Defining mononuclear phagocyte subset homology across several distant warm-blooded vertebrates through comparative transcriptomics

Thien-Phong Vu Manh[1,2,3]*‡, Jamila Elhmouzi-Younes[4]†‡, Céline Urien[4], Suzana Ruscanu[4], Luc Jouneau[4], Mickaël Bourge[5], Marco Moroldo[6], Gilles Foucras[7,8], Henri Salmon[9,10], Hélène Marty[9,10], Pascale Quéré[9,10], Nicolas Bertho[4], Pierre Boudinot[4], Marc Dalod[1,2,3]*§ and Isabelle Schwartz-Cornil[4]*§

[1] UM2, Centre d'Immunologie de Marseille-Luminy, Aix Marseille Université, Marseille, France, [2] U1104, INSERM, Marseille, France, [3] UMR7280, CNRS, Marseille, France, [4] UR892, Virologie et Immunologie Moléculaires, INRA, Domaine de Vilvert, Jouy-en-Josas, France, [5] IFR87 La Plante et son Environnement, IMAGIF CNRS, Gif-sur-Yvette, France, [6] CRB GADIE, Génétique Animale et Biologie Intégrative, INRA, Domaine de Vilvert, Jouy-en-Josas, France, [7] UMR1225, Université de Toulouse, INPT, ENVT, Toulouse, France, [8] UMR1225, Interactions Hôtes-Agents Pathogènes, INRA, Toulouse, France, [9] UMR1282, Infectiologie et Santé Publique, INRA, Nouzilly, France, [10] UMR1282, Université François Rabelais de Tours, Tours, France

Mononuclear phagocytes are organized in a complex system of ontogenetically and functionally distinct subsets, that has been best described in mouse and to some extent in human. Identification of homologous mononuclear phagocyte subsets in other vertebrate species of biomedical, economic, and environmental interest is needed to improve our knowledge in physiologic and physio-pathologic processes, and to design intervention strategies against a variety of diseases, including zoonotic infections. We developed a streamlined approach combining refined cell sorting and integrated comparative transcriptomics analyses which revealed conservation of the mononuclear phagocyte organization across human, mouse, sheep, pigs and, in some respect, chicken. This strategy should help democratizing the use of omics analyses for the identification and study of cell types across tissues and species. Moreover, we identified conserved gene signatures that enable robust identification and universal definition of these cell types. We identified new evolutionarily conserved gene candidates and gene interaction networks for the molecular regulation of the development or functions of these cell types, as well as conserved surface candidates for refined subset phenotyping throughout species. A phylogenetic analysis revealed that orthologous genes of the conserved signatures exist in teleost fishes and apparently not in Lamprey.

Keywords: comparative biology, immunology, dendritic cells, monocytes, macrophages, genomic and bio-informatic methods

Introduction

Reaching the global health objective requires to improve disease prevention and treatments in humans and in a wide variety of animal species. To achieve that goal, knowledge of the immune system, and particularly of the mononuclear phagocyte system that orchestrates the immune response, needs to be translated across species in order to develop better vaccines and immune response-targeting therapies in relevant species.

The mononuclear phagocytes encompass three main functional cell types: monocytes (Mo), macrophages (MP), and DC. The main functions of Mo are to patrol the body to detect infections and to produce microbicidal compounds including TNF, superoxide, or nitric oxide intermediates, or to differentiate into MP. The main function of MP is to preserve tissue homeostasis through trophic and scavenger functions. DCs are professional antigen-presenting cells that are key instructors of immunity, controlling tolerance to self and immune defense against pathogens. However, beyond these generic definitions, each of these mononuclear phagocyte category encompasses a complex array of different subtypes with distinct ontogeny and functions, as described extensively in mice and to some extent in humans. Mo include at least two main subsets, classical Mo (cMo) and non-classical Mo (ncMo) (1), that express different innate immune recognition receptors and mediate distinct functions, with ncMo showing the original property of patrolling blood vessels (2). Adult MP are derived either from embryonic precursors and self-renew in tissues, or in some cases are replenished from circulating Mo (2–6). The MP subtypes populating different tissues show distinct molecular and functional characteristics which are in a large part determined by their anatomical microenvironment (7, 8). Two cell types with morphologic and functional features of DC derive from the Mo/MP lineage, namely monocyte-derived DC (MoDC) and Langerhans cells (9). MoDC are generated (i) upon inflammatory stimuli *in vivo* (10), (ii) at steady-state in the skin (3), and (iii) upon culture of purified Mo or of total bone marrow cells with GM-CSF ± IL-4 *in vitro* (11, 12). Langerhans cells derive from embryonic monocytic precursors upon IL-34 signaling and populate the outer layer of epithelia (13). Finally, three types of *bona fide* DC exist, the plasmacytoid DC (pDC) and the conventional DC (cDC) cDC1 and cDC2 types which derive from a bone marrow common DC precursor and are present both in lymphoid organs and as interstitial DC in the parenchyma of non-lymphoid tissues such as skin, lung, gut, and liver (14). Comparative transcriptomic analyses pioneered by us and used by other groups, as well as functional studies, have demonstrated the existence of similar mononuclear phagocytes and DC subsets between human and mice (15–20). DC subset candidates have also been described in other mammals such as in ruminants and pigs. However, no systematic study has demonstrated the existence of a framework of homologous DC subsets throughout distant species [for review see Ref. (21)]. Overall, it remains unknown whether a similar diversity in mononuclear phagocyte subsets exists across distant mammals and vertebrates, and when during evolution this complex organization of the mononuclear phagocyte system arose.

The combination of phenotypic, functional, and ontogenic studies used in the mouse model cannot be used to define cell subsets in most other species of interest due to technical, financial, or ethical limitations. As the ontogeny and functions of cell types are instructed by specific gene-expression modules, cell type identity can be defined by its molecular fingerprinting (22). We thus reasoned that mononuclear phagocyte subset identity could be defined by gene-expression profiling, whatever the species. In addition, cell types that are homologous between species must exhibit closer molecular fingerprints and gene-expression programs than non-homologous cell types, based on the definition of homologous cell types as "those cells that evolved from the same precursor cell type in the last common ancestor" (23).

In this paper, we developed a streamlined approach (see Figure S1 in Supplementary Material) to identify homologous mononuclear phagocyte subsets in distant species with reference to the mouse, consisting in (i) designing antibody panels for sorting candidate cell subsets to high marker-based purity, (ii) generating genome expression profiling of the sorted cell subsets, and (iii) performing computational transcriptomic analyses to establish gene signatures and compare them to the transcriptomic fingerprints of the well-characterized immune cell types of the mouse referent species. Our analysis was extended to chicken cell subsets, showing that it is amenable to establish mononuclear phagocyte subset homology throughout vertebrates. We also derived gene-expression signatures and gene interaction networks that are selectively expressed in mononuclear phagocyte subsets in a conserved manner throughout distant mammals and that can be used to identify homologous subsets throughout species. The conserved gene-expression signatures and networks not only encompassed genes with known functions in mononuclear phagocyte subsets but also pointed out novel candidate genes likely involved in the ontogeny or functional specialization of these cell types. Finally, we conducted a phylogenetic analysis to examine the presence in bony fishes and in Lamprey of orthologs of genes from the transcriptomic signatures identified in mammals.

Materials and Methods

Pigs and Sheep for Blood Collection

All animal experiments were carried out under licenses issued by the Direction of the Veterinary Services of Versailles (accreditation numbers B78-93) and under approval of the Committee on the Ethics of Animal Experiments of AgroParisTech and INRA-Jouy-en-Josas (COMETHEA, authorization number 00604.01). The eight pigs (blood) used in this study (four males, four females) were around 2 years old and weighted between 60 and 85 kg. Down-sized pigs were kept at the Centre d'Imagerie Interventionnelle (Jouy-en-Josas). «Prealpe» female sheep (total 37, 50–80 kg), originate from and were raised in the «Unité Commune d'Expérimentation Animale» in Jouy-en-Josas, France. Blood (<400 ml/animal) was collected by venous puncture on sodium citrate.

Isolation of DC Subset Candidates, B Lymphocytes, and Mo from Pig Blood

PBMC were obtained from pig peripheral blood buffy coat samples by 1.076 g/ml density Percoll (GE Healthcare) gradient centrifugation (24). For B cell sorting, PBMC were surface-labeled with 2 μg/ml primary monoclonal antibody (mAb) against IgL (K139 3E1, IgG2a) followed by Alexa647-conjugated goat anti-mouse

IgG2a antibodies (Invitrogen). For pDC sorting, PBMC were surface-labeled with 2 µg/ml primary mAb anti-pig CD4 (PT90A, IgG2a), CD3 (8E6, IgG1), CD14 (CAM36, IgG1), and CD172A (74-22-15, IgG2b) followed by Alexa488, phycoerythrin (PE), or Alexa647-conjugated goat anti-mouse isotype-specific antibodies (Invitrogen). Blood pDC candidates were sorted as CD3$^-$ CD14$^-$ CD4$^+$ CD172^{int} cells, based on previously published indicative data (25). For cDC candidates and Mo sorting, PBMC were surface-labeled with 2 µg/ml mAb anti-pig IgL (K139 3E1, IgG2a), anti-pig IgG (K138 4C2, IgM), anti-pig IgM (PG145A, IgM), anti-pig CD4 (PT90A, IgG2a), anti-human and pig cross-reacting CD14 (TUK4, IgG2a), anti-pig CD172A (74-22-15, IgG1), anti-artiodactyl MHC class II (Th21A, IgG2b), and chicken anti-human and artiodactyl cross-reacting CADM1 (3E1, IgY). The primary antibodies were revealed with Alexa488, PE, or Alexa647-conjugated goat anti-mouse isotype-specific antibodies and with donkey anti-chicken IgY Peridinin Chlorophyll Protein Complex (PerCP)-conjugated IgG. The cDC2 candidates were isolated as FSC^{hi} IgL$^-$ IgG$^-$ IgM$^-$ CD4$^-$ CD14$^-$ MHC class II$^+$ CADM1$^-$ CD172^{hi} or CD172^{int} cells. The cDC1 candidates were isolated as FSC^{hi} IgL$^-$ IgG$^-$ IgM$^-$ CD4$^-$ CD14$^-$ MHC class II$^+$ CADM1$^+$ CD172^{lo} cells. Mo candidates were sorted as MHC class II$^-$ CD172^{hi} cells. Non-relevant antibodies (IgG1, IgG2a, IgG2b, and IgM) were systematically used as controls to measure the level of non-specific background signal caused by primary antibodies. The cell subsets were sorted by flow cytometry on the ImaGif Cytometry platform using the analyzer-sorter MoFlo XDP cytometer and the Summit 5.2 software from Beckman Coulter (cytometric assessment of post-sort purity >98%). The numbers of DCs that were collected per pig lay between 2 and 3 × 10^5 for pDC, 25 and 47 × 10^3 for cDC1, 20 and 40 × 10^5 for cDC2 candidates.

