See P, Dutertre C-A, Chen J, Gunther P, McGovern N, Irac SE, Gunawan M, Beyer M, Handler K, Duan K, Sumatoh HRB, Ruffin N, Jouve M, Gea-Mallorqui E, Hennekam RCM, Lim T, Yip CC, Wen M, Malleret B, Low I, Shadan NB, Fen CFS, Tay A, Lum J, Zolezzi F, Larbi A, Poidinger M, Chan JKY, Chen Q, Renia L, Haniffa M, Benaroch P, Schlitzer A, Schultze JL, Newell EW, Ginhoux F. **Mapping the human DC lineage through the integration of high-dimensional techniques**. *Science* 2017, 356(6339), 1-23.

DOI link to article:

Date deposited:

07/07/2017

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

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

Affiliations:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

# These authors contributed equally to this work.

* Correspondence to Florent Ginhoux (Florent_Ginhoux@immunol.a-star.edu.sg)

Abstract (125 max)

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

One Sentence Summary: Mapping the human dendritic cell lineage

3

Introduction (4500 max including ref, notes and caption)

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

Human blood and tissue DC, and their precursors in the BM, are identified by expression of CD135 and HLA-DR (*12-14*). Hence, we interrogated the CD135$^+$HLA-DR$^+$ fraction of human blood using several integrated high-dimensional analysis techniques (single-cell mRNA sequencing or scmRNAseq; microfluidic single cell mRNA sequencing and mass cytometry using Cytometry by Time-of-Flight mass spectrometry or CyTOF). These approaches supersede

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

Results

Unbiased identification of DC precursors by non-supervised single cell RNAseq and CyTOF

Using PBMC isolated from human blood, we first employed massively-parallel single cell mRNA sequencing (MARS-seq) (*15*) to transcriptionally profile 710 individual cells within the lineage- (CD3/CD14/CD16/CD20/CD34) negative, HLA-DR$^+$CD135$^+$ population (Fig. 1A-E; see Fig. S1A for sorting strategy, Fig. S1B-F for workflow & quality control and Table S3 for gene expression data). MARS-seq data were processed using non-linear dimensionality reduction via t-stochastic neighbor embedding (tSNE), which enables unbiased visualization of high-dimensional similarities of cells in a two-dimensional map (*16-18*). Density-based spatial clustering of applications with noise (DBSCAN) on the tSNE dimensions revealed 5 distinct clusters of transcriptionally-related cells within the selected PBMC population (Fig. 1A, S1G). To define the nature of these clusters, we generated gene signature scores for pDC, cDC1 and cDC2, and overlaid the expression of these signatures by each cell onto the tSNE visualization. Clusters #1 and #2 were identified as pDC, cluster 3 as cDC1, and cluster 5 as cDC2. The last cluster (#4) laid in between the cDC1 (#3) and cDC2 (#5) clusters and possessed a weak mixed pDC/cDC signature (Fig. 1A). We then applied connectivity MAP (cMAP) analysis (*19*), which calculates the degree of enrichment for transcripts of pDC or cDC signature genes in each

individual cell; this approach confirmed the signatures of pDC (#1 and #2) and cDC (#3 and #5) clusters, and showed that most cells in cluster #4 expressed a cDC signature (Fig. 1B). The Mpath algorithm (*20*) was then applied to the 5 clusters defined in the tSNE analysis to identify hypothetical developmental relationships based on these transcriptional similarities between cells (Fig. 1C). Mpath revealed that the 5 clusters were grouped into 3 distinct branches with one central cluster (cluster #4) at the intersection of the 3 branches (Fig. 1C, S2A). To confirm these findings, we employed the Monocle algorithm (*21*) and principal component analyses (PCA), which similarly resolved the cells into the same 3 branches with the same subset identities, and again with the cells from tSNE cluster #4 falling at their intersection (Fig. 1D, S2B). While cells from pDC clusters (#1 and #2) had a higher expression of pDC-specific markers and transcription factors (TF) then the cDC clusters (#3 and #5) and central cluster #4, cells in cluster #4 expressed higher levels of markers and TF previously-associated with all cDC lineage (Fig. 1E).. This phenotype led us to hypothesize that cluster #4 represented a population of putative intermediate uncommitted DC precursors.

