**Supplemental Information**

Proteomics of Human Dendritic Cell Subsets

Reveals Subset-Specific Surface Markers

and Differential Inflammasome Function

Kuntal Worah, Till S.M. Mathan, Thien Phong Vu Manh, Shivakumar Keerthikumar, Gerty Schreibelt, Jurjen Tel, Tjitske Duiveman-de Boer, Annette E. Sköld, Annemiek B. van Spriel, I. Jolanda M. de Vries, Martijn A. Huynen, Hans J. Wessels, Jolein Gloerich, Marc Dalod, Edwin Lasonder, Carl G. Figdor, and Sonja I. Buschow

Supplemental data
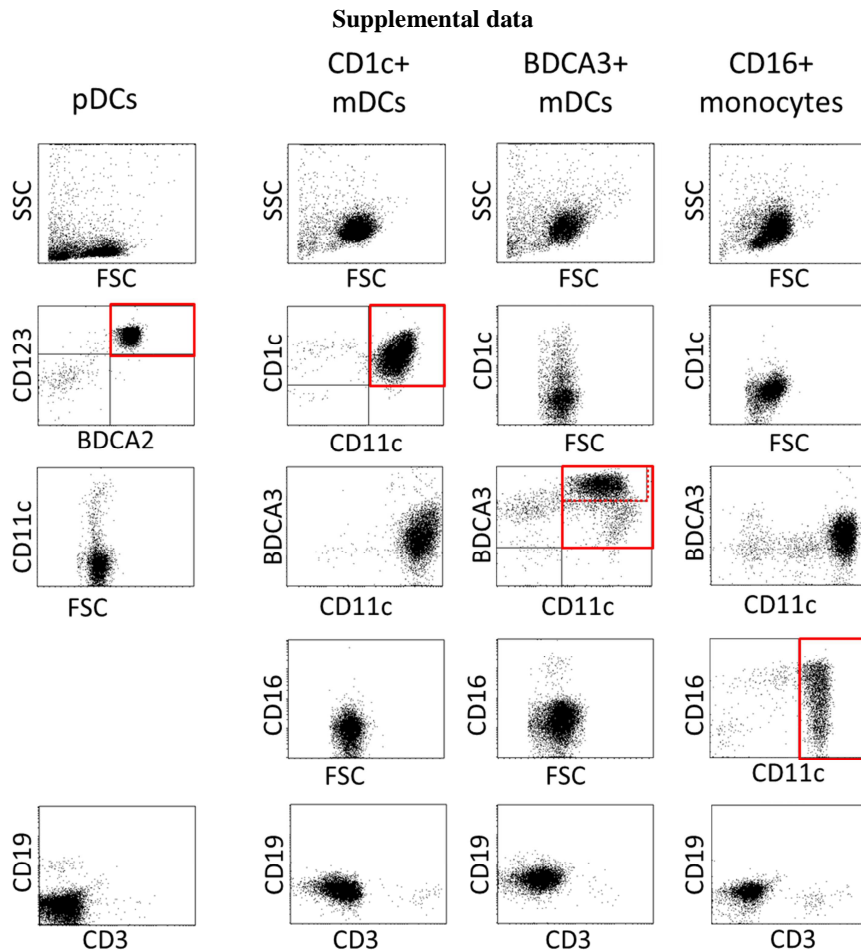
	pDC	CD1c+ mDC	BDCA3+mDC (BDCA3hi mDC)	CD16+ Monocytes
Donor 1	92%	95%	95% (45%)	98%
Donor 2	95%	98%	96% (83%)	96%
Donor 3	95%	78%	88% (49%)	82%

