expressed unique markers (*e.g.*, *CD1C*, *CLEC10A*, *FCER1A*, *FCGR2B*, and *CD1D*) enriched for antigen processing ($p < 2.66 \cdot 10^{-10}$), MHC II ($p < 1.79 \cdot 10^{-8}$) and leukocyte activation ($p < 1.14 \cdot 10^{-6}$) gene ontology (GO) terms

(**Fig. 3C; Table S6**) (6). In contrast, Mono1 cells were enriched for defense response ($p < 2.15 \cdot 10^{-14}$), inflammatory response ($p < 9.59 \cdot 10^{-14}$), and chemotaxis ($p < 6.77 \cdot 10^{-10}$) genes.

Finally, we interrogated the relationship between CD16-expressing CD141⁺CD1C⁺ cells (cluster DC4) to CD16⁺ monocytes (cluster Mono2). Although the two populations shared many genes (e.g. *FCGR3A*), they formed distinct clusters (**Fig. 3B**) defined by unique discriminative gene set (**Fig. 3C; Tables S7-S8**). DC4 cells were enriched for type I interferon signaling pathway ($p < 1.53 \cdot 10^{-13}$) and response to virus ($p < 4.77 \cdot 10^{-9}$), while Mono2 cells were enriched for immune system process ($p < 1.09 \cdot 10^{-14}$) and leukocyte migration ($p < 3.57 \cdot 10^{-8}$) GO terms. Although we conclude that monocytes and DCs are distinct from each other in the steady state, our data does not address interconversion between cell fates or distinct ontogeny.

AXL⁺SIGLEC6⁺ population and its relation to cDCs and pDCs

As described above, a population emerged from the unbiased cluster analysis (cluster DC5; **Fig. 1**), defined by expression of unique markers (e.g., *AXL*, *SIGLEC1*, *SIGLEC6*, and *CD22/SIGLEC2*) (**Fig. 4A, Fig. S5A, Tables S1-S2**). Flow cytometry analysis of PBMCs from 10 independent donors confirmed the existence of AXL⁺SIGLEC6⁺ cells ('AS DCs') within the original DC gate (**Fig. 4B**), at a 2-3% frequency consistent with what was originally observed in the initial scRNA-seq analysis (30 of 768 DCs; **Fig. 1C**). scRNA-seq profiling of prospectively sorted AS DCs cells (isolated with the gating strategy in **Fig. 4B**) showed the newly sorted cells clustering together with the original cluster (**Fig. 4C, Fig. S5B**), further validating our successful enrichment strategy.

AS DCs exhibited a spectrum of states based on gene expression (**Fig. 4D**) defined by cells enriched for a pDC-like signature (*e.g. IL3RA, IGJ, NRPI, MZB1*) and cells enriched for a cDC-like signature (*IFI30, ITGAX, LY86, GLIPR2, FGR, LYZ, ENTPD1*). We validated this observation by flow cytometry, using the surface markers CD123/*IL3RA* and CD11C/*ITGAX* that respectively correlated with pDC and cDC gene signatures (**Fig. 4B, 4D**). We exploited the combinatorial expression of AXL, SIGLEC6, CD123 and CD11C (at mRNA and protein level) to prospectively isolate the ends of this spectrum representing 2 putative AS DC subtypes (see gating strategy in **Fig. 4B**), and further validated their identities by scRNA-seq (**Fig. 4E, Fig. S5C-F**). Across all ten individuals tested, the two AS DC subpopulations represent a very small fraction of the Lin⁺HLA-DR⁺ populations: 0.1% for AXL⁺SIGLEC6⁺CD123⁺CD11C^{-/lo} cells and 0.04% for AXL⁺SIGLEC6⁺CD123^{lo}CD11C⁺ cells (**Fig. 4F**). Notably, lower levels of AXL and SIGLEC6 protein were associated with increased HLA-DR, CD11C and CD1C, while higher levels of AXL and SIGLEC6 were associated with increased CD123, CD303, CD141 and lower HLA-DR (**Fig. S5C-J**). This latter relationship was also observed by *t*-SNE analysis of flow cytometry data, where a peninsula with graded expression of AS DCs was located at the base of the CD1C⁺ DC cluster and adjacent to the pDC cluster (**Fig. 4G**). Trajectory mapping of these cells across different levels of surface markers CD123 and CD11C further highlighted that AS DC form a continuum from pDC to CD1C⁺ DC transcriptional state (**Fig. S5C-F**). Taken together, our data suggest that AXL⁺SIGLEC6⁺ DCs are related but not identical to cDCs or pDCs.

pDCs are phenotypically and functionally distinct from CD123⁺CD11C⁻ AS DCs

Since pDCs and AXL⁺SIGLEC6⁺CD123⁺CD11C^{-/lo} DCs shared expression of many genes (**Fig.**

4D-E; Fig. S6A), we assessed whether these cell types also shared functional properties. Importantly, we found that the genes specifically expressed by pDCs, but not AS DCs, were associated with the known biological properties of pDCs. This includes, for example, pathogen sensing and induction of type I IFNs (*IRF7*, *TLR7*, *SLC15A4*, *PACSINI*), secretion (e.g. *DERL3*, *LAMP5*, and *SCAMP5*), and the pDC master regulator transcription factor *TCF4*, along with its binding targets (e.g. *SLA2*, *PTCRA*, *PTPRCAP*) (**Fig. 5A; Fig. S6A**) (18-19). In contrast, CD123⁺CD11C^{-/lo} AS DCs expressed cDC markers, including *CD2*, *CX3CR1*, *CD33* (*SIGLEC3*), *CD5* and *SIGLEC1* both at protein and mRNA levels (**Fig. 5A; Figs. S6A-C**). pDCs were also morphologically distinct from AS DCs. Both AS DC subsets possessed the same cerebriform nucleus and cytoplasmic features of cDCs (**Fig. 5B**). We hypothesized that although CD123⁺CD11C^{-/lo} AS DCs expressed pDC markers, including *CD123/ILRA* and *CD303/CLEC4C* (**Fig. S5G-J**), they are functionally distinct from pDCs.

To compare the functional properties of ‘pure’ pDCs to AS DCs and cDCs, we used the markers identified in our study to isolate ‘pure’ pDCs by excluding AS DCs, CLEC9A⁺ DCs, CD11C⁺ DCs and monocytes by FACS. As expected, ‘pure’ pDCs produced their hallmark cytokines, IFN α , while AS DCs produced negligible amounts of IFN α upon TLR9 stimulation ($p<0.001$; **Fig. 5C**). In contrast, the CD123^{lo}CD11C⁺ AS DC subset secreted IL-12p70 at similar levels to other cDCs, while pure pDCs and CD123^{hi}CD11C^{-/lo} AS DCs did not produce IL-12p70 ($p<0.01$; **Fig. 5C**). Other factors, such as IL-8, were produced at high levels by the CD123⁺CD11C^{-/lo} AS DC subset but not by pDC ($p<0.001$; **FigS6D**). Finally, ‘pure’ pDCs induced undetectable or reduced levels of T cell proliferation in response to LPS or LPS+R848, respectively ($p<0.05$; **Fig. 5D**). We conclude that ‘pure’ IFN α -producing pDCs (that lack AS DCs) do not upregulate

CD86 (**Fig. S6C,E**), are diminished in their ability to induce of T cell proliferation, and that contamination of AS DCs within the traditionally defined pDC gate is likely responsible for T cell stimulation activities measured in prior reports (18-20).