To further understand the delineated sub-populations, we employed CyTOF, which simultaneously measures the intensity of expression of up to 38 different molecules at the single cell level. We designed a panel of 38 antibodies (Table S2) recognizing DC lineage/progenitor-associated surface molecules (Fig. 1F-H, S3), or molecules such as CD2, CX3CR1, CD11c and HLA-DR identified by MARS-seq as expressed by cluster #4 at the mRNA level (Fig. 1E). Using the tSNE algorithm, the CD45$^+$Lineage$^-$(CD7/CD14/CD15/CD16/CD19/CD34)HLA-DR$^+$ PBMC fraction (see Fig. S3A for gating strategy) resolved into 3 distinct clusters representing cDC1, cDC2 and pDC (Fig. 1F), with an intermediate cluster at the intersection of the cDC and pDC clusters that expressed both cDC- (CD11c/CX3CR1/CD2/CD33/CD141/BTLA) and pDC- (CD45RA/CD123/CD303) associated markers (Fig. 1G-H, S3B), corresponding to the MARS-seq cluster #4. The delineation of these clusters was confirmed by phenograph unsupervised

clustering (Fig. S3C). The location of this cluster of CD123[+]CD33[+] cells suggested again a distinct and intermediate nature; cells here showed the highest expression of CD5, CD327, CD85j, along with high levels of HLA-DR and the cDC-associated molecule CD86 (Fig. 1H-I). Taken together, these characteristics led us to ask whether CD123[+]CD33[+] cells might represent circulating human pre-DC.

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

We further analyzed the CD123[+]CD33[+] cell cluster within the Lin[-]HLA-DR[+] fraction of PBMC by flow cytometry, where we identified CD123[+]CD33[-] pDC (blue), CD45RA[+/-]CD123[-] cDC1 (red) and cDC2 (beige), as well as CD33[+]CD45RA[+]CD123[+] putative pre-DC (purple) (Fig. 2A, S4A). Putative pre-DC expressed CX3CR1, CD2, CD303 and CD304, with low CD11c expression, while CD123[+]CD33[-] pDC exhibited variable CD2 expression (Fig. 2A-B, S4B-C). Extending our analysis to immune cells from the spleen, we identified a similar putative pre-DC population, which was more abundant than in blood and expressed higher levels of CD11c (Fig. 2A, 2C, S4D). Both putative pre-DC populations in blood and spleen also expressed CD135 and intermediate levels of CD141 (Fig. S4C). Wright-Giemsa staining of putative pre-DC sorted from blood revealed an indented nuclear pattern reminiscent of classical cDC, accompanied by a perinuclear area of clearing and basophilic cytoplasm reminiscent of pDC (Fig. 2D). At the ultra-structural level, putative pre-DC and pDC exhibited distinct features, despite their morphological similarities (Fig. 2E, S4E): where putative pre-DC possessed an indented nuclear pattern, a large nucleus and lesser cytosol, pDC contained a smaller nucleus, abundant cytosol and numerous stacks of rough ER membranes decorated with ribosomes, suggesting a developed secretory apparatus in agreement with published data (*22*).

We then compared the differentiation capacity of pre-DC to that of cDC and pDC through stromal culture with FLT3L, GM-CSF and SCF, as described (*8*). After 5 days, while pDC, cDC1 and cDC2 populations predominantly remained in their initial states, the putative

pre-DC population differentiated into cDC1 and cDC2 in the proportions found *in vivo* (*14*, *23-25*) (Fig. 2F, S4F, S5). Altogether, these data suggested that CD123$^+$CD33$^+$CD45RA$^+$CX3CR1$^+$CD2$^+$ cells are circulating pre-DC with cDC differentiation potential. Recently, Breton and colleagues reported a minor population of human pre-DC (*9*), highlighted in Fig. S6A (blue), sharing a similar phenotype with the Lin$^-$CD123$^+$CD33$^+$CD45RA$^+$ pre-DC (pink) defined here (Fig. S6A-B). However, our approach extends their findings and reveals that the pre-DC population in blood and spleen is significantly larger than that identified within the minor CD303$^-$CD141$^-$CD117$^+$ fraction considered previously (Fig. S6C-D).