Figure S1 related to figures 1 and 3: Assessment of DC purity by flow cytometry: Flow cytometry results for magnetic bead purified subset samples. Displayed for one donor are the side-scatter (SSC) and forward-scatter (FSC; upper panel) and labelling for indicated subset identification markers (middle panels) and B (CD19) and T (CD3) cell markers (lower panels). For each subset the sorting gate used to quantify subset presence in the sample is marked by a red box. A summary of the percentage of each subset(of total cells) present in the samples from all donors used for MS analysis is given in the table. For BDCA3+ DCs values are provided both for total CD11c+BDCA3+ cells (red box) and CD11c+BDCA3hi cells only (dotted red box). Please note that for the CD16+ monocyte sample, CD16 itself could not be used to reliably identify CD16+ cells after purification because of a decrease of CD16 label likely due to competition between the sorting antibody and the staining antibody. Therefore presence of CD16+ monocytes in this sample was based on CD11c only, which was justified by the concomitant absence of CD1c+ and BDCA3hi cells.
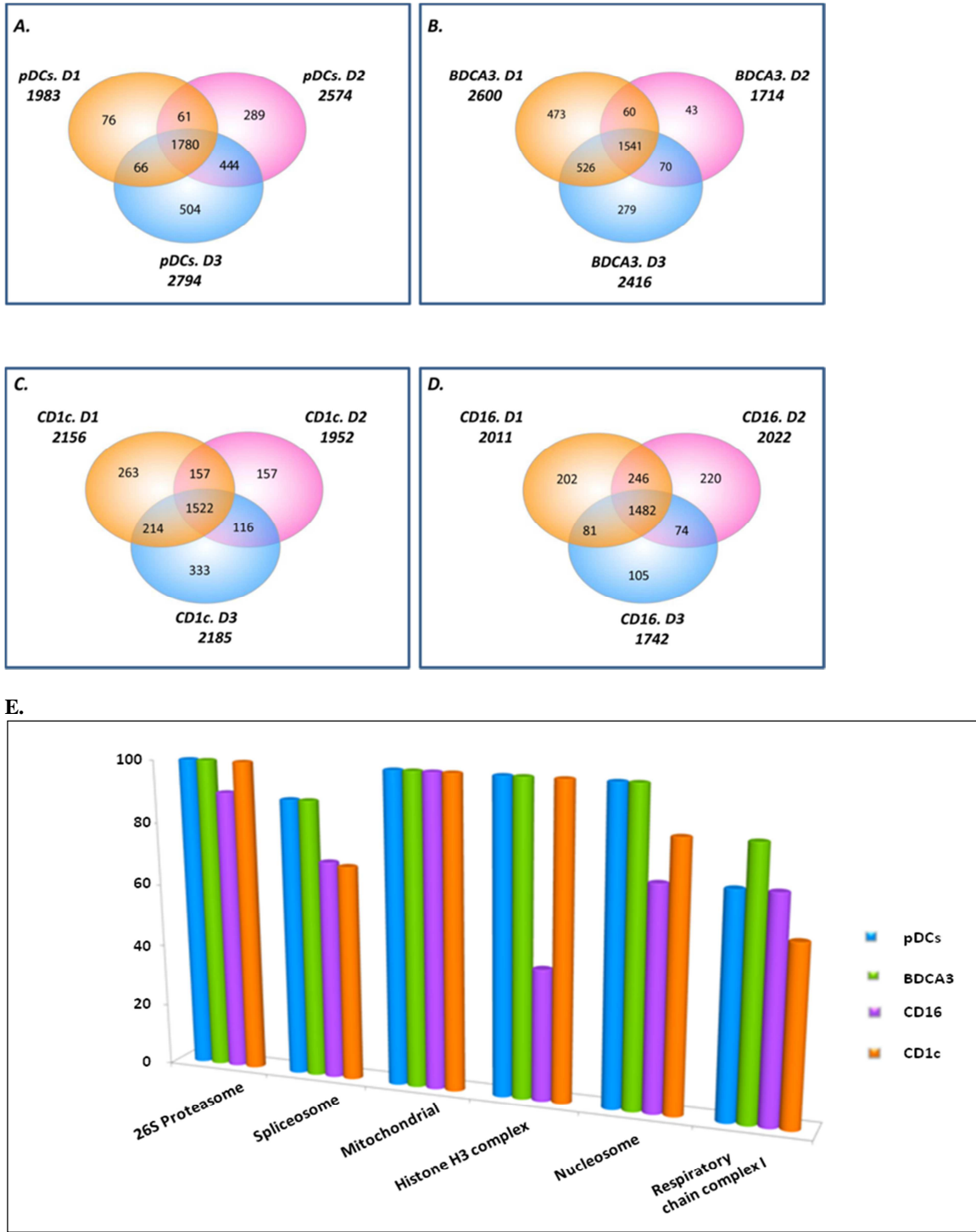
Figure S2 related to figures 1 and 3: Protein identification and coverage.