Isolation of DC Subset Candidates from Sheep Blood and B Lymphocytes and Macrophages from Sheep Spleen

Sheep PBMC were loaded on 1.065 density iodixanol gradient (Optiprep, Nycomed Pharma) to isolate low density cells from blood. Sheep pDC candidates were isolated by flow cytometry as previously described (26). For isolating sheep cDC candidates, the low density PBMC from several sheep were reacted with anti-CD11c mAb (2 µg/ml, OM1 clone, IgG1) followed by a saturating concentration of pacific blue-labeled anti-mouse IgG donkey Fab (50 µg/ml). After extensive wash, cells were further incubated anti-CD172A mAb (2 µg/ml, ILA24, IgG1) followed by a saturating concentration of Alexa488-labeled anti-mouse IgG donkey Fab (50 µg/ml). After extensive wash, cells were incubated with 2 µg/ml primary mAbs anti-ruminant B cells (DU-204, IgM), CD11b (ILA130, IgG2a), TCR1γ/δ receptor (CC15, IgG2a), CD45RB (CC76, IgG1), and chicken anti-human and artiodactyl cross-reacting CADM1 (3E1, IgY). The IgM and IgG2a primary antibodies were revealed with PE-conjugated goat anti-mouse isotype-specific antibodies, the IgG1 anti-CD45RB primary antibody was revealed with Alexa647-conjugated goat anti-mouse IgG1 antibody, and the anti-CADM1 with anti-IgY PerCP-conjugated IgG. The cDC2 candidates were isolated by flow cytometry as B$^-$ CD11b$^-$ TCR1$^-$ CD45RB$^-$ CD11c$^+$ CADM1^{lo} CD172^{hi} FSC^{hi} cells. The cDC1 candidates were isolated by flow cytometry as B$^-$ CD11b$^-$ TCR1$^-$ CD45RB$^-$ CD11c$^+$ CADM1^{hi} CD172^{lo} FSC^{hi} cells. The numbers of DCs that were collected per sheep lay between 1 and 2 × 10^5 for pDC, around 600 for cDC1, and around 4000 for cDC2. The far lower amounts of collected blood cDCs from sheep as compared to pig may probably originate from the multiple staining steps due to the necessity to separately identify several IgG1 as primary antibodies. B cells and MP were sorted by flow cytometry from isolated sheep spleen cells using the anti-ruminant B cell (DU-204, IgM) and anti-CD14 (CAM36, IgG1), respectively.

Production of Sheep MoDC

Three independent cultures of sheep MoDC were produced with GM-CSF as previously described (27).

RNA Extraction and Hybridization on Microarrays

Total RNA from subsets was extracted using the Arcturus PicoPure RNA Isolation Kit (Arcturus Life Technologies) and checked for quality with an Agilent 2100 Bioanalyzer using RNA 6000 Nano or Pico Kits (Agilent Technologies). All RNA samples had an RNA integrity number (RIN) above 8.5. When insufficient total RNA amounts for hybridization were obtained (<25 ng for sheep DNA chips, <50 ng for pig DNA chips), the RNAs from the sorted subsets of distinct animal were mixed. RNA amplification and labeling was performed using the one-color Low Input Quick Amp Labeling kit (Agilent Technologies) following the manufacturer's recommendations. Each RNA sample (25 ng for sheep and 50 ng for pig) was amplified and cyanin 3 (Cy3) labeled, and subsequently the complementary RNA (cRNA) was checked for quality on a Nanodrop and on an Agilent 2100 Bioanalyzer. The cRNAs (600 ng) were fragmented and used for hybridization on custom-designed Agilent ovine and porcine arrays. Our arrays for sheep and pig were custom-designed based on the commercial ovine Agilent arrays for these two species, as previously described (28, 29). In brief, the commercial probes with poor Sigreannot scores (30) were replaced with new probes designed using the e-array software from Agilent Technologies and including ovine or porcine orthologs of genes known to be selectively expressed in human and mouse DC subsets (15). After hybridization of the cRNAs on the custom-designed ovine array, the chips were washed according to the manufacturer's protocol and scanned using a G2565CA scanner (Agilent Technologies) at the resolution of 3 µm. The resulting .tiff images were extracted using the Feature Extraction software v10.7.3.1 (Agilent Technologies), using the GE1_107_Sep09 protocol. All the protocols used can be obtained by contacting the CRB GADIE facility[1]. The transcriptomic data from the chicken immune subsets were obtained from a previous study (31). All microarray data have been deposited in the Gene Expression Omnibus (GEO) database under reference numbers GSE9810, GSE53500, GSE55642 which have already been released and GSE66311 which is under embargo until publication of the present study.

---

[1]http://crb-gadie.inra.fr/

Computational Pipeline to Assess Cell Subset Homology Across Species

We have designed a computational pipeline in order to define cell subset homology across species, based on the analysis of gene-expression microarray data. In the current study, it has been applied to identify homology relationships between mononuclear phagocyte cells in mammalian species and then extended to the comparison with a more distant species (chicken). However, it can be applied to any cell type and to any species, provided that the annotations of the genes for each species are sufficiently well documented to allow the retrieval of the orthologous genes. In order to perform the comparison of expression profiles of cells coming from different species, thus from different platforms, we have designed two independent procedures. The first procedure (Figure S2A in Supplementary Material) is based on the assessment of the conservation of cell-specific fingerprints/signatures, as assessed by performing gene set enrichment analyses (GSEA, see below) between pairs of cell types. The second procedure (Figure S2B in Supplementary Material, see below) consists in cross-normalizing the expression datasets coming from the different species, in order to simultaneously examine the relationships between all cell types together.

Cross-Species Transcriptome Comparison by Pairwise Gene Set Enrichment Analyses

The methodological pipeline is depicted in Figure S2A in Supplementary Material, based on an example with comparison of three different species (A, B, and C). Species A is the reference species, i.e., the species for which the cell types are the most accurately described and generally also for which gene orthologous relationships can be retrieved from (mouse or human here). Species B and C are the test species, i.e., the species for which the identity of the cell types has to be established. Coming from three different platforms, the expression datasets have different numbers of probes, illustrated by boxes of different size. In brief, the strategy is to examine by GSEA whether the transcriptomic fingerprint of a given cell type (X) from the referent species A is enriched in one cell type (Y) of a test species (B for example) as compared to all other cell types of the same species. If this is the case, this would support the hypothesis that the cell type Y from the test species B is homologous to the cell type X of the referent species A. To perform these high-throughput GSEA in a processive way that could be easily reproduced and interpreted by other researchers devoid of bio-informatics expertise, we designed and implemented a dedicated software, called Bubble GUM (manuscript in preparation in which an extensive description of the software will be provided)[2]. Bubble GUM encompasses two main modules, GeneSign and BubbleMap, respectively, dedicated to the generation of gene sets and to their use for GSEA applied to multiple pairwise comparisons of samples integrated together into a simple graphical output that helps in the interpretation of the results. The first step consists in extracting from the reference species the transcriptomic fingerprints of each cell type. A cell-specific transcriptomic fingerprint can be defined as the list of genes that

are more highly expressed in the cell type of interest than in all other cell types. These fingerprints were extracted using the "Min (test) vs. Max (ref)" method [(minimum expression among all replicates for all samples for which the transcriptomic fingerprint is defined/maximum expression among all replicates used as reference) $\geq$1.5-fold] (15, 32), using the GeneSign module of Bubble GUM. These transcriptomic fingerprints, in gene symbol format, will be assessed for enrichment on the expression datasets of species B and C. Thus, it is necessary to convert the probe annotations from the arrays of species B and C into the gene symbol of their orthologous counterparts in species A. For this purpose, we used the orthology relationships defined by the Signae pipeline which annotated the pig and sheep genes with their human and mouse orthologous gene symbols (30). The genes present on the gene chips of species B and C that were not associated to an orthologous counterpart in species A remained annotated with the gene symbol corresponding to their species of origin. The statistical enrichment of the cell-specific transcriptomic fingerprints extracted from the reference species A were then calculated between pairwise comparisons of cell types from species B or C with the GSEA methodology, using gene set permutations for computing the *p*-values and false-discovery rates (FDR) (33). This was achieved, and the results graphically represented, by using the BubbleMap module of Bubble GUM.

Cross-Normalization of the Species-Specific Expression Datasets

Using the same starting expression datasets, this is an alternative strategy which is complementary to the pairwise GSEA of the species-specific expression datasets, since it allows clustering all cell types together based on the overall evaluation of the proximity of their expression patterns of hundreds to thousands of orthologous genes. The first step (Orthology Filter) consists in aligning the genes across the species (A, B, and C). It requires retaining only one representative probe per gene for each species/platform. This is needed since, in microarray designs, many genes are often each represented several times by a number of individual probes having each a different signal-to-noise ratio. However, probes have no equivalence across species, whereas genes do. In our experience, the signal-to-noise ratio is generally better for probes that have the strongest signal in positive control samples, while certain probes that have a low signal-to-noise ratio can give misleading high fold changes across conditions when using a limited number of replicates. Hence, we computed for each probe in each platform the sum of normalized expression values across all samples and kept for each gene the probe that had the highest computed value. Then, for the genes of species B and C, we retrieved the gene symbol of their orthologous counterparts in species A (reference species). In the example illustrated in Figure S2B in Supplementary Material, species A is the reference: the genes of species B and C are thus annotated using the gene symbol of the orthologous genes in the species A. The genes not represented in each of the gene chip platform used were removed from the analysis. This Orthology Filter yielded a filtered expression dataset for each species, where the number of genes and their associated symbols were similar between all species, as illustrated by boxes of the same size (Figure S2B in Supplementary Material). In order to be able

---

to rigorously merge the different datasets together, the dynamic ranges of expression values for each gene across all species must be homogenized by setting for each dataset and for each gene across all samples the mean expression to 0 and the variance to 1, a process called data centering and reduction. To prevent this mathematical transformation of the expression data to introduce noise by forcing artifactual expression changes for genes that were unregulated in the initial datasets, it is mandatory to remove all the genes that are not regulated in at least one of the datasets. This thus requires keeping only the genes that are differentially expressed between at least two cell types for each of the species studied. This was achieved in the second and third steps of the data processing. The second step (Differentially Expressed Gene, DEG, Filter) consisted in identifying independently in each dataset the genes that are differentially expressed between at least two cell types. The identification of DEG was performed by calculating the minimal ratio between each pairwise comparison of cell types and by selecting only the genes for which this minimal ratio was higher than twofold. The third step (DEG intersection) consisted in keeping only the genes that were common to all filtered DEG lists, i.e., the orthologous genes which expression was modulated across samples in each of the species studied. The fourth step consisted in data centering and reduction for each dataset, which was performed using the R statistical environment. This step consists in setting, for each dataset, the mean to 0 and the variance to 1, so that all datasets are comparable. In the fifth and final step, the different datasets were merged together simply by aligning their rows based on the common gene symbol extracted from species A. The final cross-normalized expression dataset including the data for all species was then used to perform canonical analyses for classification of samples, namely here hierarchical clustering.

Generation of Conserved Cross-Species Cell Type-Specific Signatures

For each species (human, mouse, sheep, and pig), the transcriptomic fingerprint of each cell type was generated by selecting the genes more highly expressed in the cell type of interest, as compared to all other studied cell types of the same species in the case of "absolute" transcriptomic signatures, or as compared to selected cell populations of the same species in the case of "relative" transcriptomic signatures, using the "Min (test) vs. Max (ref) ≥1-fold" method. Once the fingerprints had been obtained for each species for a given cell type, the gene identifiers were all converted into their corresponding official human gene symbol using BioMart and we selected the intersection of these four lists as the final conserved cross-species transcriptomic signature specific of that cell type, with the following exceptions. First, for certain cell types such as MoDC, data were available from only three, and not four, species. Second, in order to avoid removing putatively relevant signature genes, we kept in these signatures the genes found in all species but one, when their absence in the signature of that given species was due to absence or non-functionality of corresponding ProbeSets on the array of that species.