Pre-DC are functionally distinct from pDC

IFNα-secreting pDC can differentiate into cells resembling cDC when cultured with IL-3 and CD40L (*10*, *11*), and were considered DC precursors (*11*). However, when we used traditional ILT3$^+$ILT1$^-$ (*10*) or CD4$^+$CD11c$^-$ (*11*) pDC gating strategies, we detected a "contaminating" CD123$^+$CD33$^+$CD45RA$^+$ pre-DC sub-population within both groups that were previously considered pure pDC (Fig. S6E-F). Therefore, we questioned whether other properties of traditionally-classified "pDC populations" might be attributed to pre-DC. Stimulation of either Toll-like receptor (TLR)7/8 (CL097) or TLR9 (CpG ODN 2216) resulted in abundant secretion of IFNα but not IL-12p40 by pure cultures of pDC, whereas pre-DC readily secreted IL-12p40 but not IFNα (Fig. 2G, S7). Furthermore, while pDC were thought to induce proliferation of naïve CD4$^+$ T cells (*10*, *26*), this capacity too lies solely within the pre-DC sub-population (Fig. 2H). Therefore reports of potent allostimulatory capacity and IL-12p40 production by CD2$^+$ pDC (*26*) might also be explained by CD2$^+$ pre-DC contamination (*27*) (Fig. S6G). Pitt-hopkins syndrome (PHS), characterized by abnormal craniofacial and neural development, severe mental retardation, and motor dysfunction, is caused by an haplo-insufficiency in the *TCF4* gene which

encodes the E2-2 transcription factor, a central regulator of pDC development (*28*). We confirmed that PHS patients had a strong reduction in their blood pDC numbers and found that they retained a population of pre-DC (Fig. 2I, S4G), likely accounting for the previously ill-understood CD45RA$^+$CD123$^+$CD303^{lo} population detected in these patients (*29*). Taken together, our data indicate that, while pre-DC and pDC share some phenotypic features, they can be separated by differential expression of several markers including CD33, CX3CR1, CD2, CD5 and CD327: pDC are the *bona fide* IFNα-producing cells, but the reported IL-12 production and CD4$^+$ T cell allostimulatory capacity of pDC can likely be attributed to "contaminating" pre-DC, which can give rise to both cDC1 and cDC2.

Identification and characterization of committed pre-DC subsets

The murine pre-DC population contains both uncommitted and committed pre-cDC1 and pre-cDC2 precursors (*7*). We asked whether the same was true for human blood pre-DC using microfluidic scmRNAseq (see Fig. S8A for sorting strategy, Fig. S8B-C for workflow & quality control, Table S4 for gene expression data). The additional gene expression data relative to the MARS-seq strategy used for Fig. 1A-F (2.5 million reads/cell *vs* 60,000 reads/cell, respectively) was subjected to cMAP analysis, which calculated the degree of enrichment for cDC1 or cDC2 signature gene transcripts for each single cell (Fig. 3A). Among the 92 pre-DC analyzed, 19 and 20 cells exhibited enrichment for cDC1 and cDC2 gene expression signatures, respectively, and 53 cells showed no transcriptional similarity with either cDC subset. Further Mpath analysis showed that these "unprimed" pre-DC were developmentally related to cDC1- and cDC2-primed pre-DC, and thus their patterns of gene expression fell between the cDC1 and cDC2 signature scores by cMAP (Fig. 3B, S9). These data suggest that the human pre-DC population contains cells exhibiting transcriptomic priming towards cDC1 and cDC2 lineages, as seen previously in mice (*7*).

We next asked whether we could identify this evident heterogeneity within the pre-DC population by flow cytometry, using either pre-DC-specific markers (CD45RA, CD327, CD5) or those expressed relatively higher by pre-DC compared to cDC2 (BTLA, CD141). 3D-PCA analysis of the Lin$^-$HLA-DR$^+$CD33$^+$ population (containing both differentiated cDC and pre-DC) revealed 3 major cell clusters: CADM1$^+$ cDC1 (red region), CD1c$^+$ cDC2 (beige region) and CD123$^+$ pre-DC (purple region) (Fig. 3C, S10A). Interestingly, while cells located at the intersection of these 3 clusters (black dashed oval region in Fig. 3D) expressed lower levels of CD123 than pre-DC (though higher levels than differentiated cDC, Fig. 3C), they also expressed high levels of pre-DC markers (Fig. 3D, S10A). We reasoned that these CD45RA$^+$CD123^{lo} cells might be committed pre-DC differentiating into either cDC1 or cDC2 (Fig. 3E, where cells falling in regions defined in Fig. 3C-D are overlaid). The wanderlust algorithm (*30*), which orders cells according to their most immature to mature state into a constructed trajectory, confirmed the developmental relationship between pre-DC (early events, dark), CD45RA$^+$CD123^{lo} cells (intermediate events, grey) and mature cDC (clear) (Fig. 3F). Flow cytometry of PBMC allowed the identification of CD123$^+$CADM1$^-$CD1c$^-$ putative uncommitted pre-DC, alongside putative CADM1$^+$CD1c$^-$ pre-cDC1 and CADM1$^-$CD1c$^+$ pre-cDC2 within the remaining CD45RA$^+$ cells (Fig. 3G, S10B). These 3 populations were also present, and more abundant, in the spleen (Fig. S10C). Importantly, *in vitro* culture of pre-DC subsets sorted from PBMC did not give rise to any CD303$^+$ cells (which would be either undifferentiated pre-DC or differentiated pDC), while instead early pre-DC gave rise to both cDC subsets, and pre-cDC1 and pre-cDC2 differentiated exclusively into cDC1 and cDC2 subsets, respectively (Fig. 3H, S10D, S11).