AS DCs stimulate T cell proliferation and are present in tonsils

To understand the potential for AS DCs to stimulate T cells, we first considered their expression of costimulatory molecules and cytokines. While CD80 and CD83 were absent across all DCs, both AS DC subsets expressed CD86 at comparable levels to CD1C⁺ DCs and CLEC9A⁺ DCs (**Fig. S6E**). Strikingly, both AS DC subtypes were potent stimulators of allogeneic CD4⁺ and CD8⁺ T cell proliferation, unlike pDCs ($p < 0.01$), and were marginally superior to CD1C⁺ and CLEC9A⁺ DCs (**Fig. 5E**). These latter results are consistent with AS DCs expressing costimulatory molecule CD86 along with components of the antigen processing machinery (**Fig. S6A,C**).

Similarly to other DCs, AS DCs expressed CLA and CD62L but not CCR7 protein (**Fig. S6F**), suggesting potential homing to peripheral tissue such as skin and lymph node from the circulation. Since CD123⁺ pDCs were observed in the T cell area of the human tonsil (21), we evaluated whether CD123⁺ AS DCs were also present by staining human tonsils with antibodies to CD123 and AXL. We found AS DCs adjacent to CD3⁺ T cells, admixed with CD123⁺AXL⁻ pDCs (**Fig. 5F**). Flow cytometry confirmed this finding, showing that the CD123⁺CD11C^{-/lo} AS DCs represented 0.7% and CD123⁻CD11C⁺ AS DCs represented 1.7% of the CD45⁺LIN⁻HLA-DR⁺ fraction (**Fig. 5F**). Thus, AS DCs are both able to stimulate T cells and are present in the T cell zones of tonsils.

Identification of circulating CD100^{hi}CD34^{int} cDC progenitors

Finally, we interrogated CD11C⁻CD123⁻ cells within the HLA-DR⁺CD14⁻ gate used for isolating DCs but were not considered in the initial analysis since they were not previously thought to include DCs, (red dashed gate in **Fig. 1B** and updated gate in **Fig. 6A** used for these experiments). Analysis of CD11C⁻CD123⁻ scRNA-seq data revealed 6 clusters in this gate (**Fig. S7A-B**). Cells in cluster 6 expressed genes associated with hematopoiesis, DC progenitors, and genes essential for DC development (e.g. *SATB1*, *RUNX2*, *KIT*, *HLX*, *ID2*) (22-25), and were marked by high expression of the cell surface protein *SEMA4D* (CD100). We therefore hypothesized that cluster 6 could represent a progenitor population.

To assess the progenitor potential of this compartment, we cultured FACS-purified CD11C⁻CD123⁻ cells with MS5 stromal cells and cytokines that induce DC differentiation (6), based on a published human DC progenitor differentiation assay (26). After several days in culture, the cells were evaluated by flow cytometry, using a panel of antibodies that identify pDCs, CD1C⁺ and CLEC9A⁺ DCs (6), and by scRNA-Seq profiling of CD45⁺ immune cells for a more comprehensive assessment. For comparison, under the same conditions, we monitored the differentiation potential of isolated pDCs, CD1C⁺ and CLEC9A⁺ DCs, as well as AS DC subtypes (see **Fig. S7C-D**).

After 7 days of culture, cells isolated from the CD11C⁻CD123⁻ gate gave rise to CLEC9A⁺ and CD1C⁺ DCs, based on flow cytometry and scRNA-seq analyses (**Fig. 6B**). While at day 7 some of the cells expressed low levels of CD303 or CD123 proteins, they did not express a *bona fide* pDC transcriptional signature by scRNA-seq. We narrowed down the search for the progenitor

cells to the CD45RA⁺CD39⁻CD100⁺ pool of cells based on the unique cluster-6 marker CD100/SEMA4D (**Fig. S7B**), along with candidate markers that we tested (based on DC progenitors in the bone marrow (CD45RA) and tissue DC (CD39) markers) (**Fig. 6C, Fig S7C-F**). After iterative testing each sorted population for differentiation potential, we discovered that only the CD100^{hi}CD34^{int} cells generated CLEC9A⁺ and CD1C⁺ DCs (**Fig. 6C; Fig. S7F**). ScRNA-seq of CD100^{hi}CD34^{int} cells mapped these cells to the original cluster-6, including the expression of the same DC differentiation and progenitor function genes (**Fig. S7B**).

We validated the existence of CD100^{hi}CD34^{int} progenitors in 10 individuals, with a frequency of ~0.02% of the LIN⁻HLA-DR⁺ fraction of PBMCs (**Fig. 6D**). These cells were morphologically primitive, possessing high nuclear to cytoplasmic ratio and circular or indented nuclei (**Fig. 6D**), compared to AS DCs, pDCs, CD1C⁺ and CLEC9A⁺ DCs (**Fig. 5B**). Furthermore, CD100^{hi}CD34^{int} cells retained significant proliferative capacity ($p < 0.05$; **Fig. 6E**), in accordance with their primitive morphology, phenotype and expression profile. Although CD100^{hi}CD34^{int} cells were *CD117*⁺*KIT*⁺CD45RA⁺ and CD115⁻, CD1C⁻, CD141⁻, CD123⁻, a profile similar to previously reported circulating human DC progenitor (24, 27-28), they differ from the published progenitor in having a more primitive morphology and lacking *CD116*/*CSF2R* and *CD135*/*FLT3* expression (**Fig. S7G-H**).

While expressing some level of the co-stimulatory molecule CD86 (**Fig. S6E**), the CD100^{hi}CD34^{int} cells have low T cell stimulatory potential (**Fig. 5C**). CD100^{hi}CD34^{int} cells expressed CLA and CD62L at similar protein amounts to cDC (**Fig. S6F**). Homing *CCR7* gene was expressed in these cells (**Fig. S7B, S7H**) but protein expression was modest (**Fig. S6F**).

Both CD100^{hi} CD34^{int.} cells and AS DCs were CD45RA⁺ and CD38⁺ (**Fig. S5J**; **Fig. S6B**).

Differentiation potential of AS DCs

We seeded cultures with pDCs, CD1C⁺ and CLEC9A⁺ DCs and found that they generally retained the same phenotype throughout the differentiation assay (**Fig. 6F**, **Fig. S7I-J**). Upon observing a gene expression spectrum of AS DC states that includes pDC- and CD1C⁺-like DC signatures (**Fig. S5C-F**), we also seeded AS DCs to assess their potential to transition towards other DC subsets (ensuring no contamination with CD1C⁺ and CLEC9A⁺ DCs (**Fig. S7I-J**)). After 7 days in culture, we observed cells with high levels of CD1C (frequency 40%-50%, $n=6$ donors) and rare cells with surface CLEC9A and CADM1 (0.5-0.8%) expression (**Fig. 6F**), regardless of FLT3L concentration used (**Fig. 6F**) or if the culture was seeded with either of the two AS DC subpopulations representing both ends of the spectrum (**Fig. S7K**). Notably, both AS DCs at day 0 and the cells generated from AS DC differentiation did not express *BATF3* (transcription factor required for terminal differentiation of CLEC9A⁺ DCs), *CADM1* or *XCRI*, which are key CLEC9A⁺ DC discriminative markers (**Table S2**)(23, 29-33), consistent with the absence of expression of these genes across the spectrum of AS DCs at steady state (**Fig. S5D-E**). Importantly, AS DCs do not divide during the transition into CD1C⁺ DCs, in contrast to CD100^{hi}CD34^{int.} cells that divide and differentiate into CD1C⁺ as well as CLEC9A⁺ DCs. Furthermore, CD100^{hi}CD34^{int.} differentiation into CD1C⁺ DCs is not likely to transition through AS DCs since CD100^{hi}CD34^{int.} do not express *AXL* or *SIGLEC6* genes at day 0 or during differentiation. AS DCs are thus functional cDCs that exist in a continuum of states *in vivo* (**Fig. S5C-F**), with the potential to transition towards CD1C⁺ DCs.