Real-Time PCR

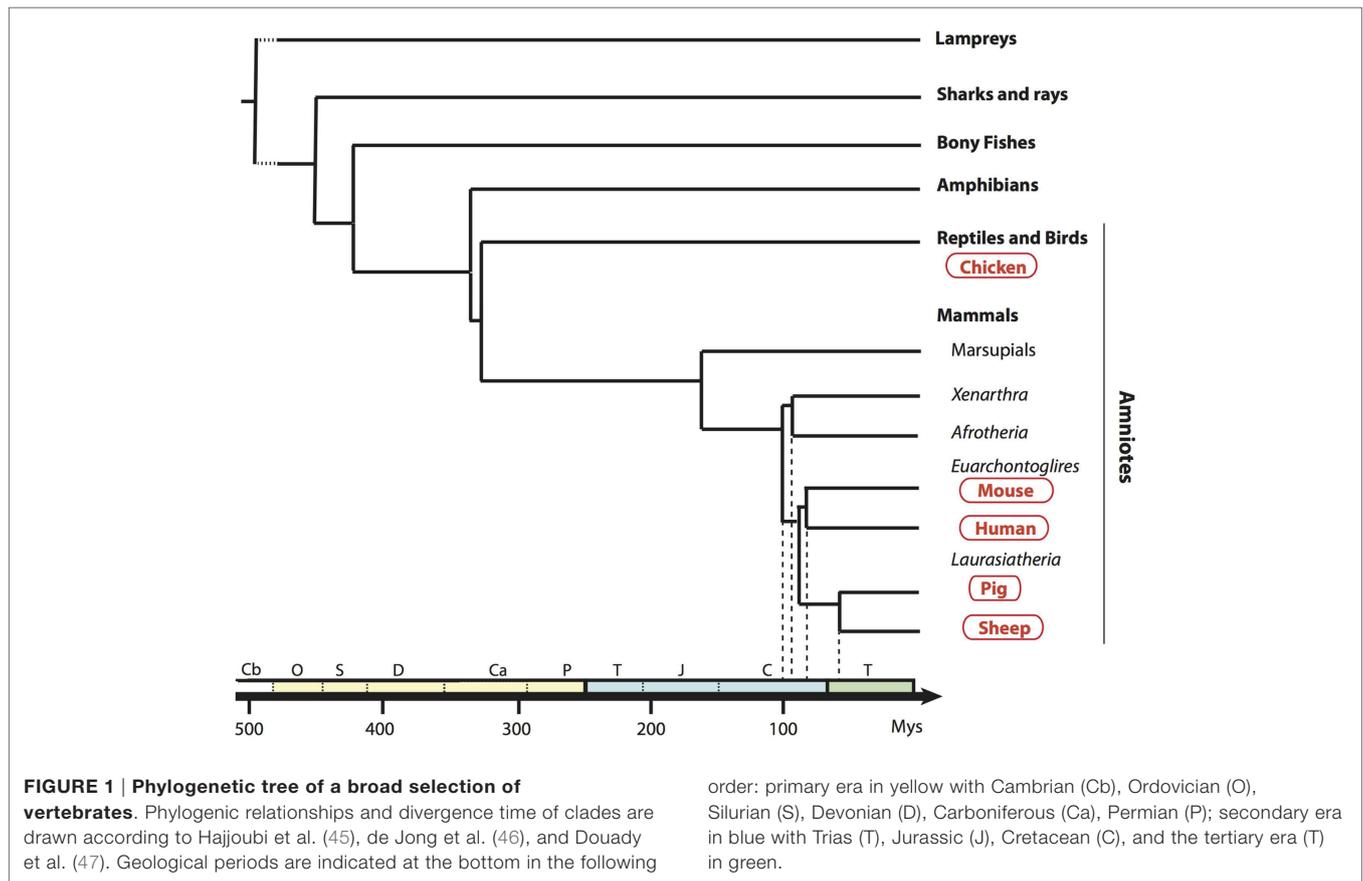For relative quantitation of gene expression in subsets, RNA was reverse transcribed using random primers and the Multiscribe reverse transcriptase (Applied Biosystems). Real-time PCR (qPCR) was carried out with 300 nM primers in a final reaction volume of 25 µl of 1 X SYBR Green PCR Master Mix (Applied Biosystems). The primers used to amplify ovine and porcine cDNA were designed with the Primer Express software (v2.0) using publically available GenBank sequences (Table S1 in Supplementary Material). PCR cycling conditions were 95°C for 10 min, linked to 40 cycles of 95°C for 15 s and 60°C for 1 min. Real-time qPCR data were collected by the Mastercycler® e0p realplex-Eppendorf system and $2^{-\Delta Ct}$ calculations for the relative expression of the different genes (arbitrary units) were performed with the Realplex software using GAPDH for normalization. All qPCR reactions showed >95% efficacy.

Results

Isolation of Mononuclear Phagocyte Subset Candidates from Artiodactyl Blood or Spleen Using a Set of Surface Markers

In order to establish a framework of homologous mononuclear phagocyte subsets across different species, we selected two mammalian artiodactyl species, sheep and pig, belonging to the Laurasatherians, a phylogenetically distant order from the Euarchontoglires that include the human and mouse species (**Figure 1**). A set of available antibodies exist to isolate cell subsets in these species of interest as food animals, hosts of zoonotic diseases, and biomedical models. We focused on blood or spleen immune subsets, because (i) a large source of transcriptomic data is available from this compartment in the human and mouse reference species, (ii) they are readily accessible with a minimum of technical biases in all species, and (iii) their gene-expression profiles are not expected to be influenced by peripheral tissue imprinting. We designed antibody panels to sort the subsets. In human and mice, cDC lack expression of T and B lymphocyte and Mo/MP markers and they abundantly express CD11c and MHC class II. Independent groups identified SIRPα as a conserved marker suitable to distinguish cDC2 from cDC1 across species (17, 34). Whereas XCR1 stands as the best marker for identifying cDC1 (34–41), appropriate reagents are not yet available in species outside human and mouse, and CADM1, whose sequence is highly conserved in evolution (42), can be used as an alternative (43, 44). cDC1 and cDC2 candidates were isolated from sheep and pig low density blood cells after exclusion of irrelevant cells (**Figure 2A** for sheep and **Figure 2B** for pigs, see Material and Methods section). The «candidate» nature of a sorted cell subset is marked by a star before the considered subset name in this paper. Due to restricted reagent availability, CD11c and MHC2 class II markers were used to isolate sheep and pig cDC, respectively.

In the case of pig, two populations being CADM1⁻ CD172⁺ or CD172^int were identified and selected as potential candidates and designated as *cDC2 and **cDC2, respectively (**Figure 2B**). We previously published the marker phenotype, morphology, and type I IFN production properties of sheep lymph and blood *pDC as CD45RB⁺ FSC^high TCRγ/δ⁻ B⁻ CD11b⁻ cells (26, 48). The sorted cells were very potent at type I IFN production upon viral-type stimulation, demonstrating at the functional level that they were highly enriched in pDC. Moreover, the sorted cells

FIGURE 1 | Phylogenetic tree of a broad selection of vertebrates. Phylogenic relationships and divergence time of clades are drawn according to Hajjoubi et al. (45), de Jong et al. (46), and Douady et al. (47). Geological periods are indicated at the bottom in the following order: primary era in yellow with Cambrian (Cb), Ordovician (O), Silurian (S), Devonian (D), Carboniferous (Ca), Permian (P); secondary era in blue with Trias (T), Jurassic (J), Cretacean (C), and the tertiary era (T) in green.

had the expected size and plasmacytoid morphology, indicating that they were not contaminated by other types of myeloid cells (48). Blood *pDC from pigs were sorted as CD3⁻ CD4⁺ CD172ᵈⁱᵐ cells based on marker phenotype, morphology, and type I IFN production properties established by others (25, 49). Pig *Mo were sorted as CD172ʰⁱᵍʰ MHC class II⁻ cells and sheep splenic *MP as CD14⁺ cells.

To decrease the risk of improper identification of sorted cell subsets, we performed a quality control consisting in examining the expression of a few control genes by qRT-PCR (**Figure 2C**) prior to performing genome-wide transcriptomic analyses. Control genes were chosen based on their high selective expression in a given subset of mononuclear phagocytes in a conserved manner between mouse and human (15, 36) and encompassed *TCF4* for pDC, *CD14* for Mo/MP, *FLT3* for cDC and pDC, *ZBTB46* for cDC, *BATF3* and *XCR1* for cDC1. As expected, *TCF4* was expressed to much higher levels in sheep (26) and pig *pDC as compared to all other cell types examined except for pig **cDC2. *CD14* was expressed at much higher levels in sheep and pig *mono/MP as compared to all other cell types examined except one of the two replicates of pig *cDC2. *FLT3* was expressed at much higher levels in sheep and pig *cDC1 and in sheep *cDC2 as compared to all other cell types examined. *BATF3* and *XCR1* were expressed at higher levels in sheep and pig *cDC1 as compared to all other cell types examined. Importantly, these control analyses have allowed us to improve

our initial strategy for sheep *cDC1 and *cDC2 sorting. In fact, in our initial sorting (Figure S3 in Supplementary Material), the CD45RB⁺ cells were not excluded to sort cDC candidates, and the *cDC1 were found to express high levels of *TCF4* mRNA, leading us to refine the sheep cDC sorting as presented in **Figure 2**. Thus, overall these control analyses validated our strategy for phenotypic identification and flow cytometry purification of sheep and pig *pDC, *Mo/MP and *cDC1 DC, and of sheep *cDC2. In the case of pig cell subsets, the nature of **cDC2 and *cDC2 was not clear since the former expressed high levels of *TCF4* and *XCR1*, and the latter expressed relatively high levels of *CD14* in one out of two replicates. Because pig **cDC2 presented a relatively high expression level of both *TCF4* and *XCR1*, we concluded that they were significantly contaminated by pDC and cDC1. Therefore, we excluded these cells from further analyses and assumed that pig *cDC2 cells were the proper candidate.

Use of Pairwise Gene Set Enrichment Analyses for Assessment of the Similarity Between Mononuclear Phagocyte Subsets Across Distant Mammal Species