Scanning electron microscopy confirmed that early pre-DC are larger and rougher in appearance than pDC, and revealed that committed pre-DC subsets closely resembled their mature cDC counterparts (Fig. 3I, S10E). Flow cytometry phenotyping of blood pre-DC (Fig.

S10F) identified patterns of transitional marker <mark>expression characterized by progressive evolution throughout the development of early pre-DC towards pre-cDC1/2 and differentiated cDC1/2: CD45RO and CD33 were acquired in parallel with the loss of CD45RA; CD5, CD123, CD304 and CD327 were expressed abundantly by early pre-DC, intermediately by pre-cDC1 and pre-cDC2, and rarely if at all by mature cDC and pDC; FcεRI and CD1c were acquired as early pre-DC commit towards the cDC2 lineage, concurrent with the loss of BTLA and CD319 expression; early pre-DC had an intermediate expression of CD141 which dropped along cDC2 differentiation but was increasingly expressed during commitment towards cDC1, with few pre-cDC1 already starting to express Clec9A;</mark> and IRF8 and IRF4, transcription factors regulating cDC lineage development (*2*, *3*), were expressed by early pre-DC and pre-cDC1, while pre-cDC2 maintained only IRF4 expression (Fig. S10G).

We next sorted pre-DC and DC subsets from blood and performed microarray analyses to define their entire transcriptome. 3D-PCA of microarray data showed that pDC were clearly separated from other pre-DC and DC subsets along the PC1 axis (Fig. 4A, S12). The PC2 axis indicated that cDC2 and pre-cDC2 exhibited similar transcriptomes, while pre-cDC1 occupied a position between early pre-DC and cDC1. Hierarchical clustering of differentially-expressed genes (DEG) confirmed the similarities between committed pre-DC and their corresponding mature subset (Fig. S13). The greatest number of DEG was found between early pre-DC and pDC (1249 genes) among which CD86, CD2, CD22, CD5, ITGAX (CD11c), CD33, CLEC10A, SIGLEC6 (CD327), THBD, CLEC12A, KLF4 and ZBTB46 were expressed by early pre-DC, while pDC showed expression of CD68, CLEC4C, TCF4, PACSIN1, IRF7 and TLR7 (Fig. 4B). An evolution of gene expression pattern was evident from early pre-DC, to pre-cDC1 and then cDC1 (Fig. 4C-D), while pre-cDC2 were similar to cDC2 (Fig. 4D, S13). Unsupervised gene expression analysis of all cDC and pre-DC subsets identified 62 <mark>commonly expressed genes</mark> including the transcription factors BATF3, ID2 and TCF4 (E2-2), and the pre-DC markers

CLEC4C (CD303), SIGLEC6 (CD327), and IL3RA (CD123) (Fig. 4E). During early pre-DC to cDC differentiation, the progressive reduction in transcript abundance of SIGLEC6 (CD327), CD22 and AXL is mirrored at the protein level (Fig. 4F). Key transcription factors involved in the differentiation and/or maturation of DC subsets showed a progressive change in their expression along the differentiation path from pre-DC to mature cDC (Fig. 4G). Finally, pathway analyses revealed that pre-DC exhibit an enrichment of cDC functions relative to pDC, and a relatively immature state compared to mature cDC (Fig. S14).