Mapping malignant cells from patients to the healthy DC atlas

We leveraged our human DC atlas to compare pathogenic cells driving blastic plasmacytoid dendritic cell neoplasm (BPDCN), a rare and aggressive hematological malignancy previously known as natural killer (NK) cell leukemia/ lymphoma (34-35), to healthy DC populations. Since the ontogeny of these cells remains unclear (34-38), we performed scRNA-seq on CD45⁺HLA-DR⁺CD123⁺ blasts from 4 BPDCN patients (n=174 cells) (6). The first principal component highlighted gene sets clustering all 4 patients together with healthy blood pDCs (**Fig. 6G**). Analysis of BPDCN samples together with healthy DCs showed highest overlap with pDC and AS DC gene expression signatures (**Fig. S8A**). Since pure pDC and AS DC subsets co-express many genes, yet have distinct biological function (**Figs. 4-5**), we further analyzed the genes overlapping between BPDCN, pure pDCs and cDCs (**Fig. S8B**). Despite sharing some pDCs genes (e.g. *NRP1*, *IL3RA*, *DERL3*, *LAMP5*, *PTCRA* and *PTPRCAP*), several key genes essential for pDC function were missing or very lowly expressed in patient cells (e.g. *GZMB*, *IRF7*, *CLEC4C/CD303*, *IRF4*, *SLC15A4*; **Fig. S8B**). Only a small number of cDC genes were expressed in patient cells, including *SIGLEC6*, *LTK*, *FCER1A*, *CD59*, *CADM1*, and *TMEM14A*. Noteworthy, all 4 patient samples shared a set of discriminative genes (**Fig. S8B; Table S9**) that included several genes expressed in B cells (e.g. *FCRLA*, *IGLL1*, *TCL1A*, *IGLL5*; **Fig. S8C**) or with hematopoietic progenitors (e.g. *SOX4* and *CLEC11A*). Collectively, our analysis suggests that while BPDCN malignant cells express some key B cell markers, they are most closely related to pDCs.

Discussion

DCs and monocytes are defined based on a combination of molecular markers, functional properties and ontogeny (39). However, an open question is whether the expression of existing markers tracks with the more complex internal states of cells. To address this question, we determined the states of blood DC/monocytes through comprehensive profiling of gene expression at single cell resolution, empirically inferred cell subtypes, identified optimal surface markers for purifying the hypothesized cell subtypes, and showed that prospectively purified cell types corresponded to inferred subtypes based on scRNA-seq. Our study has generated a more accurate taxonomy that includes 6 DC and 4 monocyte subtypes, as well as a circulating, dividing progenitor of cDCs.

Previous studies classified human blood DCs into one pDC and two cDC populations. Our study identifies 6 DC populations: DC1 corresponds to the cross-presenting $CD141/BDCA-3^+$ cDC1, which is best marked by CLEC9A; DC2 and DC3 correspond to new subdivisions of the $CD1C/BDCA-1^+$ cDC2; DC4 corresponds to $CD1C^-CD141^-CD11C^+$ DC, which is best marked by CD16 and shares signatures with monocytes; DC5 is a unique DC subtype, AS DCs; and DC6 corresponds to the interferon-producing pDC, purer than previously identified pDC population defined by standard markers (CD123, CD303/*BDCA-2*) and contaminated with AS DCs. In the process of addressing how DCs resemble monocytes, we also identified 4 monocyte subtypes – the 2 known ones, a monocyte killer subtype and a subtype characterized by cell cycle-arrest and trafficking gene sets. Although DC2/3 and DC4 shared an expression signature with monocytes, our data does not suggest how they acquired these shared modules (common precursor, interconversion or independent convergence). Finally, we derived specific expression signatures

for each DC and monocyte subtype, including transcription factors, cytokines, and cytokine receptors (**Fig. S9A-F; Tables S10A-10F**), providing a resource for further understanding of subtype functions and ontogeny.

The CD1C/*BDCA-1*⁺ DCs subdivision (DC2 and DC3) is further supported by parallel observations in their murine CD11b⁺ DC homologs (40-43) that comprise *Esam*^{lo} subset with higher expression of myeloid genes such as CD14 and potent cytokine production, and *Esam*^{hi} subset with better MHC class II-dependent priming of CD4⁺ T cells (40-41).

AS DCs, which were found within the pDC gate, formed a continuum between pDC and CD1C⁺ DC (see trajectory mapping; **Fig. S5C-F**). Consistent with this observation, AS DCs were able to transition towards the CD1C⁺ DC state *in vitro* (with <1% of differentiated AS DCs phenotypically resembling CLEC9A⁺ DCs, which could be contaminants). However, since AS DCs (at both ends of the continuum) morphologically resemble cDCs and are able to stimulate T cell proliferation, yet do not proliferate themselves, they seem less likely to serve as a progenitor that generates cDCs, but rather a functional DC variant that can be modulated to resemble CD1C⁺ DCs. While AS DCs most closely resemble CD1C⁺ cDCs in basic functional properties and expression signatures, they are likely to have distinct functions based on their localization in the T cell zone of tonsils and their expression of lectins that recognize diverse glycans, and AXL which interacts with apoptotic cells and Zika virus (44-46).

An unresolved question is the significance of AS DCs sharing an expression signature with pDCs. Consistent with our findings that AS DCs are found in the traditional pDC flow cytometry

gate, a recently described human CD2^{hi} pDC subset (20) appears to correspond to AS DCs based on expression of *CD2*, *AXL*, *CX3CR1*, *LYZ* and *CD86* (**Fig. S6C**), localization to tonsil, and a similar ability to trigger naive T cell proliferation. Furthermore, a murine study identified non-canonical CX3CR1⁺CD8 α ⁺ cDCs (nc-cDCs), which expressed pDC and cDC signatures (e.g., CX3CR1, CD11c and MHCII), do not produce IFN α , and activate T cell proliferation (47-48). Interestingly, pDC and nc-cDCs require E2-2/TCF4 to develop, and reduced levels of E2-2 lead to higher ID2 and expression of cDC genes (18, 47-48). Consistent with this finding, we observe E2-2/TCF4 expression in human pDCs (**Fig. 5A**), with decreasing levels of E2-2/TCF4 and increasing levels of ID2 as AS DCs transition to CD1C⁺ DCs (**Fig. S5C-F**). These findings suggest that AS DCs are similar to human CD2^{hi} pDCs and murine nc-cDCs.

The discovery of AS DCs led us to update the strategy for isolating pDCs. When we removed AS DCs from pDCs isolated with standard markers (e.g. CD123 and CD303), the resulting pDCs were highly attenuated in their ability to induce T cell proliferation and produce T cell stimulatory ligands (e.g., IL-12), consistent with reports that found several markers splitting pDCs into those that stimulate or do not stimulate T cells (18,20, 49-52). We thus propose that our purer pDC population corresponds more closely to the ‘natural interferon-alpha producing cells’ (21, 53). These cells also appear to share more properties with plasma B cells than DCs based on morphology, higher expression of ER/secretory machinery, known rearrangement at the Ig locus, and expression of B-cell related transcripts. We also found that BPDCN cells share the pDC signature as well as additional B cell genes (e.g. *IGLL1*, *IGLL5*, *TCL1A*). We conclude that while pure pDCs do fall into the MHC II-expressing gate, they have markers, gene signatures, and functions distinct from cDCs.