As a first approach to establish mononuclear phagocyte subset homology across species, we determined the level of similarity between artiodactyl, mouse, and human mononuclear phagocytes using pairwise GSEA, as previously performed to characterize
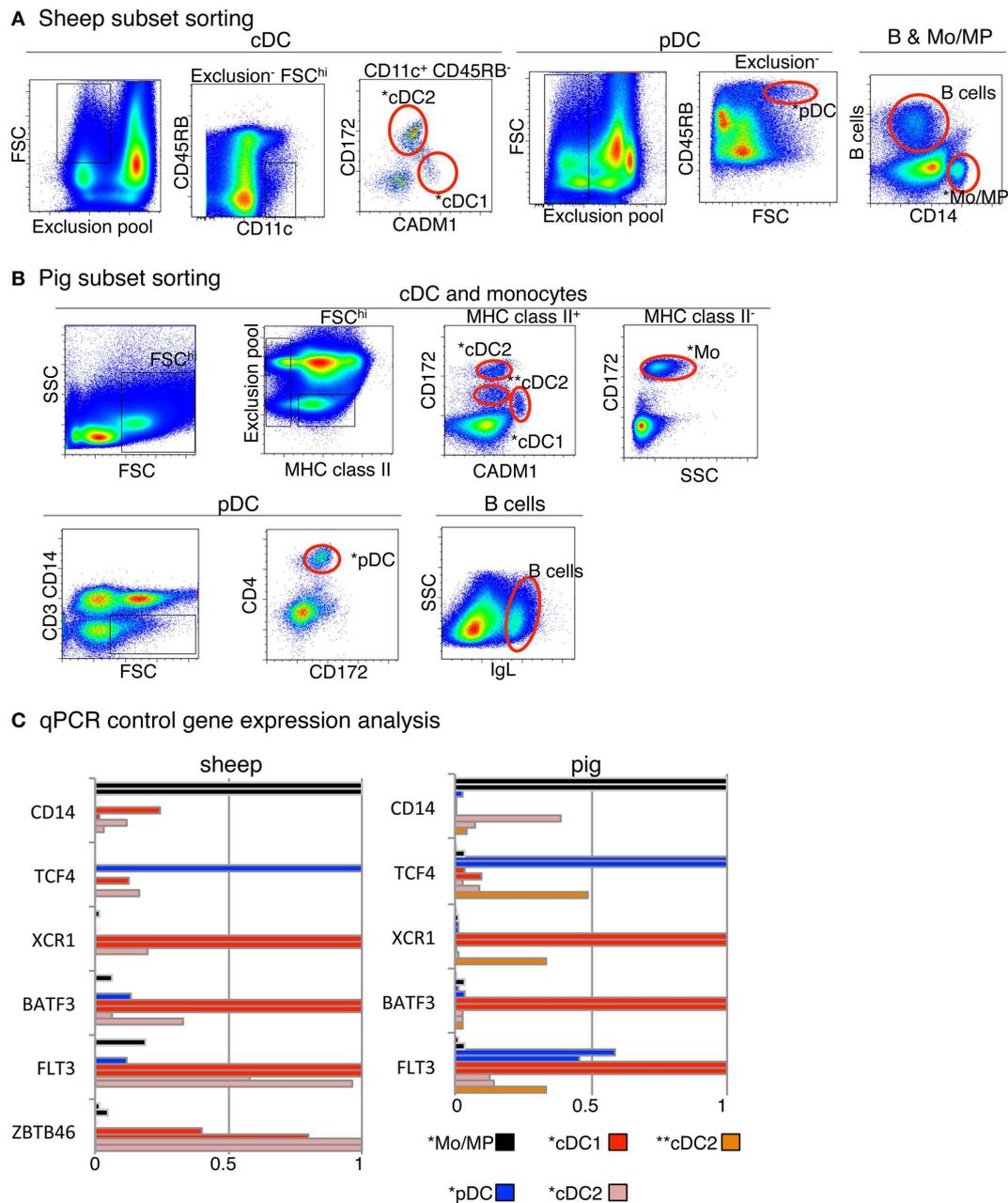
FIGURE 2 | Sorting of B cells, DC subset candidates, and Mo/MP candidates from pig and sheep blood or spleen and analysis of their expression of control genes. (A) Sheep cell subset sorting from blood and spleen. For sorting of blood cDC subset candidates, low density blood cells were gated on FSC^hi CD11c+ B− CD11b− TCR1− CD45RB− cells and analyzed for CADM1 and CD172 expression, based on isotype control references for each staining. The CADM1^hi CD172^lo (*cDC1) and CADM1^lo CD172+ (*cDC2) cells were sorted. Blood pDC candidates (*pDC) were sorted as low density FSC^hi B− CD11b− TCR1− CD8− CD11c− CD45RB+ cells. Splenic candidate *Mo/MP were sorted as CD14+ cells. Splenic B cells were identified as DU-2-104+ cells. (B) Pig cell subset sorting from blood. For cDC candidate sorting, low density PBMC were gated on FSC^hi MHC class II+ B− CD14− CD4− cells and analyzed for CADM1 and CD172 expression. One cDC1 candidate population was identified and sorted, as

CADM1+ CD172^lo (*cDC1). Two cDC2 candidate populations were identified and sorted, as CADM1− CD172^hi (*cDC2) and CADM1− CD172^int (**cDC2). Candidate Mo were sorted as CD172+ MHC2− cells (*Mo). Candidate pDC were sorted as CD3− CD14− CD4+ CD172^int cells (*pDC). B cells were identified and sorted as IgL+ cells. (C) qPCR analysis of the expression of control genes in sorted candidates from one or two animals. RNA from candidate cell subsets (left, sheep; right, pig) were subjected to detection of control transcripts by qPCR. Control transcripts were chosen based on their high selective expression in specific subsets of mononuclear phagocytes in a conserved manner between mouse and man, i.e., TCF4 for pDC, FLT3, BATF3 and ZBTB46 for cDC, XCR1 for cDC1, and CD14 for Mo/MP. Data are represented as relative expression levels normalized to maximal expression across cell types, each bar corresponding to a distinct animal.

human immune cell subsets (19) and chicken cDC (31). To that aim, we used publicly available transcriptomic data from a selection of human and mouse immune cell types (Data Sheet S1 in Supplementary Material). We established human and mouse transcriptomic fingerprints for B cells, pDC, cDC1, Mo/MP, MoDC, cMo, and ncMo as the list of genes that are expressed at least 1.5-fold higher in the index cell population than in a large number of other immune cell types (Data Sheet S2 in Supplementary Material). B lymphocytes were chosen in all species as a reference cell subset, because their phenotypic identification in each species and their homology across species are already well established, and because they are expected to share with mononuclear phagocytes a genetic program underlying their common function of antigen-presenting cells. We generated a common fingerprint for Mo and tissue MP because their gene program is very close in the mouse (9), even though tissue MP generally derive from embryonic precursors rather than from circulating blood Mo. We could not establish a human or mouse cDC2 transcriptomic fingerprint with a sufficiently large number of genes for subsequent reliable statistical analysis. We also defined relative transcriptomic signatures for cDC vs. Mo/MP as the list of genes that are 1.5-fold higher in all cDC relatively to Mo and MP from different tissues, and reciprocally (Data Sheet S2 in Supplementary Material). Finally, we identified transcriptomic fingerprints from human and mouse MoDC (12). We then tested whether the transcriptomic signatures of mouse and human immune cell types were enriched between sheep or pig candidate cell subsets using GSEA (33) (**Figures 3** and **4**). As control, since homologies between mouse and human cell subsets have been previously demonstrated by other methods of transcriptional analyses (15, 16, 18), mouse fingerprints were also used for GSEA analysis on human cells (Figure S4 in Supplementary Material) and reciprocally (Figure S5 in Supplementary Material).

As expected, sheep B cells were significantly enriched for the expression of both human and mouse B cell transcriptomic fingerprints as compared to all other sheep cell subsets examined (**Figure 3**, ❶). The sheep *pDC were enriched for the human and mouse pDC fingerprints in most comparisons (**Figure 3**, ❷), suggesting that sheep *pDC correspond to homologs of human and mouse pDC. However, both mouse and human pDC fingerprints were not significantly enriched in the comparison of sheep *pDC with *cDC2 (NES = 1.29 and 1.24, and FDR = 1.0 and 1.0, respectively), indicating that sheep pDC probably contaminate sheep *cDC2 despite exclusion of CD45RB+ cells for their purification. The sheep *pDC were also enriched for the human B cell fingerprint in most comparisons (except with sheep B cells), what can be partly explained by the known overlap between the gene-expression program of pDC and B (15, 50–52); however, the human pDC fingerprint is not enriched in the sheep *pDC comparison with B cells and the extent of the human B cell fingerprint enrichment in sheep *pDC is above the expectations provided by similar analyses in the human and mouse reference species (Figures S4 and S5 in Supplementary Material), all of this indicating that B cells are likely to contaminate sheep *pDC despite exclusion with a pan-B cell marker for *pDC selection. Finally, *cDC2 did not show a clear enrichment for any human

and mouse signatures (**Figure 3**, ❸). However, it is also the case when examining enrichment of mouse cell subset fingerprints in human cDC2 (Figure S4 in Supplementary Material, ❸) and reciprocally (Figure S5 in Supplementary Material, ❸). Hence, this GSEA approach is not very informative for identification of cDC2, due to the lack of robust human or mouse fingerprints that are specific of this cell type as mentioned earlier. Sheep *cDC1 were significantly enriched in the human and mouse cDC1 fingerprints in all comparisons (**Figure 3**, ❹). They were also enriched systematically in the mouse cDC vs. Mo/MP fingerprints. This suggested that sheep *cDC1 correspond to true homologs to human and mouse cDC1. Sheep splenic *Mo/MP were strongly enriched for the human and mouse Mo/MP vs. cDC fingerprints except when compared to MoDC, and not for the human and mouse fingerprints of B lymphocytes, pDC, or cDC (**Figure 3**, ❺). This confirmed that sheep splenic *Mo/MP belong to the monocytic lineage and not to the B nor DC lineages. However, their precise identity remained unclear as they were enriched for the mouse cMo fingerprint but not for the human cMo or ncMo fingerprints. When mouse fingerprints were applied on human immune cell subsets comparisons and vice versa, there was also no consistent alignment of ncMo between the two species (Figures S4 and S5 in Supplementary Material, highlights ❺ and ❻). Finally, sheep *MoDC that were derived from bone marrow cells in GM-CSF (27), were systematically and strongly enriched in the human and mouse MoDC signatures (**Figure 3**, ❻), confirming the homology between these three populations.

A similar analysis for pig candidate cell subsets also clearly established similarities with their putative human and mouse equivalents for B cells (**Figure 4**, ❶), pDC (**Figure 4**, ❷), and cDC1 (**Figure 4**, ❹) but not for cDC2 (**Figure 4**, ❸). Pig *Mo were clearly enriched for human and mouse fingerprints of cells of the monocytic lineage, and not for human and mouse signatures of B lymphocytes, pDC, or cDC (**Figure 4**, ❺).

Thus, altogether, GSEA analysis of the sheep and pig data for the fingerprints of human and mouse immune cell subsets gave results as informative as those obtained when comparing together human and mouse cell types, and clearly established similarities between sheep and pig cell subset candidates and their putative human and mouse equivalents for B cells, pDC, cDC1, and MoDC. Further analyses are necessary to precisely identify the nature of sheep and pig *cDC2 and *Mo/MP subsets.

Confirmation and Extension of the Conclusions on the Similarity Between Mononuclear Phagocyte Subsets Through Global and Simultaneous Analysis of the Gene-Expression Profiling of All Cell Types from Mammalian Species Using Hierarchical Clustering