Committed pre-DC subsets are functional

We then asked to what extent the functional specializations of DC (*1*, *31*) were acquired at the precursor level. We stimulated PBMC with TLR agonists and measured their cytokine production (Fig. 5A). Pre-DC produced markedly more TNFα and IL-12p40 following exposure to CL097 (TLR7/8 agonist) or CpG ODN 2216 (TLR9 agonist), than to LPS (TLR4 agonist) or polyI:C (TLR3 agonist). We confirmed that pDC were uniquely capable of robust IFNα production in response to CL097 and CpG ODN 2216. CpG ODN 2216 stimulation also triggered IL-12p40 and TNFα production by early pre-DC, pre-cDC1, and to a lesser extent pre-cDC2. Although TLR9 transcripts were detected only in early pre-DC (Fig. S15A), these data indicate that, contrary to differentiated cDC1 and cDC2, pre-cDC1 and pre-cDC2 do express functional TLR9 protein. Interestingly, while pre-cDC2 resembled cDC2 at the gene expression level, their TLR-responsiveness was intermediate between that of early pre-DC and cDC2. Pre-DC subsets also expressed T cell costimulatory molecules (Fig. 5B) and induced proliferation and polarization of naïve CD4+ T cells to a similar level as mature cDC (Fig. 5C, S15B).

Unsupervised mapping of DC ontogeny using scmRNAseq and CyTOF

We performed an unsupervised isoMAP analysis of human BM cells, obtained from CyTOF analysis, focusing on the Lin$^-$CD123^{hi} fraction (Fig. 6A, S16A). Such analysis automatically identified CD123^{hi}CD34$^+$ CDP (phenograph cluster #5), from which branched CD34$^-$CD123$^+$CD327$^+$CD33$^+$ pre-DC (clusters #1 and #2) or CD34$^-$CD123$^+$CD303$^+$CD68$^+$ pDC (clusters #3 and #4), both progressively acquiring their respective phenotypes; cells in the pre-DC branch increasingly expressed CD2, CD11c, CD116 and, at a later stage, CD1c. In the peripheral blood, two parallel lineages, corresponding to either pre-DC or pDC were also evident in the isoMAP analysis of lineage$^-$CD123$^+$ cells, in which a CDP population was not detected (Fig. 6C). IsoMAP and phenograph analysis of pre-DC extracted from the isoMAP analysis of Fig. 6A (BM, clusters #1 and #2) and Fig. 6B (blood, cluster #6) revealed the 3 distinct pre-DC subsets (Fig. 6C as defined by their unique marker expression patterns (Fig. S16B-C). In summary, we were able to trace the developmental stages of DC from the BM to the peripheral blood through CyTOF. The CDP population in the BM bifurcates into two pathways, developing into either the pre-DC or pDC found in blood (Fig. 6A-C): the pre-DC population is heterogeneous and exists as distinct subsets detectable in both blood and BM (Fig. 6C, S16B-C). Alongside, we uncovered an intriguing heterogeneity in blood and BM pDC that warrants further investigation (Fig. 6C, S16D-E).

Discussion

Using unsupervised single-cell RNAseq and CyTOF analyses, we unraveled the complexity of the human DC lineage at the single cell level, revealing a continuous process of differentiation that starts in the BM with CDP, and diverges at the point of emergence of pre-DC and pDC potentials, with subsequent maturation of both lineages in the blood. A previous study using traditional surface marker-based approaches had suggested the presence of a minor pre-DC population in PBMC (9), but the combination of high-dimensional techniques and unbiased

analysis employed here revealed that this minor population had been significantly and importantly underestimated. Our data extended the published findings and revealed the existence of a population of pre-DC that overlapped with those observed by Breton and colleagues within the $CD117^+CD303^-CD141^-$ fraction of PBMC. However, we also showed that the full pre-DC population includes more than ten-fold the number of cells in peripheral blood than was originally estimated, and is considerably more diverse.

Recent work in mice found uncommitted and subset-committed pre-DC subsets in BM (*7*, *32*); similarly, we identified 3 functionally- and phenotypically- distinct pre-DC populations in human PBMC, spleen and BM: uncommitted pre-DC and 2 populations of subset-committed pre-DC. In line with the concept of continuous differentiation from the BM to the blood, the proportion of uncommitted cells was higher in the BM pre-DC population than in that of the blood. Altogether, these findings support a two-step model of DC development in which a central transcriptomic subset-specific program is imprinted on DC precursors from the CDP stage onwards, conferring a core subset identity irrespective of the final tissue destination; then in the second step, peripheral tissue-dependent programming is added to ensure site-specific functionality and adaptation (*7*, *32*). Future studies will be required to reveal the molecular events underlying DC subset lineage priming, the tissue-specific cues that regulate peripheral programming as well as to design strategies that specifically target DC subsets at the precursor level. In addition, how the proportions of uncommitted pre-DC versus committed pre-DC are modified in acute and chronic inflammatory settings warrants further investigation.