(A-D) Venn Diagrams representing the proteins identified in each donors for each subsets. (E)The percentage of complex components found back in each subset and donor for the indicated 6 essential protein complexes retrieved from the CORUM database. Coverage of these complexes provides a measure for the completeness of the measured proteome.
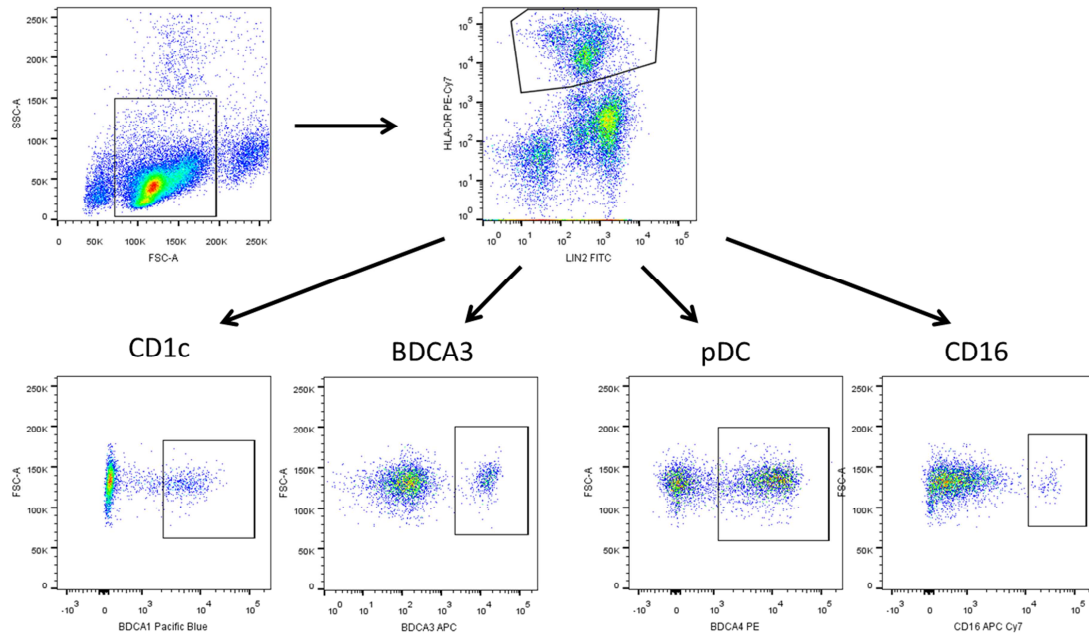
Figure S3 related to figure 6: Gating strategy for flow sorting of DC-subsets for Western blotting and functional analysis. A DC enriched fraction was obtained by magnetic bead depletion of non-DC PBMC constituents. This enriched fraction was subsequently sorted using a pool of lineage markers (CD3, CD14, CD16, CD19, CD20, and CD56; all excluded), HLA-DR (included) and the subset identification markers BDCA1, BDCA3, BDCA4 and CD16.
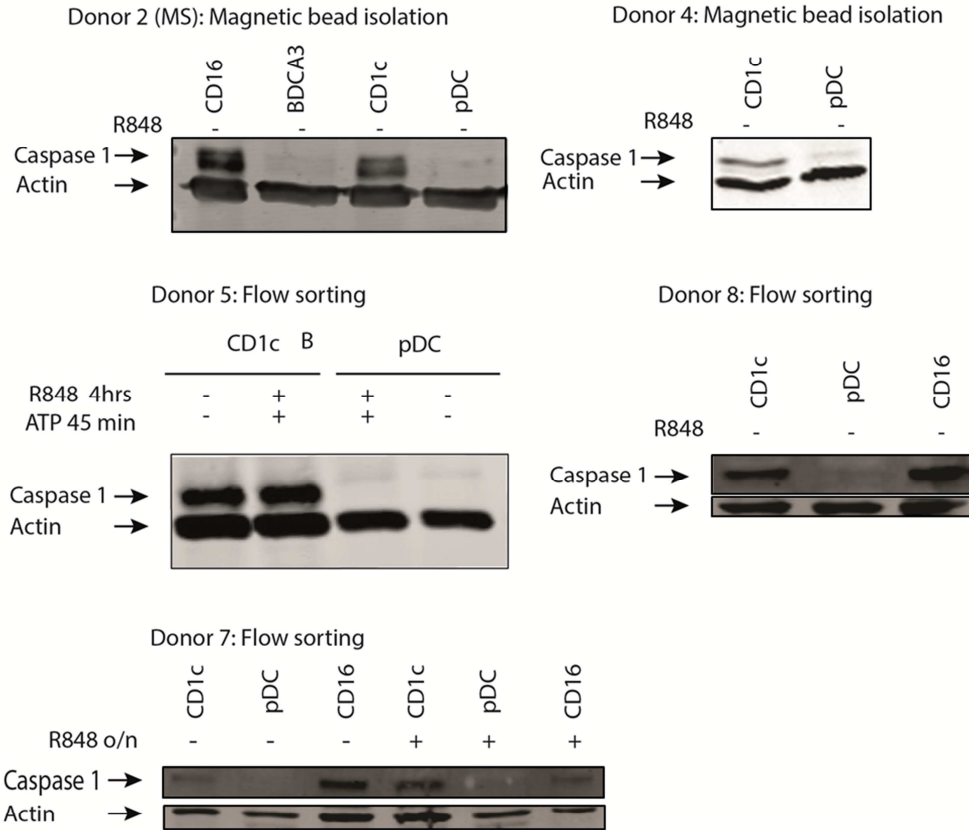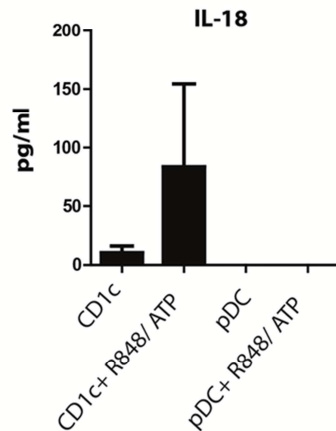
Figure S4 related to figure 6: Additional evidence for the absence of Caspase-1 and Caspase-1 dependent cytokine secretion in pDCs. (A) Additional Western blots for Caspase-1 and actin of magnetic bead isolated or flow cytometry sorted resting or R848 (ATP) stimulated DC-like subsets from one donor also used for MS analysis (donor 2) and 4 additional novel donors (4,5,7,8; Western blots of donors 6 and 9 are depicted in figure 6 of the main manuscript). (B) Presence of IL-18 in the supernatant of CD1c+ mDCs or pDCs measured by ELISA after stimulation for 4 hours with R848 followed by 45 minutes of ATP. Shown is the mean amount secreted by 3 donors +/- SEM.