In order to confirm the identification of homologous mononuclear phagocytes across species as deduced from GSEA analyses, and to potentially gain more insights into the exact nature of pig and sheep *cDC2 and *Mo/MP, we next processed all the data together for global analysis by hierarchical clustering (**Figure 5**). Only the genes that showed significant variation in their expression across subsets in each species were selected
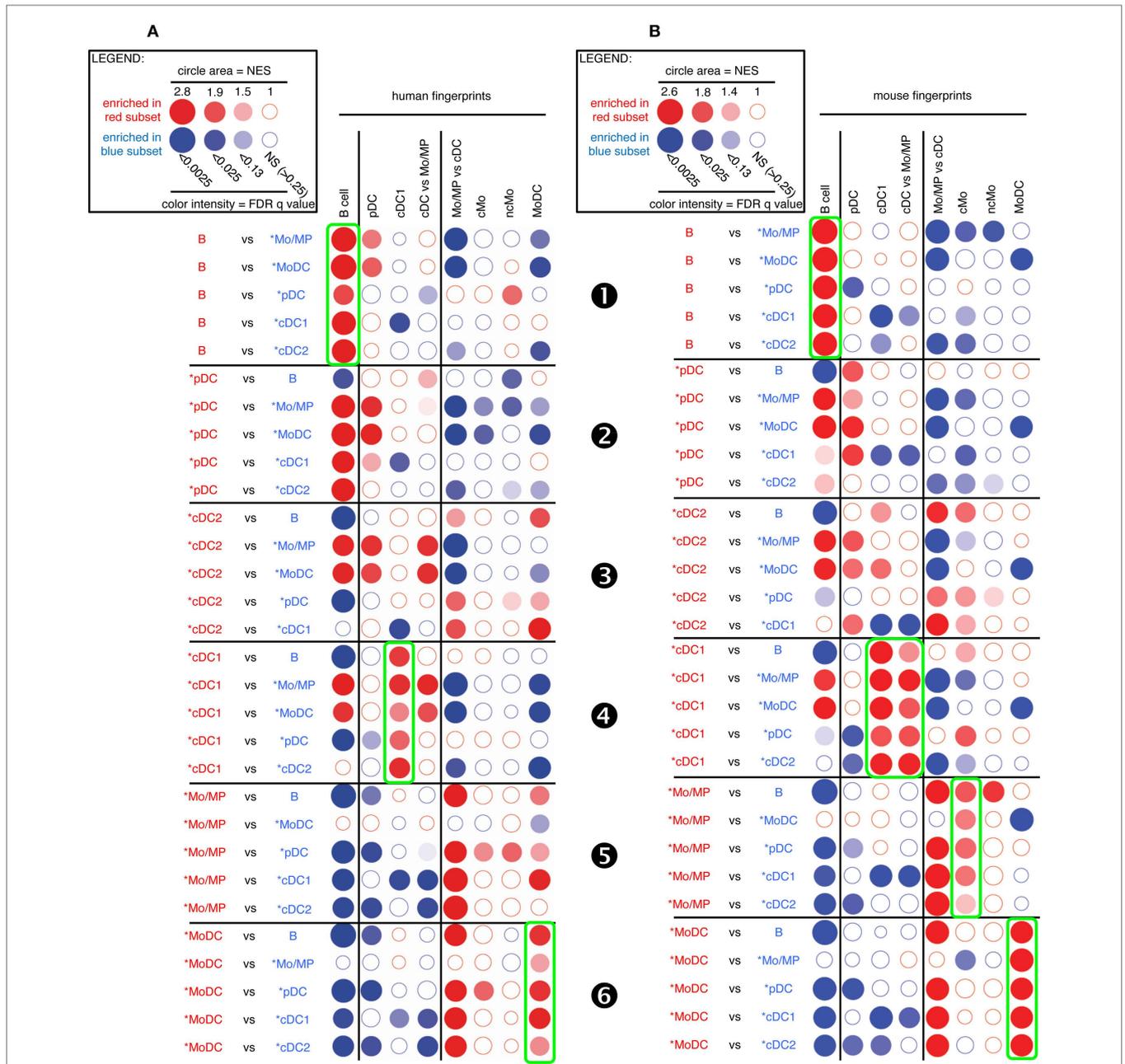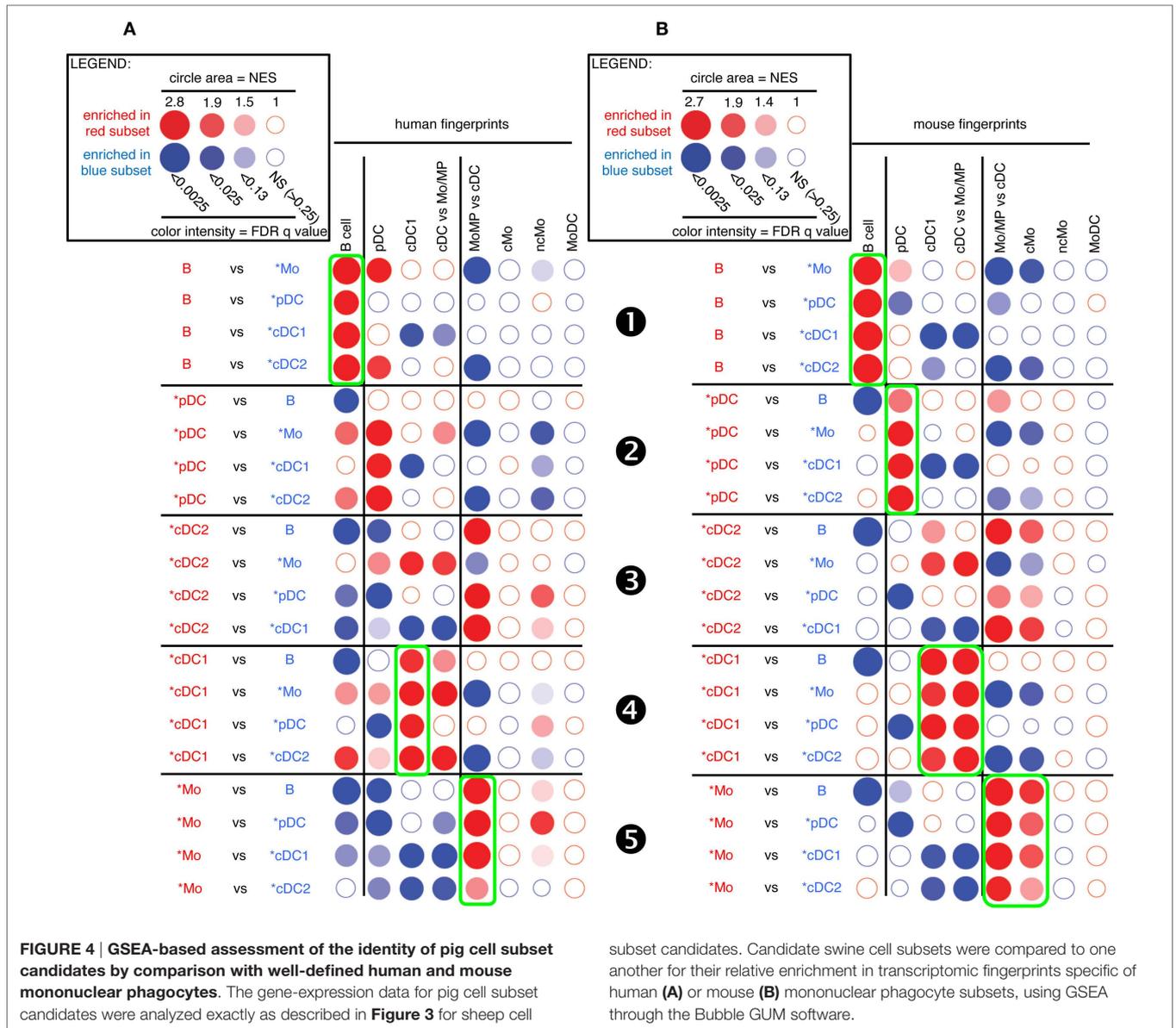
FIGURE 3 | GSEA-based assessment of the identity of sheep cell subset candidates by comparison with well-defined human and mouse mononuclear phagocytes. Candidate sheep cell subsets were compared to one another for their relative enrichment in transcriptomic fingerprints (GeneSets) specific of human (A) or mouse (B) mononuclear phagocyte subsets, using GSEA through the Bubble GUM software. The human and mouse GeneSets were defined through the same approach based on pre-existing knowledge of equivalency between human and mouse mononuclear phagocytes. A GeneSet specific for B cells was included as a control for the methodology, since the identity of this cell type is clearly established in all species and its homology across species is undisputed. The GeneSets used were named and defined as follows. The transcriptomic fingerprints "B cell," "pDC," "cDC1," "cMo," "ncMo," and "MoDC" consisted in the lists of human/mouse genes showing a high selective expression in the eponym human/mouse cell subset as compared to many other leukocytes

(see Materials and Methods for further details, Data Sheet S2 in Supplementary Material). The transcriptomic fingerprints "cDC vs. Mo/MP" and "Mo/MP vs. cDC" consisted in the lists of human/mouse genes expressed in cDC to higher levels than in Mo/MP, and reciprocally (Data Sheet S2 in Supplementary Material). All possible pairwise comparisons between sheep cell subsets were performed to assess their respective expression of the transcriptomic fingerprints of human and mouse mononuclear phagocyte subsets, using the Bubble GUM software for calculations and graphical output. Results are represented as bubbles, in a color matching that of the cell subset in which the GeneSet was enriched. Stronger and more significant enrichments are represented by bigger and darker bubbles, as illustrated in the legend box of the figure. Specifically, the surface area of bubbles is proportional to the absolute value of the normalized enrichment score (NES). The color intensity of dots is indicative of the false-discovery rate (FDR) statistical value.

FIGURE 4 | GSEA-based assessment of the identity of pig cell subset candidates by comparison with well-defined human and mouse mononuclear phagocytes. The gene-expression data for pig cell subset candidates were analyzed exactly as described in **Figure 3** for sheep cell subset candidates. Candidate swine cell subsets were compared to one another for their relative enrichment in transcriptomic fingerprints specific of human **(A)** or mouse **(B)** mononuclear phagocyte subsets, using GSEA through the Bubble GUM software.

and the resulting datasets were normalized across species. All B cells from the four mammalian species grouped together in a specific branch of the tree, rather than each with other immune cells of the same species. This finding validates hierarchical clustering as an alternative method for identifying homologous mononuclear phagocytes across species. A closer examination of the dendrogram shows that the different cell types grouped in two major branches. The first one encompassed all the known and candidate cells of the monocytic lineages and pig *cDC2, and split further into two subgroups, one including all the identified or candidate MoDC, and the other one including all the identified or candidate Mo/MP and pig *cDC2. The second branch encompassed all the other cell types known or hypothesized not to belong to the monocytic lineage. This branch further split into two sub-branches, one constituted of the group of B cells and of the group of identified or candidate pDC, and the other

constituted of identified or candidate cDC subsets except pig *cDC2. The common clustering of B and pDC transcriptome can be explained by the shared gene-expression program between B and pDC as mentioned above. Hence, this analysis confirmed the conclusion already drawn from the GSEA analyses, namely the monocytic nature of sheep and pig *Mo/MP and *MoDC, as well as the homology between pig, sheep, mouse, and human *pDC/ pDC. Moreover, the hierarchical clustering analysis allowed to better define the nature of sheep and pig *cDC2. Specifically, it confirmed the hypothesis that sheep *cDC2 belong to the cDC family, while, on the contrary to our *a priori* assignment, it shows that pig *cDC2 rather resemble Mo than cDC. However, within the branch of monocytic cells, this analysis grouped Mo/MP by species of origin rather than by cMo vs. ncMo subsets. Similarly, this analysis grouped cDC by species rather than by cDC1 vs. cDC2 subsets.
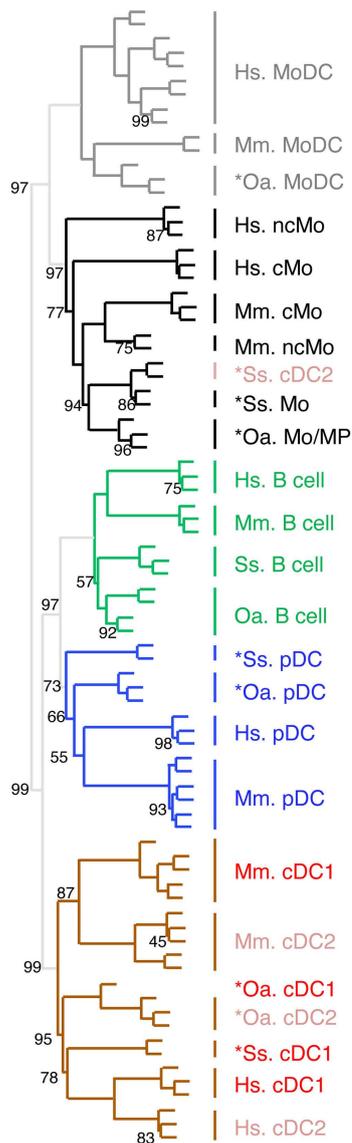
FIGURE 5 | Confirming and completing homology assignment of sheep and pig candidate Mo/MP, pDC, and cDC by unsupervised hierarchical clustering with human and mouse cell types. The datasets of each species were filtered and cross-normalized in order to allow mixing them all together for global analysis of the relationships between sheep, pig, mouse, and human mononuclear phagocyte subsets by using unsupervised hierarchical clustering. In brief, this analysis is focused on 1926 unique orthologous genes (i) for which a functional and specific ProbeSets was present on the microarrays for each species and (ii) which were found to be differentially expressed in each species between at least two subsets of mononuclear phagocytes. For each species and each of these 1926 genes, the expression data was then transformed to a mean = 0 and a variance = 1, in order to cross-normalize expression values to a similar dynamic range between the different datasets. For each cell type, the initials of the scientific name of the species of origin are indicated as a prefix: Hs, human; Mm, mouse; Ss, pig; and Oa, sheep. The robustness of the tree was tested by multiscale bootstrap resampling using Pearson's correlation as distance and average linkage as cluster method, with 1000 iterations at 10 different dataset sizes comprised between 50 and 140% of the complete dataset. An AU (approximately unbiased) p-value (percentage) was calculated and placed on the nodes of the cluster dendrogram. Missing percentages correspond to 100%.