An important aspect of unbiased analysis is that cells are not excluded from consideration on the basis of preconceptions concerning their surface phenotype. In this case, we found that pre-DC express most of the markers classically used to define pDC, such as CD123, CD303 and CD304. Thus the phenotypic strategy used to identify and isolate pDC in many functional studies

has led to the inadvertent inclusion of CD123[+]CD33[+] pre-DC within the pDC population. While this calls us to urgently reconsider some aspects of "pDC population" biology, it may also explain earlier findings, including reports suggesting that: pDC cultures possess cDC potential , acquiring cDC-like morphology (*10*, *11*), also recently observed in murine BM pDC (*33*); pDC can mediate Th1 immunity through production of both IFNα and IL-12 (*10*, *34-38*); pDC exhibit naïve T-cell allostimulatory capacity (*26*, *36*); and that pDC express costimulatory molecules and exhibit antigen-presentation/cross-presentation capabilities at the expense of IFNα secretion (*34*, *39*). These observations could well be attributed to the undetected pre-DC in the pDC populations described by these studies; indeed it was speculated that the IL-12 production observed in these early studies might be due to the presence of contaminating CD11c[+] cDC (*40*). Seeking clarification, we separated CX3CR1[+]CD33[+]CD123[+]CD303[+]CD304[+] pre-DC from CX3CR1[-]CD33[-]CD123[+]CD303[+]CD304[+] "pure" pDC and showed that pDC could not polarize nor induce proliferation of naïve CD4 T cells, while pre-DC were able to do so: pDC were unable to produce IL-12, unlike pre-DC, but were the only cells capable of strongly producing IFNα in response to TRL7/8/9 agonists, as initially described (*41*). Thus, it is of paramount importance that pre-DC be excluded from pDC populations in future studies, particularly when using commercial pDC isolation kits. Finally, if pDC are now stripped of all their cDC properties, it raises the question of whether they truly belong to the DC lineage, or rather are a distinct type of innate IFN-I-producing lymphoid cell. Importantly, it also remains to be shown whether the BM CD34[+]CD123[hi] CDP population is also a mixture of independent *bona fide* cDC progenitors and pDC progenitors.

Despite their classification as precursors, human pre-DC appear functional in their own right, being equipped with some T-cell co-stimulatory molecules, and with a strong capacity for naïve T cell allostimulation and cytokine secretion in response to TLR stimulation (Fig. 2, Fig. 5 and S7, S14). Pre-DC produced low levels of IFN-α in response to CpG ODN 2216 exposure,

and secreted IL-12 and TNF-α in response to various TLR ligands. Hence, it is reasonable to propose that pre-DC have the potential to play a role in the context of disease, particularly in inflammatory and autoimmune diseases where dysregulation of their differentiation continuum or their arrested development could render them a potent source of inflammatory DC ready for rapid recruitment and mobilization.

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

Figure legends:

Fig. 1. MARS-seq and CyTOF® identifies rare CD123⁺CD33⁺ putative pre-DC. (**A-E**) 710 sorted Lin(CD3/CD14/CD16/CD20/CD34)⁻HLA-DR⁺CD135⁺ PBMC (see Figure S1A for gating strategy) were subjected to MARS-seq. (**A**) tSNE plot of the 710 cells, colored by (left panel) clusters identified by tSNE + Seurat clustering (clusters labeled #1-#5), or colored by the relative signature score for (from left to right) pDC, cDC1 and CD2. (**B**) cMAP analysis showing the degree of enrichment for transcripts of pDC or cDC signature genes for the 5 tSNE/Seurat clusters. (**C**) Mpath analysis applied to the 5 tSNE/Seurat clusters defining their developmental relationship. (**D**) Principal component analysis (PCA) representation of the 710 cells highlighting the tSNE/Seurat clusters identified in (**A**). (**E**) Violin plots showing the expression of pDC and cDC signature genes for tSNE/Seurat pDC clusters (#1+#2, blue), the central cluster #4 (purple) and cDC clusters (#3+#5). Displayed genes were selected for their differential expression between the tSNE/Seurat clusters #4 (purple) and pDC (clusters #1+#2). Adjusted p values were calculated with the Kruskal-Wallis test followed by the Dunn's multiple comparisons test. (**F, G**) tSNE plots of CyTOF data from CD45⁺Lin(CD7/CD14/CD15/CD16/CD19/CD34)⁻HLA-DR⁺ PBMC, (**F**) gates defining CD123⁺CD33⁺ cells and DC subsets, and (**G**) relative expression level of selected markers. (**H**) Subsets defined in **Fig. 1F** were overlaid onto 2D contour plots to for phenotypic comparison.