Supplemental methods

Cells

For proteome analysis, DCs were isolated from aphaeresis products obtained from healthy volunteers after written informed consent and according to institutional guidelines. Peripheral Blood Mononuclear Cells (PBMCs) were purified from aphaeresis products via ficoll density gradient centrifugation (Lucron Bioproducts, Sint Martens-Latem, Belgium). To obtain peripheral blood leukocytes (PBLs), monocytes were depleted from PBMCs via adherence to plastic culture flasks. CD1c+ mDCs and CD16+ mDCs were isolated from PBMCs with a CD1c+ DC isolation kit and CD16+ monocyte isolation kit, respectively. BDCA3 myeloid DCs were isolated from PBLs by selection for BDCA3+ cells with a CD141 (BDCA3) isolation kit. Plasmacytoid DCs were purified from PBLs by positive selection using anti–BDCA-4–conjugated magnetic microbeads (all Miltenyi Biotec, Bergisch Gladback, Germany). DC purity was assessed by flow cytometry by staining for identification markers as indicated in Figure S1. Antibodies used for CD11c, CD1c, CD16+, BDCA3, BDCA2 and CD123 were all from Miltenyi Biotec and as described previously (Schreibelt et al., 2012; Tel et al., 2013). Contamination with T cells and B cells was assessed by double staining of CD19 and CD3 (BD Biosciences).