In contrast to AS DCs, the CD100^{hi}CD34^{int} cells appear to be cDC progenitors based on their primitive morphology, absence of cDC functions and signatures, and their potent ability to proliferate and generate a large and equal number of CD1C⁺ DC and CLEC9A⁺ DC within 7 days of culture. The recently identified human pre-cDC (24-28), which has proliferative capacity and differentiates into CD1C⁺ and CD141⁺ DCs, appears to have some functional and phenotypical similarities with our CD100^{hi}CD34^{int} progenitors, though our cells appear to be morphologically more primitive and lack the expression of CD116 and CD135, which were previously reported as markers (24). Single cell profiling studies are needed to determine whether and how these precursors are related.

CD100^{hi}CD34^{int} cells also appear to be different from peripheral blood CD34^{hi} HSCs. Culturing of CD100^{hi}CD34^{int} cells gives rise only to CLEC9A⁺ DCs and CD1C⁺ DCs in 7 days (and no other cell types). In contrast, peripheral blood CD34^{hi} HSCs under the same culture conditions for up to 14 days did not give rise to CLEC9A⁺ cDCs. Furthermore, CD100^{hi}CD34^{int} cells have a distinct transcriptional signature from blood CD34^{hi} HSCs. Mapping CD100^{hi}CD34^{int} to other bone marrow progenitors may help resolve the origin of these cells.

Our results have several implications. The discovery of several DC subsets will enable a more complete understanding of DCs in tissues, inflammation and disease. Furthermore, the identification of circulating CD100^{hi}CD34^{int} progenitors provides a well-defined cell type for generating DCs *in vitro* and for therapeutic targeting. Our new strategy for isolating pure pDCs, combined with the knowledge that the functions of contaminating AS DCs were incorrectly attributed to pDCs, should lead to more definitive annotation of pDC functions with implications

for their therapeutic application (54-56). More generally, our use of the DC atlas to understand BPDCN cells illustrates how single cell analysis can pinpoint relationships of diseased to healthy cells. Finally, some susceptibility genes identified in human genetics association studies are expressed in the DCs and monocytes subsets defined in this study, suggesting new potential roles in disease (**Fig. S10A-B; Table S11A-C**).

Using single cell transcriptome profiling, we deconvoluted admixtures of cell types (e.g., pDCs, ‘intermediate’ monocytes, cDC progenitors), revealed rare cell types (e.g., AS DCs) and elucidated complex relationships between cell types (e.g., spectrum of states for AS DCs) – thus addressing limitations in the existing classification that relies on a small number of markers (39). Nevertheless, some DC/monocyte subtypes were likely missed because they either do not express MHC Class II at rest, can only be defined by non-RNA molecules, are distinguished by lowly expressed transcripts, or are only present during inflammation, disease or within tissues. To build a comprehensive immune cell atlas, future studies will need to address these challenges as well as localize these cell types within lymphoid and non-lymphoid tissues.

Materials and Methods

Study subjects

The study was performed in accordance with protocols approved by the institutional review board at Partners (Brigham and Women's Hospital, Massachusetts General Hospital, Dana-Farber Cancer Institute; Boston, USA) and Broad Institute (USA), as well as the Newcastle upon Tyne Hospitals (UK) Research Ethics Committee. All patients provided written informed consent for the genetic research studies and molecular testing. Healthy donors were recruited

from the Boston-based PhenoGenetic project, a resource of healthy subjects that are re-contactable by genotype (57), and the Newcastle community. Individuals were excluded if they had a history of cancer, allergies, inflammatory disease, autoimmune disease, chronic metabolic disorders or infectious disorders. All healthy donors were non-smoker, had a normal BMI and normal blood pressure, and were between 25-40 years of age.

Cell isolation, flow cytometry staining, cell sorting, and analysis

For profiling of healthy cells, peripheral blood mononuclear cells (PBMCs) were isolated from fresh blood within 2hrs of collection, using Ficoll-Paque density gradient centrifugation as previously described (58). Single-cell suspensions were stained per manufacturer recommendations with different panels of antibodies (**Table S12**) designed to enrich for certain population for single cell sorting and single cell RNA-sequencing (scRNA-seq). Refer to supplementary materials and method section for further details (6). Flow cytometry and FACS-sorting of PBMC was performed on a BD Fortessa or BD FACS Fusion instrument, and data analysed using FlowJo v10.1. Single-cells were sorted into 96-well full-skirted eppendorf plate chilled to 4°C, pre-prepared with lysis buffer consisting of 10µl TCL buffer (Qiagen) supplemented with 1% beta-mercaptoethanol. Single-cell lysates were sealed, vortexed, spun down at 300g at 4°C for 1 minute, immediately placed on dry ice and transferred for storage at -80°C. Tonsil was mechanically disrupted to obtain single-cell suspension.

Single-cell RNA-sequencing

Smart-Seq2 protocol was performed on single sorted cells as described (7-8), with some modifications (6). For DCs, a total of 8 x 96-well plates (768 single DCs) were initially profiled

from the same blood draw and sort from the index volunteer and subsequent validation performed on an additional ten healthy individuals. For monocytes, a total of 4 plates were profiled (372 single monocytes and 12 population samples). An additional 975 single cells were profiled to further characterize the CD1C⁺ DC subsets (n=125), AXL⁺SIGLEC6⁺ cells (n=372), CD11C⁻CD123⁻ compartment at day 0 (n=164), differentiation assay outputs (n=218), CD100^{hi}CD34^{int.} cells (n=96), and BPDCN patient samples (n=269). Noteworthy, some of these single cells were excluded from the analysis after applying QC filters and analytically confirming cell type (6). Refer to supplementary materials and methods for further details on the scRNAseq processing (6).

Single-cell RNA sequencing analyses

Raw sequencing data were processed as previously described (60) (see **Table S13-S16** for cell identities that accompanies raw data and gene expression matrices). Briefly, short sequencing reads were aligned to the UCSC hg19 transcriptome. These alignments were used to estimate transcriptomic alignment rates, and were also used as input in RSEM v 1.2.1 to quantify gene expression levels (transcripts per million; TPM) for all UCSC hg19 genes in all samples. We filtered out low-quality cells from our dataset based on a threshold for the number of genes detected (a minimum of 3000 unique genes/cell for cells sequenced at HiSeq depth, and 2000 unique genes/cell for cells sequenced at MiSeq depth). All genes that were not detected in at least 0.5% of all our single cells were discarded, leaving 21,581 genes for all further analyses. Data were log transformed (log(TPM+1)) for all downstream analyses, most of which were performed using the R software package Seurat (<https://github.com/satijalab/seurat>;

<http://satijalab.org/seurat/>). Refer to supplemental materials and method section for further details, including R script used to generate clusters (6).

DC differentiation assay on MS5 stromal cells

DC differentiation assay was performed as previously described (23-25) with minor adaptation. Briefly, 1×10^4 purified progenitors, DC, and monocyte subsets were cultured in 96-well flat bottomed plate layered with 4×10^4 murine MS5 stromal cells (DSMZ, Germany) in the presence of human FLT3-ligand (FL; 100 ng/ml; Miltenyi Biotec), recombinant human SCF (20 ng/ml; R&D Systems) and recombinant human GM-CSF (10 ng/ml; Peprotech). MS5 stromal cells were seeded 24 hours prior to co-culture. Growth factors were replenished on day 3 of culture. Cells were in culture for up to 7 days prior to harvesting by physical dissociation on ice. Cells were then stained on ice either for flow cytometry analysis (see output panel in **Table S12**) or single cell index sorting of CD45⁺ cells for scRNAseq of culture output analysis.