Fig. 2. **Characterization of human pre-DC.** (**A**) Identification of pre-DC (purple), pDC (blue), cDC1 (red) and cDC2 (beige) within PBMC and spleen. (**B**) Expression of CD303, CD304, CD123 and CD11c by pre-DC and DC subsets. (**C**) Relative numbers of pre-DC in peripheral blood (n=6) and spleen (n=3). (**D**) Wright-Giemsa staining of blood pre-DC (left panels) and

pDC (right panels). (**E**) Representative electron micrographs showing morphological characteristics of pre-DC and pDC; boxed areas are shown at higher magnification below. Putative pre-DC exhibited thinner cytoplasm with homogeneously-distributed mitochondria (m) and relatively less rough endoplasmic reticulum (RER, arrowheads) compared to pDC. pDC exhibited an abundant cytoplasm, packed mitochondria and a well-developed cortical RER, organized in parallel cisterna at a pole of the cell (note the presence of the centriole (C) and microtubules (small arrows) near RER cisterna). (**F**) DC subset and pre-DC populations co-cultured for 5 days with MS-5 feeder cells supplemented with FLT3L, GM-CSF and SCF were investigated for the presence of CD123$^+$CD172α$^-$ pDC (blue), Clec9A$^+$CADM1$^+$ cDC1 (red) or CD172α$^+$CD1c$^+$ cDC2 (beige). (**G**) Detection of intracellular cytokine production by DC subsets and pre-DC following TLR stimulation. 2-D plots of IFNα and IL-12p40 production by pDC (blue) and pre-DC (purple) are shown, alongside mean relative frequency of IFNα$^+$-, IL-12p40$^+$- and TNFα$^+$- producing cells within pre-DC and DC subsets following exposure to LPS, LPS+IFNγ (L+I), polyI:C (pI:C), CL097 (CL) or CpG ODN2216 (CpG) (n=4). (**H**) Proliferation of naïve CD4$^+$ T cells after 6 days of culture with allogenic CD123$^+$CD33$^+$ pre-DC or CD123$^+$CD33$^-$ pDC (n=2). (**I**) Graphical representation of the frequency of pDC (upper panel) and pre-DC (lower panel) within CD45$^+$ PBMC from control (Ctrl, n=11) and Pitt-Hopkins patients (PH, n=4). P-values calculated by Mann-Whitney test.

Fig. 3. Identification of committed human pre-DC subsets. (**A-B**) Single-cell mRNAseq of 92 sorted Lin(CD3/14/16/19/20)$^-$HLA-DR$^+$CD33$^+$CD123$^+$ cells (see Fig. S9A (not S9A – couldn't see it) for the sort gating strategy). (**A**) cMAP enrichment score of cells for cDC1- or cDC2-specific signatures. (**B**) Mpath analysis showing the development relationship between cells defined as "unprimed" (central grey cluster) or cDC1- (green) and cDC2- (orange) primed in the cMAP analysis in (**A**). (**C-E**) 3D-PCA of Lin$^-$HLA-DR$^+$CD33$^+$ PBMC, analyzed by flow cytometry. (**C**) Dot plot of PCA components 1-3 revealed 3 major cell clusters corresponding to

pre-DC (purple circle), CADM1[+] cDC1 (red circle) and CD1c[+] cDC2 (beige circle). Relative expression of CADM1, CD1c and CD123 is displayed. (**D**) Magnification of the intersect between pre-DC, cDC1 and cDC2 in the 3D-PCA plot showing the relative expression of CD45RA, BTLA, CD327, CD141 and CD5. Dashed black circle delineates the intermediate CD45RA[+] population. (**E**) CD45RA/CD123 dot plots showing overlaid cell subsets defined in the 3D-PCA plot (far left panel) with the relative expression of BTLA, CD327, CD141 and CD5. <mark>(**F**) The Wanderlust dimension (early to dark events progressively switching from dark to clear) was overlaid onto the 3D-PCA and the CD45RA/CD123 dot plots of panels C-E.</mark> (**G**) Sequential gating of flow cytometric data starting from Live CD45[+]Lin(CD3/14/16/19/20)[-]CD34[-]HLA-DR[+] PBMC to definition of putative pre-DC subsets among CD33[+]CD45RA[+] cDC. (**H**) Pre-DC subsets co-cultured for 5 days on MS-5 feeder cells supplemented with FLT3L, GM-CSF and SCF were analyzed for their capacity to differentiate into Clec9A[+]CADM1[+] cDC1 (red or CD1c[+]CD11c[+] cDC2 (beige). (**I**) Morphology of sorted pre-DC and DC subsets visualized by SEM.