Identification of Similarity Between Subsets of DC and of Mo Across Species Through Hierarchical Clustering Analyses Focused on These Cell Types

The expression patterns of genes outside of the cell types of interest may mask similarity between cDC or Mo subsets, as previously reported (15). Hence, we further evaluated the similarities between subsets of cDC on the one hand, and of Mo/MP on the other hand, by re-analyzing their gene-expression profiles focusing only on the genes that showed significant variation in their expression across DC subsets (**Figure 6**) or Mo/MP (**Figure 7**) in each species. Pig data were not used in the analysis focused on cDC, because, pig *cDC2 belonged to the monocytic branch and not to the DC branch of **Figure 5**. Sheep data were not used in the analysis focused on Mo/MP, because only one subset of sheep Mo/MP had been purified. Remarkably, these focused analyses grouped samples by cell types rather than by species. The cDC-focused hierarchical clustering confirmed the conclusion drawn from GSEA that sheep, mouse, and human cDC1/*cDC1 are homologs, and refined our understanding of the identity of sheep *cDC2 by showing their homology to mouse and human cDC2 (**Figure 6**). The Mo/MP-focused hierarchical clustering allowed to newly identify pig homologs to mouse and human cMo vs. ncMo (**Figure 7**). Pig *cDC2 correspond to ncMo and pig *Mo correspond to cMo. In a complementary phenotypic FACS analysis, we confirmed that likewise human ncMo as compared cMo, pig *cDC2 express higher membrane levels of CD16 and CD163 as compared to pig *Mo (Figure S6 in Supplementary Material).

Altogether, our comparative analyses of the gene-expression profiles of mononuclear phagocyte subsets across mammals
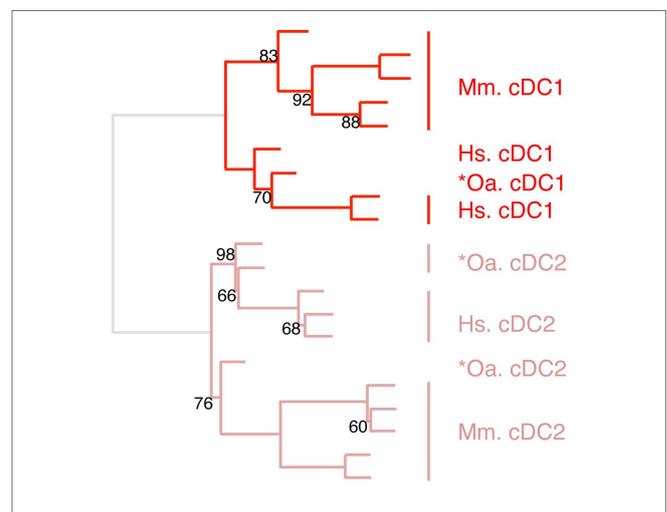

FIGURE 6 | Confirming homology assignment of sheep cDC1 and cDC2 candidates by unsupervised cross-species hierarchical clustering focused on cDC subsets. An unsupervised cross-species hierarchical clustering analysis was performed as described in **Figure 5**, but focused only on cDC subsets. The corresponding filtered dataset included 868 unique orthologous genes found regulated between cDC1 and cDC2 from human (Hs), mouse (Mm), and sheep (Oa). Pig cDC could not be included in this analysis due to the lack of data on proper pig cDC2.
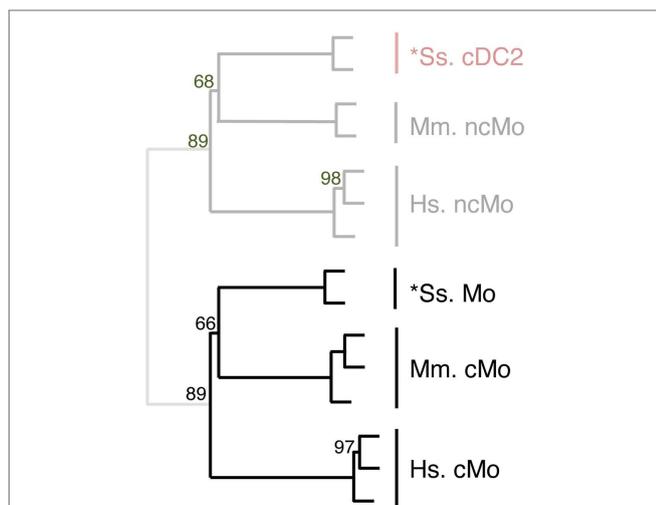
FIGURE 7 | Completing homology assignment of pig cDC2 DC candidate to non-classical Mo subset by unsupervised cross-species hierarchical clustering focused on Mo subsets. An unsupervised cross-species hierarchical clustering analysis was performed as described in **Figure 5**, but focused only on cells from the monocyte branch of the tree obtained in **Figure 5**. The corresponding filtered dataset included 191 unique orthologous genes found regulated between cMo and ncMo from human (Hs), mouse (Mm), and pig (Ss). Sheep data could not be included in this analysis due to lack of data on subsets of sheep monocytes.

indicated that the complex specialization of these cells into distinct subsets is conserved across mammals for both DC and Mo. Subset grouping did not indicate existence of a relationship between transcriptomic proximity of subsets and phylogenetic closeness of species. The conserved organization across distant mammals suggests that the mononuclear phagocyte complexity arose in a common mammalian ancestor and that the different subsets can be considered as homologous subsets across mammals.

Evidences for Homologous cDC and Mo/MP Lineages Across Warm-Blooded Vertebrates

We recently generated the transcriptomic profile of MP, total cDC, and B cells from chicken spleen and found similarities with human and mouse corresponding immune cell subsets by GSEA (31). In order to extend our subset homology analysis to non-mammalian vertebrates, we normalized and processed the transcriptomic data in a hierarchical clustering analysis as described above, using mammalian and chicken Mo/MP, B cells, and cDC subsets (**Figure 8**). There again, a tree consisting of two main branches was obtained, corresponding to a split between Mo/MP and B cells/DC. In the cDC branch, the cDC1 subset clustered together and included the chicken total cDC. The chicken MP grouped with the mammalian Mo/MP. Whereas this analysis is still partial due to limited knowledge and availability on marker sets for sorting immune cell subsets in chicken, it shows that our transcriptomic comparative approach can be used to define subset homology throughout vertebrates. It also further supports that separation of mononuclear phagocytes into Mo/MP

and cDC occurred early during vertebrate evolution and must already have been in place in the common ancestor of reptiles (including birds) and mammals.

Identification of Mononuclear Phagocyte Gene-Expression Signatures Across Mammals

Taking advantage of our multi-species microarray data, we sought to identify core gene-expression signatures that should universally define at the molecular level each of the mononuclear phagocyte subset and that should hold biological relevance based on their selective and conserved expression in homologous subsets throughout mammalian evolution. Absolute signatures ["Min (test) vs. Max (ref)" method, see Materials and methods] encompassed all genes selectively expressed at higher levels in the cell subset of interest (index population) as compared to all the other cell subsets studied (comparator populations), in all species studied. An absolute signature was computed for B cells in order to validate the approach by comparison of the gene list obtained with the advanced knowledge available on the biology of this lymphocyte population. Absolute signatures were also found for pDC, cDC1, and MoDC. Relative signatures encompassed genes selectively expressed to higher levels in one or several cell subsets of interest (index population) as compared to a selection of other cell subsets (comparator populations). The choice of index and comparator populations was largely based on the branching of different cell subsets in hierarchical clustering (**Figure 5**), or on known sharing of specific functions between cell subsets in mouse or human. The conserved absolute and relative gene-expression signatures in mononuclear phagocyte subsets are listed in **Table 1** and Data Sheet S3 in Supplementary Material. In several instances, Ingenuity Pathway Analysis (IPA) mapped a high proportion of the genes to gene interaction networks (**Figure 9** for the DC lineage subsets, **Figure 10** for the monocytic lineage subsets and Figure S7 in Supplementary Material), and revealed predicted upstream regulators (**Figure 11A**) and canonical pathways and functions (**Figure 11B**) that are described thereafter for B cells, DC lineage subsets, and Mo/MP categories. Although certain functions or pathways were enriched in several gene signatures, the genes responsible for the enrichments differed (Data Sheet S4 in Supplementary Material) and pointed out to different, complementary contributions of the distinct cell types to the corresponding functions or pathways.

The conserved B cell signature that we use as our reference subset (**Table 1**) includes a regulatory gene network directed to immunoglobulin production (Figure S7 in Supplementary Material), with PAX5 as an upstream regulator ($p = 10^{-5.8}$) (**Figure 11A**). SOX11 ($p = 10^{-8}$) and FOXO1 ($p = 10^{-7}$) are predicted to be other upstream regulators in the conserved B signature (**Figure 11A**), in agreement with existing knowledge. As expected, this signature is associated to B lymphocyte ontogeny and functions [e.g., "development of B lymphocytes" $p = 10^{-10.4}$, "antibody response" ($p = 10^{-7.5}$), "proliferation of B lymphocytes" ($p = 10^{-10.5}$), and "morphology of B lymphocytes" ($p = 10^{-7.5}$) as well as to the "B cell receptor signaling" pathway ($p = 10^{-7.3}$)] (**Figure 11B**). The B cell signature also pinpoints to genes without any known function in B cells yet, such as the cell cycle gene
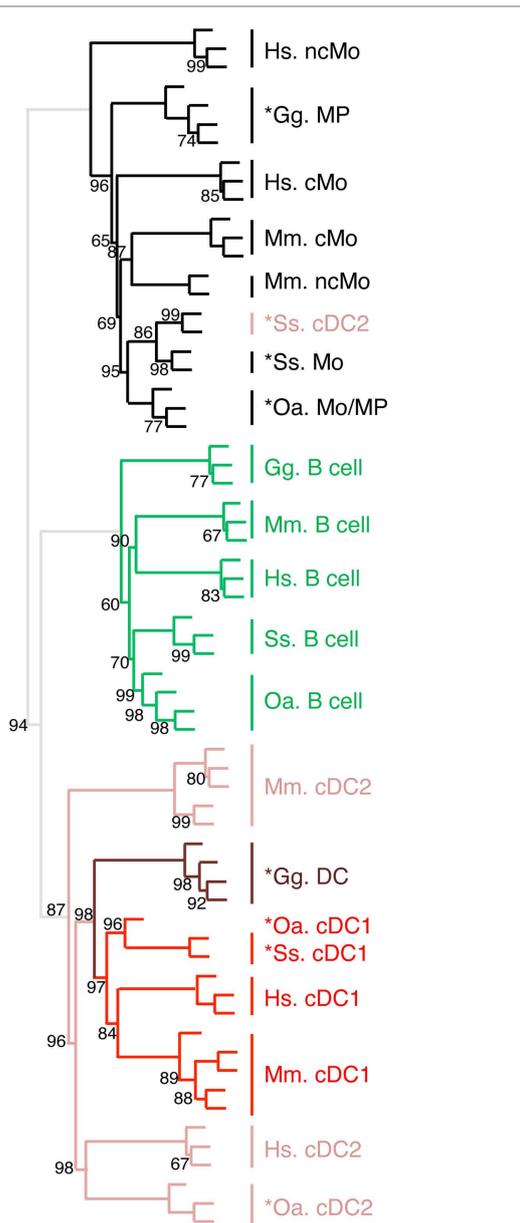
FIGURE 8 | Unsupervised cross-species hierarchical clustering including a chicken dataset demonstrates a conserved organization of vertebrate mononuclear phagocytes in the two main lineages of Mo/MP vs. cDC. An unsupervised cross-species hierarchical clustering analysis was performed as described in **Figure 5**, but including gene-expression data from chicken (Gg prefix for *Gallus gallus*) and focused only on the cell types commonly sorted in all five vertebrate species, i.e., B cells, Mo/MP, and cDC. The corresponding filtered dataset included 388 unique orthologous genes found regulated across cell subsets in each species.