For Western blotting and ex vivo stimulation, DCs were isolated from PBMCs obtained from buffy coats of healthy donors after written informed consent and according to institutional guidelines. For these experiments, pDCs and CD1c mDCs were either isolated by flow sorting after an initial DC-enrichment using a DC-enrichment kit (Miltenyi Biotec) and subsequent DC-identificationusing Lin1-FITC (BD Biosciences; containing a pool of antibodies for CD3, CD14, CD16, CD19, CD20, and CD56; to be excluded) anti-HLA-DR-PE-Cy7 (positive selection) and anti-BDCA3-APC, anti-BDCA4-PE (B, pDCs) (all from Miltenyi), anti-CD16-APC-Cy7 (BD Biosciences) and anti-CD1c-PB (Biolegend). See figure S3A for the gating strategy. The four subsets were sorted using a FACS Aria II (BD Biosciences).For measuring the subset specific expression of 5 novel selected surface markers, PBMCs from healthy donors were stained for each of the surface marker using specific or isotype control antibodies conjugated to PE (below) and either a cocktail containing CD45-APC-Vio770, CD14-VioGreen, BDCA-3-APC, Clec9a-VioBrightFITC, CD20-PE-Vio770 (all from Miltenyi) CD1c-BV421 (Biolegend) to identify CD1c mDC (CD45+, CD20-, CD14-, BDCA1+) and BDCA3+ mDC (CD45+, CD20-, CD14-, BDCA3+, CLEC9A+), or a cocktail containing CD45-V450 (BD Biosciences), BDCA-2-APC, CD123-APC-Vio770 (both from Miltenyi), CD16-PE-Cy7, HLA-DR-BV510 and Lin2(CD56, CD3, CD14, CD20, CD19)-FITC (all from BD Biosciences) to identify pDCs (CD45+, Lin2-, HLA-DR+, BDCA2+, CD123+) and CD16+ monocytes (CD45+, Lin2-, HLA-DR+, CD16+). Selected novel surface markers were stained using PE-conjugated monoclonal antibodies against CD93 (Miltenyi Biotec), Siglec-9 (R&D systems), SIRPα (Biolegend), CD163 (BD Biosciences) and Siglec-10 (Biolegend).

Protein Extraction/Sample Preparation for MS

The purified and isolated DC subsets were resuspended in homogenisation buffer (20 mM Hepes (Roche) pH 7.5, 250 mM sucrose (Baker) and complete protease inhibitor cocktail (Roche)) and disrupted by three cycles of freezing and thawing. The cells were centrifuged at 10,000 g to separate soluble and insoluble fractions and separated by SDS-PAGE using precasted 4-20% TRIS/Bis ready Gels (Biorad). Both soluble and insoluble fractions were loaded onto gels for MS analysis. For each subset in total 8ug of protein was loaded corresponding to $5-10.10^5$ cells. After electrophoreses the protein gel was stained with Novex Colloidal Blue (Invitrogen) and each lane was cut in to ten fractions. Gel fractions were subsequently treated with dithiothreitol (DTT) and iodoacetamide and digested by trypsin. Digested samples were acidified to a final concentration of 0.5% HAc and purified by STAGE tips as described before (Rappsilber et al., 2003).

Liquid chromatography tandem mass spectrometry

All the samples were analyzed using liquid chromatography (Easy n-LC; Thermo Fisher scientific) coupled to a 7-T linear ion trap Fourier-Transform ion cyclotron resonance mass spectrometer model (LTQ FT Ultra, Thermo Fisher Scientific). Chromatography was performed with PicoTip columns (New Objective, Woburn, USA) of 15 cm 100 μm in size and was packed with 3 μm Reprosil C18 beads (Dr. Maisch). Tryptic peptides were separated using a 90 min gradient from 12% buffer B to 40 % buffer B (buffer B contains 80% acetonitrile in 0.5% acetic acid) with a flow-rate of 300 nL/min. The LTQ-FT instrument was operated in data-dependent mode. Full-scan MS spectra of intact peptides (m/z 350-1500) with an automated gain control accumulation target values of 1.000.0000 ions were acquired in the Fourier transform ion cyclotron resonance cell. The four most abundant ions were sequentially isolated and fragmented in the linear ion trap by applying collisional induced dissociation using an accumulation target value of 10.000, a capillary temperature of 100 ° C, and a normalized collision energy of 27%. A dynamic exclusion of ions previously sequenced was enabled. All unassigned charges states and singly charges ions were excluded from sequencing. A minimum of 200 counts was required for MS2

selection. Maximum injection times were set at 500 ms and 400 ms respectively for FT MS and IT MS/MS measurements.