Cytokine production measurements

Purified subsets were cultured at 5×10^3 cells/well in 96 well round bottom plates in the presence of LPS (100ng/ml; Invivogen) and ODN2395 (1 μ M; Invivogen) or ODN5328 (ODN2395 control, 1 μ M; Invivogen), or in the presence of LPS, poly (I:C) (25 μ g/ml; Invivogen) and R848 (2.5 μ g/ml; Enzo Life Sciences). Culture supernatants were harvested after 24 hours and analyzed using a multiplexed cytokine assay (ProcartaPlex, eBioscience), or by leveraging the 92 inflammatory-related protein biomarker panel and 4 controls provided by Olink Proteomics (Uppsala, Sweden). Refer to supplementary methods for further details (6).

Assessing T cell stimulatory potential

DC, monocyte, and progenitor subsets were purified from peripheral blood of healthy donors by FACS sorting (BD FACS Fusion; see **Table S12** for sorting panels and antibodies). For T cell stimulatory potential, purified DCs, monocytes, AXL⁺SIGLEC6⁺ subsets, and progenitor subset were cultured at 5x10⁴/well cell density. All purified cell subsets were matured with LPS (100ng/ml, Sigma) and R848 (2.5ug/ml, Invivogen), or with just LPS (100ng/ml, Sigma), for 24 hours prior to co-culture with 5x10⁵ CFSE-labelled allogeneic unfractionated CD3⁺ T cells at a 1:10 DC:T cell ratio. T cell proliferation was assessed by measuring CFSE dilution on day 5 of culture.

Cytospin and immunostaining

Cytospin of FACS-purified cells were prepared as previously described (76) using Shandon Cytospin 4 (Thermo Scientific). Giemsa-Wright staining was performed using Advia S60 (Siemens) and imaged using Axioimager.Z2 microscope with Axiovision software v4.8 (Carl Zeiss, Germany). Human tonsil paraffin sections were immunostained with the following antibodies (clone: manufacturer): anti-AXL (MM0098-2N33: Abcam), CD123 (BR4MS: Leica Biosystems) and CD3 (LN10: Leica Biosystems) using a Ventana Benchmark XT instrument.

Monitoring cell proliferation

PBMCs were labeled with Cell Trace Violet (CTV, Life Technologies) according to manufacturer's protocol. CTV-labeled FACS-purified progenitors and DC subsets were cultured on murine MS5 stromal cells as described above and analyzed on day 5 to assess proliferation measured by CTV dilution.

References

1. M. Haniffa, V. Bigley, M. Collin, *Semin. Cell Dev. Biol.* **41**, 59-69 (2015).
2. A. Mildner, S. Jung, *Immunity* **40**(5), 642-656 (2014).
3. B.U. Schraml, C. Reis e Sousa, *Curr Opin Immunol.* **32**,13-20 (2015).
4. C. Trapnell, *Genome Res.* **25**(10), 1491-1498 (2015).
5. D. Grün, A. van Oudenaarden, *Cell* **163**(4), 799-810 (2015).
6. Materials and methods are available as supplementary materials at the Science website.
7. S. Picelli *et al.*, *Nat. Methods.* **10**, 1096–1098 (2013).
8. J. J. Trombetta *et al.*, *Curr. Protoc. Mol. Biol.* **107**, 4.22.1–4.22.17 (2014).
9. E.Z. Macosko *et al.*, *Cell* **161** (5), 1202-1214 (2015).
10. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. **9**, 2579–2605 (2008).
11. J.H. Levine *et al.*, *Cell* **162**(1),184-197 (2015).
12. H. Liu *et al.*, *BMC Bioinformatics.* **17**, 171 (2016).
13. V. Pascual, D. Chaussabel, J. Banchereau, *Annu. Rev. Immunol.* **28**, 535-571 (2010).
14. B. Mesko, S. Poliska, L. Nagy, *Trends Mol. Med.* **17**(4), 223-233 (2011).
15. R.S. Welner *et al.*, *Blood* **109** (11), 4825-4931 (2007).
16. J. Taieb *et al.*, *Nat. Med.* **12**(2), 214-219 (2006).
17. C.W. Chan *et al.*, *Nat. Med.* **12**(2), 207-113 (2006).
18. M. Swiecki, M. Colonna, *Nat. Rev. Immunol.* **15**(8), 471-485 (2015).
19. M. Cheng *et al.*, *Sci. Rep.* **5**, 10752 (2015).
20. T. Matsui *et al.*, *J. Immunol.* **182** (11), 6815-6823 (2009).
21. G. Grouard *et al.*, *J. Exp. Med.* **185**(6), 1101-1111 (1997).
22. J.C. Miller *et al.*, *Nat. Immunol.* **13**(9), 888-899 (2012).

23. A.T. Satpathy, X. Wu, J.C. Albring, K.M. Murphy, *Nat. Immunol.* **13**(12), 1145-54 (2012).
24. G. Breton *et al.*, *J. Exp. Med.* **212** (3), 401-413 (2015).
25. J. Lee J *et al.*, *J. Exp. Med.* **212** (3), 385-399 (2015).
26. G. Breton, J. Lee, K. Liu, M.C. Nussenzweig, *Nat. Protoc.* **10** (9),1407-1422 (2015).
27. F. Notta *et al.*, *Science* **351** (6269), aab2116 (2016).
28. S. Doulatov, F. Notta, E. Laurenti, J.E. Dick, *Cell Stem Cell* **10** (2), 120-136 (2012).
29. K. Hildner *et al.*, *Science* **322** (5904), 1097-1100 (2008).
30. L.F. Poulin *et al.*, *J. Exp. Med.* **207** (6), 1261-1271 (2010).
31. K. Crozat *et al.*, *J. Exp. Med.* **207** (6), 1283-1292 (2010).
32. S.L. Jongbloed *et al.*, *J. Exp. Med.* **207** (6), 1247-1260 (2010).
33. M. Haniffa *et al.*, *Immunity* **37**(1), 60-73 (2012).
34. W. Riaz, L. Zhang, P. Horna, L. Sokol, *Cancer Control.* **21** (4), 279-289 (2014).
35. F. Garnache-Ottou, J. Feuillard, P. Saas, *Br. J. Haematol.* **136** (4), 539-548 (2007).
36. M.R. Sapienza *et al.*, *Leukemia* **28** (8), 1606-1616 (2014).
37. Y. Osaki *et al.*, *PLoS One* **8** (11), e81722 (2013).
38. H. Yu *et al.*, *Protein Cell* **6** (4), 297-306 (2015).
39. M. Guilliams *et al.*, *Nat. Rev. Immunol.* **14** (8), 571-578 (2014).
40. K.L. Lewis *et al.*, *Immunity* **35** (5), 780-791 (2011).
41. A.T. Satpathy *et al.*, *Nat. Immunol.* **14** (9), 937-948 (2013).
42. A. Schlitzer *et al.*, *Immunity* **38** (5), 970-983 (2013).
43. E.K. Persson *et al.*, *Immunity* **38** (5), 958-69 (2013).
44. M.S. Macauley, P.R. Crocker, J.C., *Nat. Rev. Immunol.* **14**(10), 653-666 (2014).