RAD17 (53) or the *SP140* gene that encodes a nuclear body protein (54) (**Table 1**). Altogether, the results of the functional analysis of the conserved signature of B cells support the biological relevance of the conserved gene signatures generated by our approach.

In the conserved signatures corresponding to the DC lineage, the pDC signature is restricted to few genes including *RUNX2*,

which encodes for a major known regulator of pDC development (56) and other genes whose role is not yet known in this subset, with three of them coding for potential cell surface markers or targeting molecules, i.e., the low density lipoprotein receptor-related protein 8 (*LRP8*), tetraspanin 13 (*TSPAN13*), and a zinc-family transporter protein member (*SLC30A5*) (**Table 1**). These genes, except *SLC30A5*, map to a common network (**Figure 9A**). No functional annotation was found significantly enriched in the pDC absolute signature due to the low number of associated genes. Interestingly, the pDC vs. cDC relative signature includes genes belonging to a regulatory network pointing to IFN $-\alpha/\beta$ production (**Figure 9B**) and retrieves as a major putative upstream regulator X-box binding protein 1 (*XBP1*) ($p = 10^{-15}$) (**Figure 11A**), a transcription factor involved in mouse DC development (57). The pDC vs. cDC relative signature was also enriched for "proliferation of B lymphocytes" ($p = 10^{-4}$), "morphology of B lymphocytes" ($p = 10^{-5}$), and "B cell receptor signaling" pathway ($p = 10^{-2.9}$), similarly to the conserved B cell signature (**Figure 11B**). These observations are consistent with the known usage downstream of mouse and human pDC endocytic receptors of a signaling pathway akin to that of the B cell receptor (58). This known pDC signaling pathway involves the products of *SYK*, *BLNK*, and *PIK3AP1*, three of the six genes responsible for the enrichment of the "B cell receptor signaling" pathway in the conserved pDC vs. cDC gene signature (Data Sheet S4 in Supplementary Material), as well as *CARD11* which contributes to the enrichment for the annotation "proliferation of B lymphocytes" in the pDC vs. cDC signature. This strongly suggests that this signaling pathway is conserved in pDC of all mammalian species. Beside *TCF4* which encodes for a major known regulator of both B and pDC development (52), several other genes associated to B cell biology are found in the pDC vs. cDC relative signature (**Table 1**), namely *CD79B*, *PTPRCAP*, *SEMA4D*, *CTCF*, *IFR1*, and *MEF2C*. This suggests that additional biological processes shared between B cells and pDC remain to be identified.

No absolute signature could be generated for cDC but interesting informations were obtained with relative signatures, i.e., the cDC vs. Mo/MP and cDC vs. pDC. The cDC vs. Mo/MP signature includes *FLT3*, a key gene in mouse DC development (59) as well as many genes of a regulatory network including *BCL11A*, *HLA-DOA*, *HLA-DRA*, *HLA-DMB*, *HLA-DOB*, *CD74*, the axone guidance neuron navigator *NAV1* and the MHC class 2 transcription regulator *RFX5* (**Figure 9C**). In relation to this network, *IL27* ($p = 10^{-4.4}$), *IFNG* ($p = 10^{-2.6}$), and NFkB ($10^{-2.8}$) were retrieved as putative upstream regulators (**Figure 11A**). The cDC vs. Mo/MP signature was enriched for canonical pathways such as "antigen presentation" ($p = 10^{-9.6}$), "DC maturation" ($p = 10^{-4.6}$), and "T helper cell differentiation" ($p = 10^{-6.2}$) (**Figure 11B**). The cDC vs. pDC signature includes a main regulatory network encompassing *PIK3CB*, *ICAM1*, *CLEC7A*, *HLA-DRA*, *IL1B*, and *LGALS3* (**Figure 9D**) and is enriched for "functions of antigen-presenting cells" ($p = 10^{-11.1}$), "inflammatory response" ($p = 10^{-8.9}$), "bacterial infection" ($p = 10^{-7.9}$), "migration of cells" ($p = 10^{-11.4}$), and "clathrin-mediated endocytosis signaling" pathway ($p = 10^{-4.6}$) (**Figure 11B**). TNF ($p = 10^{-9}$), RELA ($10^{-5.8}$), NFKB1 ($10^{-6.1}$),

TABLE 1 | Conserved gene signatures for mammalian mononuclear phagocytic cell subsets.

Cell subset gene signatures	genes conserved in 3/3 or 4/4 species[a]	genes conserved in 2/3 or 3/4 species[b]
B cell	*TRAF5; SP140; RAD17; MEF2C; MBD4;* **FCRL1**[c,d,e]*;* **CD19**	**VPREB3**; *RFX5;* **PAX5**; *BACH2; AFF3; SWAP70; PLEKHA2;* **MS4A1**; *DMXL1;* **CR2**; **CD79B**; *CD22;* **BLK**; *ELL3; STRBP;* **EBF1**
cDC vs Mo/MP	*NAV1; MSI2;* **HLA-DMB**; **FLT3**; *BCL11A*	*RFX5; PLEKHA5;* **HLA-DOA**; *BCAT2; AFF3; FAM149A; APOBEC3H; UVRAG; SPINT2; PDXP;* **HLA-DOB**; **CD74**; *CD5; AP1S3;* **HLA-DRA**
cDC vs pDC	*WDR41; WDFY3; TPM4;* **TLR2**; *SPI1; SNX14; SNX10; SERPINB1;* **SAMHD1**; *RIN3; REL; RAB32; NHSL1; NCOR2; NAV1; MARCKS; LYZ; LGALS3; KLF3;* **JAK2**; *ITGA5; IL4I1;* **IL1B**; **IFNGR1**; **IFI30**; **ID2**; **ICAM1**; **HLA-DMB**; *GCA; FGL2; F11R; ETV3; DOCK7; DENND4A; CXCL16; CLEC7A; CHSY1;* **BATF3**; **ATP2B1**; **ARRB1**; *ARHGAP22;* **ANPEP**; **AIM1**; *AIF1; AHR; ADAM8*	*YWHAH; TPCN1; TDRD7; SNX21; SLC7A10; SIPA1L3; RGS12; MYO1D; MRC2; METRNL; MEA1; LRRK2; LRRC8C; LOXL3;* **HLA-DQB2**; *HAVCR2; FGF17; EHF; DOK1; DGKH; ATXN1; ASB2; ARHGAP26;* **ACTR3**; *RNF144B; PLEKHO2; MYOF; LPCAT2; KANK1; FAM114A1; DENND5A; ZNF524; VASP; SULT1A1; SPRED1; SNX8; SH3BP1; SH3BGRL;* **RELB**; *RALB; RAC1; PTPN12; PLEKHO1; PIK3CB; PAK1; NR4A1; NAB2; LFNG; JUNB;* **IFNGR2**; *IER2; HFE; FAM49B; EPSTI1; EGR1; EFHD2; DHRS3; CTBP2; COTL1;* **CD74**; *CD63; CBFB; C9ORF72; C1ORF21;* **BCL6**; *BASP1; ANXA5; SR140; PKM2;* **HLA-DRA**; *RGS4; TMSB4X; GMIP; MAST2; CXCL9; DNAJA4; KIF14; MTUS1; RABGGTA; RTN1; SYNJ1; TBX3*
DCs vs (Mo/MP & MoDC)	*MSI2; BCL11A*	*RAB34;* **PDCD1LG2**; *CHCHD7; CCL17; CARM1; AUH;* **VEGFA**; *UBA3; TUBA1A; TSKU; TMEM159; SLC48A1; SIGMAR1; RNF181; PTGR1;* **NOS2**; *IKBIP; FAM162A; BHLHE40*
MoDC	*TPI1; NDUFV2;* **FCGR2B**; *CD200R1; ALDOA*	
MoDC vs Mo/MP	*TPI1; SLC2A1; SLAMF1; PRNP; PPA2; POLR1D; PLAU; PALLD; NDUFV2; NARF; MRPL4;* **IL1R2**; **FCGR2B**; *EGLN3; DGKA; CSNK2B; CISH; CD200R1; AVPI1; ALDOA; ADAMTSL4*	*ZNF747; ZNF219; WIBG; VDR; SLC45A4; ROGDI; RASSF7; RAB34; RAB33A; PDE6D;* **PDCD1LG2**; *PBX2; NAGS; KCNK6;* **ICOSLG**; **HRH1**; *GOLGA8B; GOLGA8A; ETHE1; ERCC6; DVL2; DGUOK; CLEC10A; CHN2; CHCHD7;* **CD209**; *CCNG2; CCL17; CARM1; C1ORF122; AUH; ANKRD37; ZEB1;* **VEGFA**; *UBA3; TUBA1A; TSKU; TMEM159; TCTEX1D2; STRA13; SPATA24; SNRNP27; SLC48A1; SIGMAR1; S1PR3; RNF181; RMND1; RAB7A; PTGR1; PIGU; PI4K2A; OST4; NSL1;* **NOS2**; *NAE1; MT1A; MORN4; LMF2; JKAMP; IKBIP; IFT46; HAUS4; GLTPD1; GATC; FAM162A; FAM13A; FAM134A; ESYT1; ERI2; EEPD1; DNLZ; DHRS11; DCTPP1; CENPW; BHLHE40; APOO; AKIP1; CD1B; CGREF1; NOSTRIN; OLFM4;* **GAS6**; *SLC27A3*
(Mo/MP & MoDC) vs DCs	**CEBPB**; *CCDC93;* **C5AR1**	**TLR8**; *FTL; DOK3;* **CD68**
Mo/MP vs cDC	**TLR4**; **SOD2**; *RBMS1;* **LAMP2**; *GLUL; FNDC3B;* **CYBB**; **CEBPB**; *CCPG1; CCDC93;* **C5AR1**	**TLR8**; *SNX27; RHOQ; OSTM1; KIF1B; FTL; DUSP6; DOK3;* **CTSD**; **CTSB**; **CD68**; *HERC5; IPMK; DPYD*
Mo/MP vs MoDC	*WDR33; VPS13D; UBE2D2; TRA2A; STAG2; SFPQ; NSD1;* **NFKB1**; *NADK; ITPR1; CFLAR; ARFGEF1*	*ZNF407; VPS13C; USP31; SLC16A4; SKAP2; PRKCH; PPFIA1; PIAS2; MDN1; MAP3K5; LRRC8D; CHM; AKAP13; ACTR3; SFRS2IP; RAD51L1; NAT12; MYST3; CDC2L5; ZNF830; ZBED5; TPPP3; TMEM164; TGS1; TBC1D8B; SNRNP35; SMEK1; SLC38A10; SHISA2; RSRC2; REV1; RALGAPB; PWWP2A; PRRC2B; PBRM1; NLRC5; MOGS; MAP7D1; LUC7L3; LIMCH1; KDM4C; ISY1; IP6K1; HNRNPUL2; HNRNPU; HNRNPK; HNRNPH2; HNRNPH1; HNRNPD; HNRNPA1; FOXN3; FAM173B; FAM159B; ERVW-1; CELF2; C9; NUP210L; PDZK1; ALMS1; LAMB1; METTL3; PAIP1*
pDC	**RUNX2**	*LRP8; INPP4A; TSPAN13; SLC30A5; GPM6B*
pDC vs cDC	*UBR2; UBE2H; TMED3;* **TCF4**; *TARBP1;* **SYK**; *STT3B; SPCS3; SNX5; SLC39A7; SIT1; SEMA4D; SEC61A1; SCYL3; SCAMP2; SAP30BP;* **RUNX2**; *RDH11; RASGRP2; RABAC1; PPAPDC1B; PGM3; PARN; PAG1; OGT; NUCB2; MSI2; MEF2C; LMAN2; IQCB1; IFT52; HBS1L; GPAM; GORASP2; FKBP2; FAM3C; EIF2AK3; DERL1; DDOST; DAD1; CYBB; COPA; CDC42SE2; CD4; CD164; BTRC;* **BLNK**; *BCL7A; ATP2A3; ATG5*	*ZXDC; VPS13A; UEVLD; TNRC6B; TMEM63A; TAF9B; TAF1A; SUSD1; STOML1; ST6GALNAC4; SSR2; SRPRB; SPG20; SLC38A6; SLC38A1; SLC25A36; SGCB; SERPINI1; SEC24C; SAP130; RAPGEF2; RALGPS1; RAB28; RAB11FIP2; PTAR1; PIK3AP1; OSTM1; NRP1; MYB; MGAT4A; MCOLN2; MCOLN1; LRP8; KIF13B; KIAA0226;* **IRF7**; *INPP4A; IMPACT; HIVEP1; FKBP8; FANCD2; FAM122B; DMTF1; CSTF1;* **CREB3L2**; *COBLL1; CBX4; CANX; ATG4D; ANKRD28; ANKIB1; AGBL3; AFF3; TPRG1L; RNF144A; IFI27L1; FAM65B; ELMOD3; DCAF7; CARS2; ZMYND11; YPEL3; USP24; TUBGCP6; TSPAN13; TRAM1; TOE1; TMEM138; TM9SF1; TCTA; SURF4; STAMBPL1; SSR3; SPCS2; SPATA13; SNX9; SLC7A5; SLC44A2; SLC30A5; SEPP1; SCAND1; SCAMP3; RHOH; RHBDF2; RHBDD1; REXO2; QDPR; PYCR2; PTPRCAP; PRMT7; POLD1; PEX5; NSUN3; MTMR9; LPGAT1; INTS7; IFNAR1; HM13; GRAP; GANAB; FNDC3A; FASTK; EXOC7; ELOF1; ELMOD2; CTCF; COPE; COMMD6; CNP; CIRBP; CDS2; CD79B;* **CARD11**; *C19ORF10; C16ORF80; C10ORF88; BTD; BET1; ARHGAP12; AHI1; WDR51B; SAPS3; MLF1IP; KIAA1370; CYBASC3; CEP110; CCDC111; ANUBL1; MME; PTPRS; ATF2; GPM6B; MON2; PPM1A; TM7SF3; TMCO1; UGCG; ZDHHC14; ZNF521; TMED10; PAIP1*
cDC2		*FCER1A*
cDC1	**XCR1**; *WDFY4; FNBP1;* **FLT3**; **CADM1**	*SNX22; GCET2*
cDC2 vs (pDC & cDC1)	*TRPS1; STK24; SLC16A3;* **SIRPA**; *SIGLEC8; S100A4; RIN2;* **REL**; **PILRA**; **NFAM1**; *NCF2; MAFB; LRP1; ITGAM;* **IL1R2**; **IL1B**; *IGSF6;* **IFI30**; *FHL3; EPB41L3; DOCK4; DHRS3;* **CSF3R**; **CSF1R**; *CLEC4A;* **CD300A**; *C19ORF59; ADRBK2; TREM1*	*TNFRSF1B;* **TLR8**; *TICAM2; STK10; SP2; SLFN12; SIGLEC9; SIGLEC7; RNASE2; PHF21A; LST1; LIMD2; CLEC6A;* **LILRB2**; **LILRB1**; **LILRA6**; **LILRA3**; *IFITM2; GNGT2; GBP4; FAM111A; EMR1; DPP10; DENND1A;* **DDX58**; *CDKN2B; CD300LF; CD300LB;* **CD209**; *C10ORF11; ADAP1; CLEC6A; DAGLB; WDR45L; SIGLEC5; SFRS5; S100A12; PLEC1; MYST1;* **MX2**; **MS4A8B**; *LRRC33; HSPA6; GK3P; GAPDH; FAM45B; CEBPD;* **CD1E**; *CD1B; FCER1A; KSR1;* **OAS2**; *PTGER3*