Mass spectrometry data processing and protein identification
Proteins were identified and quantified from raw mass spectrometric files using MaxQuant software version 1.3.0.5 (Cox and Mann, 2008). Peak lists were generated to contain the six most intense peaks per 100 Dalton mass window. Database search was performed in Andromeda search engine (Cox et al., 2011) against Human Uniprot database (86,749 entries, June 2012) supplemented with sequences of contaminant proteins. We included cysteine carbamidomethylation as a fixed modification and oxidized-methionine and protein N-terminal acetylation as variable modifications. The minimum peptide length was six amino acids and up to two tryptic mis-cleavages were considered for identification. The time window of 2 minutes was allowed to match peptides across different LC-MS/MS run from each fraction on basis of mass and retention time. At least two peptides, of which one was unique, were required for the protein identification. Protein quantification was based on both unique and "razor" peptides. The protein abundance was determined by MaxLFQ that is based on comparison of individual peptide intensities over all samples as described by (Cox et al., 2014) and LFQ intensities values were normalized across biological replicates using median value. For ANOVA and hierchical clustering (below) missing values were filled by random values from the lower end of the expression value spectrum using Perseus software (version 1.4.2.23; Tyanova et al., 2016). Raw MS data are available at proteome central (accession number PXD004678) (http://proteomecentral.proteomexchange.org)

Statistical analysis of protein data
Since DC subsets derived from the third donor were of lower purity further use of these samples is only justified when sufficiently stringent criteria are used. Therefore, to extract reliable and informative data, observed protein expression differences were only used when present in at least two out of three donors or when confirmed by additional data from independent studies.
The list of differentially expressed proteins for hierarchical clustering and signature generation was obtained by merging subsets specifically expressed proteins (present in at least 2 out of 3 donors and absent in all other subsets in all donors) to proteins that were called significantly differentially expressed by 3-group one-way ANOVA (p-0.05) on the imputation supplemented protein data. On the ANOVA p-values a bonferoni-Holms correction for multiple testing was performed (Table S7). Furthermore, to prevent false positive results, proteins expressed below a measured LFQ value of 2^{17} were excluded from ANOVA because for these proteins the imputed LFQs to fill missing values may exceed that of the measured values. ANOVA was performed in the R programming environment.

Data processing of publically available transcriptome data
The raw microarray data (CEL files) were downloaded from ArrayExpress (accession: E-TABM-34) for human DC subsets (pDCs, BDCA3+ DCs, CD1c+ DCs, CD16+ cells). The raw data was normalized using the RMA normalization function of the affy package and mapped using hgu133plus2.db annotation package (Gautier et al., 2004). After data normalization, only the genes that had an expression level above background in at least 1 out of 4 populations with a differential expression between at least 2 cell types were considered. In cases where multiple probes were related to a gene, only the probe expressed highest across all samples was considered.

RNA sequencing
For RNA sequencing RNA was extracted from subsets lysed immediately after harvesting in Trizol reagent (Thermo fisher) and forcing the lysate trough a 25 gauge syringe for 10 times. RNA was subsequently isolated using and RNeasy kit and on-column DNase treatment (both from Qiagen). Thereafter for each sample 250 ng of total RNA was treated by Ribo-Zero rRNA Removal Kit (epicentre) to remove ribosomal RNAs (according to manufacturer instructions) and mRNA was purified using the (Zymo research). Subsequently purified RNA was fragmented using (5x) fragmentation buffer (200 mM Tris acetate pH 8.2, 500 mM potassium acetate and 150 mM magnesium acetate) and incubated at 95°C for 90 seconds. After a new round of purification, first strand cDNA was synthesized from fragmented RNA with SuperscriptIII (Invitrogen). First strand cDNA was purified by MinElute cleanup kit (Qiagen) and second strand cDNA was prepared in the presence of dUTP instead of dTTP and using random hexamers. Double stranded cDNA was purified by Qiagen mini elute columns and used for Illumina sample prepping and sequenced according to the manufacturer's instructions. The RNAseq reads were mapped to HG18 and used to calculate RPKM values (reads per kilobase of gene length per million reads) (Mortazavi et al., 2008).

Hierarchical clustering
The hierarchical clustering of samples and genes was performed based on 1-Pearson correlation in combination with average linkage clustering using GENE-E software (http://www.broadinstitute.org/cancer/software/GENE-E).Proteome and transcriptome data were mapped based on similarity of the gene symbol. Both datasets were

separately normalized and z-scored in the R programming environment (mean to 0 and variance to 1). From the merged dataset differentially expressed proteins we extracted (present in at least 2 out of 3 donors and absent in all other subsets in all donors plus proteins called significantly differentially expressed by 3-group one-way ANOVA (p-0.05). The thus obtained dataset was used for hierarchical clustering of samples.