45. C.V. Rothlin, E.A. Carrera-Silva, L. Bosurgi, S. Ghosh, *Annu. Rev. Immunol.* **33**, 355-391 (2015).
46. T.J. Nowakowski *et al.*, *Cell Stem. Cell.* **18** (5), 591-596 (2016).
47. L. Bar-On *et al.*, *PNAS* **107**(33), 14745-14750 (2010).
48. C.M. Lau *et al.*, *J. Exp. Med.* **213** (3), 415-31 (2016).
49. T.R. Wilhelm *et al.*, *Clin. Immunol.* **163**, 42-51 (2016).
50. N. Schwab *et al.*, *J. Immunol.* **184** (9), 5368-5374 (2010).
51. Q. Du *et al.*, *Cell Mol. Immunol.* **8** (5), 441-444 (2011).
52. C. Bryant *et al.*, *Immunol. Cell Biol.* **94** (5), 447-57 (2016).
53. M. Cella *et al.*, *Nat. Med.* **5** (8), 919-923 (1999).
54. J. De Vries, C. Figdor, *Nature* **534** (7607), 329-331 (2016).
55. J. Tel *et al.*, *Blood* **120** (19), 3936-3944 (2012).
56. J. Tel *et al.*, *Cancer Res.* **73** (3), 1063-1075 (2013).
57. Z. Xia *et al.*, *J. Immunol.* **188** (7), 3315-3322 (2012).
58. M.N. Lee *et al.*, *Science* **343**(6175), 1246980 (2014).
59. E.C. Townsend, *et al.*, *Cancer Cell* **29** (4), 574-586 (2016).
60. A.K. Shalek *et al.*, *Nature* **510** (7505), 363-369 (2014).
61. F. Buettner *et al.*, *Nat. Biotechnol.* **33**(2), 155-160 (2015).
62. N.C. Chung, J.D. Storey, *Bioinformatics.* **31**(4), 545-554 (2015).
63. J.T. Leek, J. D. Storey. *Statistical Applications in Genetics and Molecular Biology* 10.1 (2011).
64. C. Xu, Z. Su, *Bioinformatics* **31**(12), 1974-1980 (2015).
65. L. Waltman, N.J. Van Eck, *European Physical Journal B.* **86** (11), 471 (2013).
66. A. McDavid *et al.*, *Bioinformatics* **29**(4), 461-467 (2013).
67. C. Trapnell *et al.*, *Nat. Biotechnol.* **32** (4), 381–386 (2014).

68. J.M. Vaquerizas *et al.*, *Nat. Rev. Genet.* **10** (4), 252-263 (2009).
69. M. Uhlén *et al.*, *Science* **347** (6220), 1260419 (2015).
70. D. Bausch-Fluck *et al.*, *PLoS One* **10**, e0121314 (2015).
71. L.A. Hindorff *et al.*, A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed [June 2016].
72. M. Kanehisa *et al.*, *Nucleic Acids Res.* **44** (D1), D457-462 (2016).
73. M. Kanehisa, S. Goto, *Nucleic Acids Res.* **28** (1), 27-30 (2000).
74. E. Eden *et al.*, *BMC Bioinformatics*.**10**, 48 (2009).
75. S. Darmanis *et al.*, *Cell Reports* **14**, 380-389 (2016).
76. M. Haniffa *et al.*, *J. Exp. Med.* **206** (2), 371-385 (2009).
77. G. X.Y. Zheng *et al.*, *Nat. Comm.* **8**, 14049 (2017).

List of Supplementary Materials

Materials and Methods

References (57-77)

Fig. S1 – S11

Table S1 – S16 are provided as a separate Excel file

Acknowledgments:

We thank the subjects in the PhenoGenetic Project for donating the blood used in our study. We thank Elodie Segura, Sebastian Amigorena, Christophe Benoist and David Puyraimond-Zemmour for advice on the sorting strategy; Michael Waring and the Ragon Institute Imaging Core-Flow Cytometry Facility (which is supported in part by the Harvard University Center for AIDS Research (CFAR), an NIH funded program (5 P30 AI060354-10)); Trevor Booth and Newcastle University Bioimaging Unit for assistance with microscopy; and Leslie Gaffney for illustrations. We also thank Jonathan Scott, Katie Best, Anna Long (MRC/EP SRC Molecular Pathology Node), Claire Jones, David Lieb, Christopher Ford, David Gennert, John Trombetta, and Alexandra Schnell for help and advice. This was supported by the National Human Genome Research Institute Centers of Excellence in Genomics Science P50 HG006193 (N.H., A.R.), Manton Foundation (A.R., N.H.), Howard Hughes Medical Institute (A.R.), NIH BRAIN grant (A.R.), Klarman Family Foundation (A.R., N.H.), Banting Postdoctoral Fellowship (A.C.V.), NIH Director's New Innovator Award Program DP2-HG-009623 (R.S.), American Society of Hematology Scholar grant (AAL), ISAC SRL-EL program (AJF), and Wellcome Trust WT088555 and WT107931/Z/15/Z (M.H.). A.C.V., R.S., A.R., N.H., The Broad Institute, and Massachusetts General Hospital have filed a U.S. provisional patent application (62/376,007) that relates to products and methods useful for modulating and evaluating immune responses. A.R. is a scientific advisory board member for ThermoFisher Scientific and Syros Pharmaceuticals, and a consultant for Driver Group. Processed scRNA-seq data are available through the Gene Expression Omnibus accession number GSE94820. Raw RNA sequencing data is available through Database of Genotypes and Phenotypes (dbGaP) accession number phs001294.v1.p1.

Figure Legends

Figure 1. Human blood DC heterogeneity delineated by single-cell RNA-sequencing.

(A) Workflow of experimental strategy: isolation of human PBMC from blood; sorting single DC (8x96-well plates) and monocytes (4x96-well plates) into single wells using an antibody cocktail to enrich for cell fractions; single cell transcriptome profiling. (B) Gating strategy for single cell sort: DCs were defined as live, lineage (LIN:CD3,CD19,CD56)⁻CD14⁻HLA-DR⁺ cells. Three loose overlapping gates were drawn as an enrichment strategy to ensure a comprehensive and even sampling of all populations: CD11C⁺CD141⁺ (CD141; turquoise), CD11C⁺CD1C⁺ (CD1C; orange), CD11C⁺CD141⁻CD1C⁻ ('Double Negative'; blue), CD11C⁻CD123⁺ plasmacytoid DCs (pDCs; purple). 24 single cells from these 4 gates were sorted per 96-well plate. A fifth gate (CD11C⁻CD123⁻, red dashed gate) was subsequently investigated (see **Fig. 6**). (C) *t*-SNE analysis of DCs (*n* = 742). Number of successfully profiled single cells per cluster include: DC1 (*n* =166); DC2 (*n*=105); DC3 (*n*=95); DC4 (*n* =175); DC5 (*n*=30); DC6 cluster (*n* =171). The number of discriminative genes with AUC cutoff ≥ 0.85 is reported in bracket next to each cluster ID. Up to 5 top discriminators are listed next to each cluster; number in bracket refers to AUC value. Colors indicate unbiased DC classification via graph-based clustering. Each dot represents an individual cell. (D) Heatmap reports scaled expression (log TPM (Transcripts Per Million) values) of discriminative gene sets for each cluster defined in **Fig. 1C** with AUC cutoff ≥ 0.85 . Color scheme is based on z-score distribution from -2.5 (yellow) to 2.5 (purple). Right margin color bars highlight gene sets specific to the respective DC subset.