Fig. 4. DC and pre-DC subset gene expression analysis. Microarray data from sorted DC and pre-DC subsets in Fig. 3 were analyzed by (**A**) 3D-PCA using probe-filtered genes (upper panel) or DEG (lower panel). For PCA dimension, the loading value was indicated as relative number. (**B-D**) Heat maps of DEG between <mark>(**B**) early pre-DC/pDC</mark> (**C**) early pre-DC/pre-cDC1/cDC1 and (**D**) early pre-DC/pre-cDC2/cDC2. (**E**) Profile analysis of all common DEGs between early pre-DC, pre-cDC1, cDC1, pre-cDC2 and cDC2. (**F**) Expression level of CD327, CD22 and AXL proteins by DC and pre-DC subsets evaluated by flow cytometry. MFI are indicated. (**G**) Profile analysis of selected transcription factor-coding genes.

Fig. 5. Functional analysis of DC and pre-DC subsets. (**A**) Frequency of cytokine production by pre-DC and DC subsets upon TLR stimulation was measured by intracellular flow cytometry. Left panel shows dot plots of IFNα, IL-12p40 and TNFα production by pDC (blue), early pre-DC (purple), pre-cDC2 (orange), cDC2 (beige), pre-cDC1 (green) and cDC1 (red); right panels show mean relative numbers of pre-DC and DC subset cells producing IFNα^+, IL-12p40$^+$ or TNFα^+ in response to LPS, LPS+IFNγ (L+I), pI:C, CL097 (CL) or CpG ODN2216 (CpG) (n=4). (**B**) Expression level (MFI) of costimulatory molecules (CD40, CD80, CD83, CD86) by blood pre-DC and DC subsets (n=4). (**C**) Proliferation of naïve CD4$^+$ T cells after 6 days of culture with allogenic pre-DC and DC subsets (n=3).

Fig. 6. Unsupervised mapping of DC ontogeny using CyTOF. CyTOF data from (A**, C**) BM and (**B, C**) blood (PBMC) were analyzed using isoMAP dimensionality reduction to compare overall phenotypic relatedness of cell populations, and were automatically subdivided into clusters using the phenograph algorithm. (**A, B**) IsoMAP1-2 plots showing expression level of CDP-, pDC- and pre-DC/cDC-specific markers within (**A**) BM, and (**B**) blood Lin(CD3/CD7/CD14/CD15/CD19/CD34)$^-$HLA-DR$^+$CD123$^+$ cells. (**C**) Phenotypic relationship of Lin-HLA-DR$^+$, CD123^{hi} BM marrow and CD123$^+$ PBMC, showing progression from CDP (light blue) towards pDC (blue) or pre-DC (pink) in the BM, and the clear separation of pDC (blue) and pre-DC (pink) in the blood. Both in the BM (upper panels) and in the blood (lower panels), cells within the pre-DC phenograph clusters (#1 and #2 in the BM, and #6 in the blood) and cells within the pDC phenograph clusters (#3 and #4 in the BM, and #7 in the blood) were further analyzed by isoMAP to define pre-DC subsets (left panels, and Fig. S16C, D) and heterogeneity among pDC (right panels, and Fig. S16D, E).

REFERENCES

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

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

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

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

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

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

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

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

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

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

11.   G. Grouard *et al.*, The enigmatic plasmacytoid T cells develop into dendritic cells with interleukin (IL)-3 and CD40-ligand. *Journal of Experimental Medicine*. **185**, 1101–1111 (1997).

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

13.   J. D. Griffin, K. D. Sabbath, F. Herrmann, Differential expression of HLA-DR antigens in subsets of human CFU-GM. ... (1985).

14.   A. Dzionek, A. Fuchs, P. Schmidt, S. Cremer, BDCA-2, BDCA-3, and BDCA-4: three markers for distinct subsets of dendritic cells in human peripheral blood. *The Journal of Immunology*. **165**, 6037–6046 (2000).

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

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

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

18.   L. Van der Maaten, G. Hinton, Visualizing data using t-SNE. *Journal of Machine Learning ...* (2008).

19.   J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. **313**, 1929–1935 (2006).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

37. T. Ito *et al.*, Plasmacytoid dendritic cells prime IL-10-producing T regulatory cells by inducible costimulator ligand. *Journal of Experimental Medicine*. **204**, 105–115 (2007).

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

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

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

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