Signature generation

For proteins highly or lowly expressed in each of the 3 subsets (pDCs, CD1c+ mDC, CD16+ monocytes), protein signatures were generated by combining protein and transcriptome data in 4 levels of evidence according to the criteria below (Table S7).

For the <u>high expression</u> signatures for each subset the criteria were:
<u>Level I)</u> specific proteins with RNA support. Protein: Is specifically expressed in the subset in 3 out of 3 donors but absent in all other subsets. RNA: MinMax >1.5 fold for linear values (for Log2 values, FC = $Log2(Min)_{cluster-subset(s)}$-$Log2(Max)_{other\ subsets}$> 0.58). Proteins specifically expressed only in 2 out of 3 donors were "rescued" and included in the signatures because of RNA support (MinMax >1.5 fold).
<u>Level II)</u> differentially expressed proteins with RNA support. Protein: 3-group one-way ANOVA proteomics with multiple testing correction (BH) <0.05 and expression in the subset(s) in at least 2 out of 3 donors. Post-hoc t-test comparisons proteomics for the high expressing subset(s) against other subset p<0.05 and a fold change of >2. RNA: MinMax >1.5 fold for linear values. ANOVA significant proteins (p<0.05) lost by multiple testing correction but meeting all other criteria and supported by RNA expression (MinMax >1.5 fold) were included in Table S7 as low confidence DEPs.
<u>Level III)</u> Specific protein expression without RNA support. Protein: As level I. RNA: Criteria Level I are not met or no data available. Proteins specifically expressed in only 2 out of 3 donors were also included Table S7 as low confidence DEPs
<u>Level IV)</u> Differential protein expression without RNA support. Protein as in level II; RNA: Conditions level II are not met or no data available.

For the proteins <u>lowly expressed</u> in subset the criteria were as above with the following modifications:
<u>Level I)</u> Protein: Absent in the cluster-subset but present in all other subsets; RNA: Minimum expression in other subsets over maximum in the subset > 1.5 fold. Proteins absent in the cluster-subset in 3/3 donors but present in only 2/ 3 donors in (any of) the other subsets are
included in Table S7 as low confidence DEPs.
<u>Level II)</u> Protein: Presence in the other subset in at least 2 out of 3 donors.
<u>Level III)</u> Protein as for level I. Proteins absent in the cluster-subset in 3/3 donors but present in only 2/ 3 donors in (any of) the other subsets were also included in Table S7 as low confidence DEPs.
<u>Level IV)</u> Protein as for level II. RNA: Did not meet criteria level II or no data available.

Finally, proteins identified in at least two donors in one subset but not identified as differentially expressed according to the signature criteria above, yet found differntially expressed based on RNA expression data are given in supplementary table S8.

Extracting DEPs between BDCA3+ mDCs and CD1c+ mDCs:
To extract DEPs we aplied roughly the same structure as for signature generation but now based on t-tests and with the follwing specific rules:
Level I: present in > 2/3 donors in one of the 2 subsets, no presence in the other subset (3/3), also significant at RNA level by Min Max (> 1.5 fold)
Level II: t-test significant (p-value <0.05, FC >2), also significant at RNA by Min max (> 1.5 fold)
Level III: present in > 3/3 donors in one of the 2 subsets, no presence in the other subset (3/3), no RNA data or significance. Proteins only specifically present in 2/3 donors are only present in Table S7 as low confidence DEPs
Level IV: t-test significant (p-value < 0.01, FC>2); no presence in the other subset (3/3), no RNA data or significance. Proteins with a t-test p-value between 0.01 and 0.05 are only present in Table S7 as low confidence DEPs

Protein-protein Interaction, GO and pathway analysis
Signatures depicted in figure 2 were used as input for protein-protein interaction (PPI) analysis using the STRING PPI web tool (http://string-db.org/; version 10). A confidence score of 0.4 was used as a cut-off for protein-protein interaction allowing all sources of evidence. Obtained confidence scores for each interaction can be found in table S9. The generated interaction networks were uploaded in Biolayout express 3D (version 3.3) for graphical representation (Theocharidis et al., 2009). Signatures were mapped onto functional categories

using the functional annotation clustering algorithm in the DAVID web tool (https://david.ncifcrf.gov/) using standard settings. Full functional annotation output can be found in Table S9.