Figure 2. Definition and validation of CD1C⁺ DC subsets.

(A) Heatmap showing scaled expression (log TPM values) of discriminative gene sets defining each CD1C⁺ DC subset with AUC cutoff ≥ 0.75 . Color scheme is based on z-score distribution, from -2.5 (yellow) to 2.5 (purple). Violin plots illustrate expression distribution of candidate genes across subsets on the x-axis (orange for CD1C_A/DC2; green for CD1C_B/DC3). In red are 3 markers used for subsequent enrichment strategy: *CD163*, *CD36* and *FCGR2B/CD32B* (AUC =0.63). (B) Enrichment gating strategy of CD1C⁺ DC subsets (LIN(CD3,CD19,CD56)⁻ HLA-DR⁺CD14⁻CD1C⁺CD11C⁺). CD1C_A/DC2 subset was further enriched by sorting on the 10% brightest CD32B⁺ cells (orange gate), while CD1C_B/DC3 subset was enriched by sorting on CD32B⁻CD163⁺CD36⁺ cells (green gate), or on CD32B⁻CD163⁺. Right panel: overlay of the two sorted CD1C⁺ DC populations. 47 single cells were sorted from the green and orange gates in a 96-well plate for profiling. (C) Heatmap reporting scaled expression (log TPM values) of scRNAseq data from three cell subsets defined by CD1C⁺CD32B⁺, CD1C⁺CD36⁺CD163⁺, and CD1C⁺CD163⁺. Combining either CD1C⁺CD36⁺CD163⁺ or just CD1C⁺CD163⁺ recapitulated the CD1C_B/DC3 signature. (D) Proliferation of allogeneic CD4⁺ and CD8⁺ T cells five days after co-culture with CD14⁺ monocytes, pDCs, CD1C_A/DC2 DCs (CD1C⁺CD32B⁺), CD1C_B DC3 (CD1C⁺CD163⁺). Left panel depicts representative pseudocolor dot plot and right panel bar graphs of composite data (n=3, mean \pm SEM, * p <0.05, paired t-test).

Figure 3. Human blood monocyte heterogeneity.

(A) Gating strategy for monocyte single cell sorting: monocytes were enriched by first gating on LIN(CD3, CD19, CD56)⁻CD14^{+/lo}, followed by three loose overlapping gates defined by relative expression of CD14 and CD16 for comprehensive sampling of CD14⁺⁺CD16⁻ (yellow gate), CD14⁺⁺CD16⁺ (purple), CD14⁺CD16⁺⁺ (blue). 32 cells from each gate were sorted per 96-well plate profiled. Bottom right dot plot shows overlay of the sorted populations. (B) *t*-SNE analysis incorporating monocytes (*n*=337 successfully profiled) and DCs (*n*=742). Number of successfully profiled single monocytes per transcriptionally defined clusters includes: Mono1, *n*=148; Mono2, *n*=137; Mono3, *n*=31; Mono4, *n*=21. The number of discriminative genes with AUC cutoff ≥ 0.85 (combined analysis of DC and monocyte datasets) is reported in bracket next to cluster ID. Up to 5 top discriminators are listed next to each cluster; number in bracket next to each gene refers to AUC value. Colors indicate unbiased DC and monocyte clustering from unbiased graph-based clustering. Each dot represents an individual cell. (C) Heatmap reporting scaled expression (log TPM values) of discriminative gene sets for each monocyte subsets with AUC cutoff ≥ 0.85 (see **Fig. S3B** for detailed heatmap). Color scheme is based on z-score distribution, from -2.5 (yellow) to 2.5 (purple). Right margin color bars highlight gene sets of interest.

Figure 4. Identification of AXL⁺SIGLEC6⁺ DCs (AS DCs).

(A) Violin plots showing expression distribution of surface markers *AXL* and *SIGLEC6*. Other populations are depicted on the x-axis and each dot represents an individual cell. (B) Flow cytometry gating strategy to identify AXL⁺SIGLEC6⁺ cells within human blood LIN(CD3, CD19, CD20, CD161)⁻ and HLA-DR⁺ mononuclear fraction. AXL⁺SIGLEC6⁺ cells were further distinguished by the relative expression of CD123/IL3RA and CD11C/ITGAX (1=CD123⁺CD11c^{-lo} (pink) and 2=CD123^{lo}CD11c⁺ (blue)). Data shown is a representative analysis of ten healthy individuals. (C) *t*-SNE analysis of all DCs (*n*=742), along with prospectively profiled AXL⁺SIGLEC6⁺ single cells (*n*=105) using gating strategy in panel B (sorted from purple gate). Newly isolated AS DCs overlap with the originally identified DC5 cluster (*n*=30), purple dashed circle. (D) Heatmap reporting scaled expression (log TPM values) of discriminative gene sets (AUC cutoff ≥ 0.75), highlighting the expression continuum of AS DCs. Top bar graph defines the AS DCs population purity score based on the top 10 most discriminative genes (i.e. *AXL*, *PPP1R14A*, *SIGLEC6*, *CD22*, *DAB2*, *S100A10*, *FAM105A*, *MED12L*, *ALDH2*, *LTK*). (E) Heatmap reporting scaled expression (log TPM values) of prospectively enriched AS DCs populations (*n*=90) isolated by relative CD11C/ITGAX and CD123/IL3RA expression levels (red in panel D). 43 single AXL⁺SIGLEC6⁺CD11C⁻ (pink gate, panel B) and 47 single AXL⁺SIGLEC6⁺CD11C⁺ (blue gate, panel B) were sequenced. The average expression values of the original CD1C⁺ (combined DC2 and DC3), CD141⁺/CLEC9A⁺ (DC1) and pDC (DC6) single cells were used as reference to highlight enrichment of cDC-like and pDC-like gene sets. Top bar graph represents AS DC purity score. (F) Frequency (% mean \pm SEM) of AXL⁺SIGLEC6⁺CD123⁺CD11C^{-lo} (population 1 (pink): 0.1 \pm 0.014) and AXL⁺SIGLEC6⁺CD123^{lo}CD11C⁺ (population 2 (blue): 0.04 \pm 0.01) as a % of LIN(CD3, CD19,

CD20, CD161)⁻HLA-DR⁺ PBMCs. Scatter plot includes data from nine healthy individuals. (G) *t*-SNE analysis of flow cytometry data for LIN(CD3,CD19,CD20,CD161)⁻HLA-DR⁺CD14⁻CD16⁻ PBMCs based on the protein expression levels of AXL, SIGLEC6, CD1C, CD11C, CD22, CD33, CD34, CD45RA, CD100, CD123, CD303 and HLA-DR (see **Fig. 6** for CD100^{hi}CD34^{int.} population). Overlay of populations defined by conventional flow cytometry gating on clusters derived by *t*-SNE analysis shown in the following colors: 1=AXL⁺SIGLEC6⁺CD123⁺CD11C⁻ (pink); 2=AXL⁺SIGLEC6⁺CD123⁺CD11C^{lo} (blue); 3=AXL⁺SIGLEC6⁺CD123^{lo}CD11C⁺ (green); 4=CD1C⁺ DCs (purple) cells; 5=pDCs (dark blue); 6=CLEC9A⁺ DCs (red); 7= CD34⁺CD45RA⁻ (yellow); 8= CD34⁺CD45RA⁺ (brown); 9= CD100^{hi}CD34^{int.} (beige); 10= CD34⁻CD100^{lo} (cyan).

Figure 5. Phenotypic and functional characterization of AS DCs and ‘pure’ pDCs.

(A) Heatmap reporting scaled expression (log TPM values) of gene sets common between AS DCs (DC5) and cDCs (clusters DC1-DC4), and genes uniquely expressed in pDCs (DC6). Gene sets were generated through K-means clustering using the doKmeans function in the Seurat package. (B) Morphology of pDCs, CD1C⁺ DCs, CLEC9A⁺ DCs, AXL⁺SIGLEC6⁺CD123⁺CD11C^{-/lo} and AXL⁺SIGLEC6⁺CD123^{lo}CD11C⁺ by Giemsa-Wright stain. Scale bar=10μm. (C) IFNα (left panel) and IL-12p70 (right panel) concentration in culture supernatant 24 hours after CpG and LPS stimulation (*n*=8) or after LPS, R848 and poly (I:C) stimulation (*n*=4) of CD14⁺CD16⁻ monocytes, pDCs, CLEC9A⁺ DCs, CD1C⁺ DCs, AXL⁺SIGLEC6⁺CD123⁺CD11C^{-/lo} (1, pink), AXL⁺SIGLEC6⁺CD123^{lo}CD11C⁺ (2, blue), and CD100^{hi}CD34^{int} cells (3, beige) cells. Composite data from four to eight donors is shown (mean ±SEM, ***p*<0.01, *** *p*<0.001, Mann-Whitney U test). (D) Proliferation of allogeneic CD4⁺ and CD8⁺ T cells five days after co-culture with pDCs contaminated with AXL⁺SIGLEC6⁺ cells compared with pDCs devoid of AXL⁺SIGLEC6⁺ cells, in the context of LPS or LPS+R848 stimulation. Top panel depicts representative pseudocolor dot plot and bottom panel bar graphs of composite data (*n*=4, mean ± SEM, **p*<0.05, paired t-test). (E) Proliferation of allogeneic CD4⁺ and CD8⁺ T cells five days after co-culture with CD14⁺CD16⁻ monocytes, pDCs, CLEC9A⁺ DCs, CD1C⁺ DCs, AXL⁺SIGLEC6⁺CD123⁺CD11C^{-/lo} (1, beige), AXL⁺SIGLEC6⁺CD123^{lo}CD11C⁺ (2, blue) cells, and CD100^{hi}CD34^{int} (3, beige) cells. Top panel depicts representative pseudocolor dot plot, and bottom panel bar graphs of composite data (*n*=7, mean ±SEM, ***p*<0.01, paired t-test). (F) Top panel: immunohistochemical staining of human tonsil with AXL (brown), CD123 (purple) and CD3 (green). Brown arrows depict AXL⁺CD123⁺ cells adjacent to CD3⁺ T cells. Data shown is representative of four donors. Scale bar=50μm.

Middle panel: frequency of AXL⁺SIGLEC6⁺CD123⁺ and CD123^{lo/-} cells in human tonsil determined by flow cytometry analysis, as a percentage of CD45⁺LIN(CD3, CD19, CD20, CD56, CD161)⁻HLA-DR⁺ cells (mean \pm SEM of three donors shown; AXL⁺SIGLEC6⁺CD123⁺: 0.7% \pm 0.2%, and AXL⁺SIGLEC6⁺CD123^{lo/-}: 1.7% \pm 0.2%). Bottom panel: representative pseudocolor dot plot of AXL⁺SIGLEC6⁺CD123⁺ (pop. 1, pink) and AXL⁺SIGLEC6⁺CD123^{lo/-} (pop. 2, blue) cells in human tonsil by flow cytometry analysis ($n=3$).

Figure 6. Identification and characterization of circulating CD100^{hi}CD34^{int} cDC progenitor.

(A) Flow cytometry gating strategy to isolate DC subsets: CLEC9A⁺ DCs (red), CD1C⁺ DCs (blue), pDCs (green), AXL⁺SIGLEC6⁺ (purple), and CD123⁺CD11C⁻ cells (red gate) for differentiation assays. Data shown is a representative analysis of at least ten healthy individuals.

(B) Differentiation assays readout (flow cytometry for CLEC9A⁺ DCs, CD1C⁺ DCs and pDC, and scRNA-seq profiling of CD45⁺ cells) after seven days of co-culturing LIN(CD3, CD19, CD20, CD161)⁻HLA-DR⁺CD14⁻CD16⁻AXL⁻SIGLEC6⁻CD123⁺CD11C⁻ cells on MS5 stromal cell line supplemented with GM-CSF, SCF and FLT3LG. Top panel shows representative overlay dot plots. Overlay of pDC (green), and output cells (grey) for CD123 and CD303 expression shown in the far right (in green). Population 3 (in beige) represents CD100^{hi}CD34^{int} at day 0. Top right are composite bar graphs for CLEC9A⁺ and CD1C⁺ DCs differentiated from culture by flow cytometry analysis ($n=4$, mean \pm SEM). Heatmap in bottom panel reports scaled expression (log TPM values) signature from culture output by scRNA-seq ($n=132$), confirming differentiated CLEC9A⁺ (red) and CD1C⁺ (blue) DC transcriptional identities. Original transcriptional signatures from DC1 (CD141⁺/CLEC9A⁺ DC), DC2 (CD1C_A subset), and DC3 (CD1C_B subset) clusters are used as reference set.

(C) Top panel: flow cytometry gating strategy used to identify the CD100^{hi}CD34^{int} subset. All cell fractions in dashed-gate were tested for differentiation potential (see **Fig. S6A-F**). **(C, F)** Representative culture outputs on day seven and composite bar graphs (mean \pm SEM; $n=6$ donors). **(C)** Bottom panel – output from CD100^{hi}CD34^{int} fraction (population 3, beige gate). **(D)** Frequency of CD100^{hi}CD34^{int} subset as of LIN(CD3, CD19, CD20, CD161)⁻HLA-DR⁺ PBMCs ($n=9$ healthy donors). Morphology of CD100^{hi}CD34^{int} cell by Giemsa-Wright stain. Scale bar=10 μ m. **(E)** Proliferative capacity of peripheral blood Cell Trace Violet (CTV)-labeled CD34⁺ HSCs (purple), CD100^{hi}CD34^{int}.

(beige), AXL⁺SIGLEC6⁺CD123⁺CD11C^{-/lo} (pink), and AXL⁺SIGLEC6⁺CD123^{lo}CD11C⁺ (blue), as measured by CTV dilution after five days in culture on MS5 stromal cell line supplemented with GM-CSF, SCF and FLT3LG. Left panel shows representative overlay histogram; right panel shows composite bar graphs illustrating % of proliferated cells and number of proliferations undergone from three donors shown (**p*<0.05, paired t-test). (F) Output from differentiation assays seeded with CLEC9A⁺ DCs, CD1C⁺ DCs, pDCs, and AXL⁺SIGLEC6⁺ cells isolated using gating strategy in panel A. AXL⁺SIGLEC6⁺x2 = double FLT3LG concentration. (G) PCA analysis incorporating monocytes (*n*=339), DCs (*n*=742), and 4 BPDCN patient samples (*n*=174) using the R software package Seurat. PC1 vs. PC2 demonstrates the close transcriptional proximity between all 4 BPDCN samples and pDCs (dashed black circle), with overlapping cells (black bracket). PC1 and PC2 variance is 3.8%. Each dot represents an individual cell and colored legend for each subset is shown on the right.