Western blotting

Reduced cells lysates were loaded on a 10% SDS-gel and transferred to a PVDF membrane. Caspase-1 was stained with the caspase-1-p10 (C-20): sc-515 antibody (Santa Cruz Biotechnology) and Actin with anti-Actin (20-33) antibody (Sigma). Blots were analyzed on the Odyssey imaging system.

ELISA

Cytokine production of IL-1β (R&D Systems) and TNF (eBioscience) was measured by ELISA. IL-18 production was analyzed by using Luminex. To monitor activation, cells were stained for 30 minutes with anti-CD83-FITC (BD Biosciences) and analyzed by flow cytometry.

Supplemental references

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics *13*, 2513-2526.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. Journal of proteome research *10*, 1794-1805.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307-315.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem *75*, 663-670.

Schreibelt, G., Klinkenberg, L.J., Cruz, L.J., Tacken, P.J., Tel, J., Kreutz, M., Adema, G.J., Brown, G.D., Figdor, C.G., and de Vries, I.J. (2012). The C-type lectin receptor CLEC9A mediates antigen uptake and (cross-)presentation by human blood BDCA3+ myeloid dendritic cells. Blood *119*, 2284-2292.

Tel, J., Schreibelt, G., Sittig, S.P., Mathan, T.S., Buschow, S.I., Cruz, L.J., Lambeck, A.J., Figdor, C.G., and de Vries, I.J. (2013). Human plasmacytoid dendritic cells efficiently cross-present exogenous Ags to CD8+ T cells despite lower Ag uptake than myeloid dendritic cell subsets. Blood *121*, 459-467.

Theocharidis, A., van Dongen, S., Enright, A.J., and Freeman, T.C. (2009). Network visualization and analysis of gene expression data using BioLayout Express(3D). Nature protocols *4*, 1535-1550.

Overview of supplemental Tables:

Table S1: Correlation between biological and technical replicates_related to figures 1 and 4
Tab1: Technical replicates
Tab2: biological replicates
Table S2: Identified proteins_related to figures 1 and 4: Protein table of all proteins identified in all 4 subsets and all 4 donors (at least 2 in one subsets, no LFQ restrictions)

Table S3: Identified peptides_related to figures 1 and 4: Peptide table of all peptides identified in all 4 subsets in all 4 donors

Table S4: Pairwise comparison of subsets_related to figures 1 and 4: t-test of all combinations of all 4 subsets

Table S5: RNA sequencing data donor 2_related to figures 1 and 4

Table S6: Merged RNA and protein data_related to figures 1 and 4

Tab1: merged table RNA and protein data (MA and RNA seq)

Tab2: correlation of RNA (MA and RNA seq)

Table S7: Signatures_related to figures 2 and 4: Signatures for all 6 groups in the 3 subset comparison and for BDCA3 compared to CD1c+ mDC

Tab1: pDC vs CD1c & CD16 up

Tab2: pDC vs CD1c & CD16 down

Tab3: CD1c vs pDC& CD16 up

Tab4: CD1c vs pDC% CD16 down

Tab5: CD16vs pDC&CD1c up

Tab6: CD16 vs pDC&CD1c down

Tab7: BDCA3 vs CD1c up

Tab8: CD1c vs BDCA3 up

Table S8: Table S8_DEGs not confirmed by proteomics_related to figure 2 and 4

Tab1: pDC vs CD1c & CD16 up

Tab2: pDC vs CD1c & CD16 down

Tab3: CD1c vs pDC& CD16 up

Tab4: CD1c vs pDC% CD16 down

Tab5: CD16vs pDC&CD1c up

Tab6: CD16 vs pDC&CD1c down

Tab7: BDCA3 vs CD1c up

Tab8: CD1c vs BDCA3 up

Table S9: PPI and FA analysis_ related to figure 3

Tab1: pDC vs CD1c & CD16 up

Tab2: pDC vs CD1c & CD16 down

Tab3: CD1c vs pDC& CD16 up

Tab4: CD1c vs pDC% CD16 down

Tab5: CD16vs pDC&CD1c up

Tab6: CD16 vs pDC&CD1c down

Tab7: BDCA3 vs CD1c up

Tab8: CD1c vs BDCA3 up