TABLE 1 | Continued

Cell subset gene signatures	genes conserved in 3/3 or 4/4 species[a]	genes conserved in 2/3 or 3/4 species[b]
cDC2 vs (pDC & cDC2)	*XCR1*; *WDFY4*; *ST3GAL5*; *RAB32*; *PPT1*; *PPA1*; *LRRC1*; *KIAA1598*; *FNBP1*; *FLT3*; *CALM1*; *CADM1*	*SNX22*; *PPAP2A*; *PLEKHA5*; *GRAMD2*; *DENND1B*; *CLEC1A*; *ATXN1*; *FAM114A1*; *HEPACAM2*; *PI4K2A*; *PLEKHO2*; *WDR91*; *TRIO*; *RALB*; *PKP4*; *PDLIM7*; *G3BP2*; *BCL6*; *ATPIF1*; *GCET2*; *BRWD2*; *FGD6*; *MYO9A*
ncMo vs cMo[f]	*ACAT2*; *ACE*; *ACOT9*; *ADRBK2*; *ANKRD42*; *APOA2*; *ASB2*; *BDKRB2*; *BGLAP*; *C1ORF112*; *C1ORF56*; *C20ORF112*; *CAPZB*; *CBX4*; *CD4*; *CD83*; *CDH24*; *CHD5*; *CSF1R*; *CYP2R1*; *DCBLD1*; *DDB2*; *DDIT4*; *DLGAP4*; *FBP1*; *GABBR1*; *GLMN*; *GNE*; *GNPNAT1*; *GPT*; *GRHPR*; *HEY1*; *HN1*; *IL12RB1*; *IL17A*; IL2RG; *KCNMA1*; *KCTD11*; *KNDC1*; *LMX1B*; *LUZP1*; *MAFF*; *MPZL1*; *MUTYH*; *MYOD1*; *NCAPH2*; *NCOR2*; *NFKBIA*; *NPAS2*; *NUB1*; *PCK1*; *PDCD4*; *PGR*; *PITPNM1*; *PLEKHH1*; *PMF1*; *PMVK*; *POLR3H*; *RAB25*; *RAD52*; *RFC5*; *RHOF*; *RSAD1*; *RWDD3*; *SECISBP2*; *SERPINA1*; *SH2D3C*; *SIRT5*; *SLC37A1*; *SMS*; *ST3GAL1*; *ST3GAL5*; *TBC1D8*; *TCF7L2*; *TNNC1*; *U2AF1L4*; *UNG*; *WDR76*	
cMo vs ncMo[f]	*AACS*; *ABHD5*; *AGTPBP1*; *ALDH2*; *ALOX5AP*; *ANXA1*; *AOAH*; *ARL8B*; *ATP6V1A*; *ATP6V1B2*; *ATP6V1C1*; *AUH*; *B4GALT1*; *C19ORF59*; *C5ORF15*; *CCR1*; *CD164*; *CD84*; *CETN2*; *CLTA*; *COPB2*; *CSF3R*; *CYP27A1*; *DCLRE1A*; *DNAJC10*; *ECE1*; *EHD4*; *EIF2AK2*; *EIF2AK3*; *ENSA*; *ENTPD7*; *ERP29*; *EXOC5*; *F13A1*; *F5*; *FAM102B*; *FAM63A*; *FBXL5*; *FBXO9*; *FN1*; *GBE1*; *GNA12*; *GNPAT*; *GSN*; *GYS1*; *HMGB2*; *IL1R2*; *IL1RN*; *ITM2B*; *KEAP1*; *LACTB*; *LCN2*; *LEO1*; *LMAN1*; *LMNB1*; *LYZ*; *MBD5*; *MBIP*; *MGA*; *MPP1*; *NHLRC2*; *NISCH*; *NKRF*; *NPC1*; *NSF*; *NUCB2*; *PAM*; *PARP8*; *PDE2A*; *PGD*; *PLCB1*; *PNPLA8*; *PON2*; *PREPL*; *PRKAR1A*; *PRUNE*; *PSMA1*; *PSTPIP1*; *PUM2*; *PXK*; *PYGL*; *RAB27A*; *RAB3D*; *RABGAP1L*; *RARS*; *RHOT1*; *RMI1*; *RNF130*; *RPGR*; *RSC1A1*; *S100A8*; *SCRN3*; *SCYL1*; *SDCBP*; *SEC22C*; *SELL*; *SENP5*; *SERPINB1*; *SHB*; *SIGLEC1*; *SLC16A7*; *SLC25A44*; *SLC35B3*; *SLC39A9*; *ST8SIA4*; *TBC1D2*; *TEX2*; *TGM1*; *TM6SF1*; *TMEM161B*; *TMEM71*; *TPCN1*; *TREML2*; *TRIP11*; *TSHZ1*; *UBE4A*; *UMPS*; *USP10*; *UXS1*; *VAPB*; *VNN3*; *VPS37B*; *WDTC1*; *XBP1*; *ZMYM4*	

[a] *Genes conserved in 4/4 species, or in 3/3 species for cDC2, cMo, and ncMo since only three species could contribute to the analysis.*

[b] *Genes conserved in 3/4 species or 2/3 species.*

[c] *Genes in bold were previously demonstrated to play a significant role in the development or functions of the population of interest.*

[d] *Underlined genes were annotated as located in "plasma membrane" according to Ingenuity Pathway Analysis.*

[e] *Genes highlighted in gray have been previously identified as signatures genes for the corresponding mouse and human cell populations in our earlier study (15).*

[f] *Signature genes of the relative cMo vs. ncMo and of the ncMo vs. cMo signatures were provided only for the 3/3 species selection since the gene lists for the 2/3 species selection encompassed hundreds of genes.*

P38 MAPK ($10^{-5.5}$), IFNG ($10^{-5.4}$), and to a lesser extent STAT3 ($10^{-4.7}$) and CSF2 ($10^{-4.3}$) are predicted as putative upstream regulators in this signature (**Figure 11A**). In addition, this cDC vs. pDC signature includes *BATF3*, a gene highly expressed in cDC that is key in cDC1 development in mouse and human (60), as well as *ARHGAP22*, a gene involved in actin cytoskeleton regulation (61), that was initially described as a top gene of the absolute cDC signature common to human and mouse (15). Altogether, the relative gene signatures of cDC emphasize their nature of highly endocytic, motile, and expert antigen-presenting cells throughout species.

The conserved cDC1 signature encompasses genes with known contribution in the biology of this lineage, such as *XCR1*, *FLT3*, and *CADM1* (59), as well as additional genes which biological function in this subset remains enigmatic, such as the germinal center B-cell-expressed transcript 2 protein (*GCET2*),
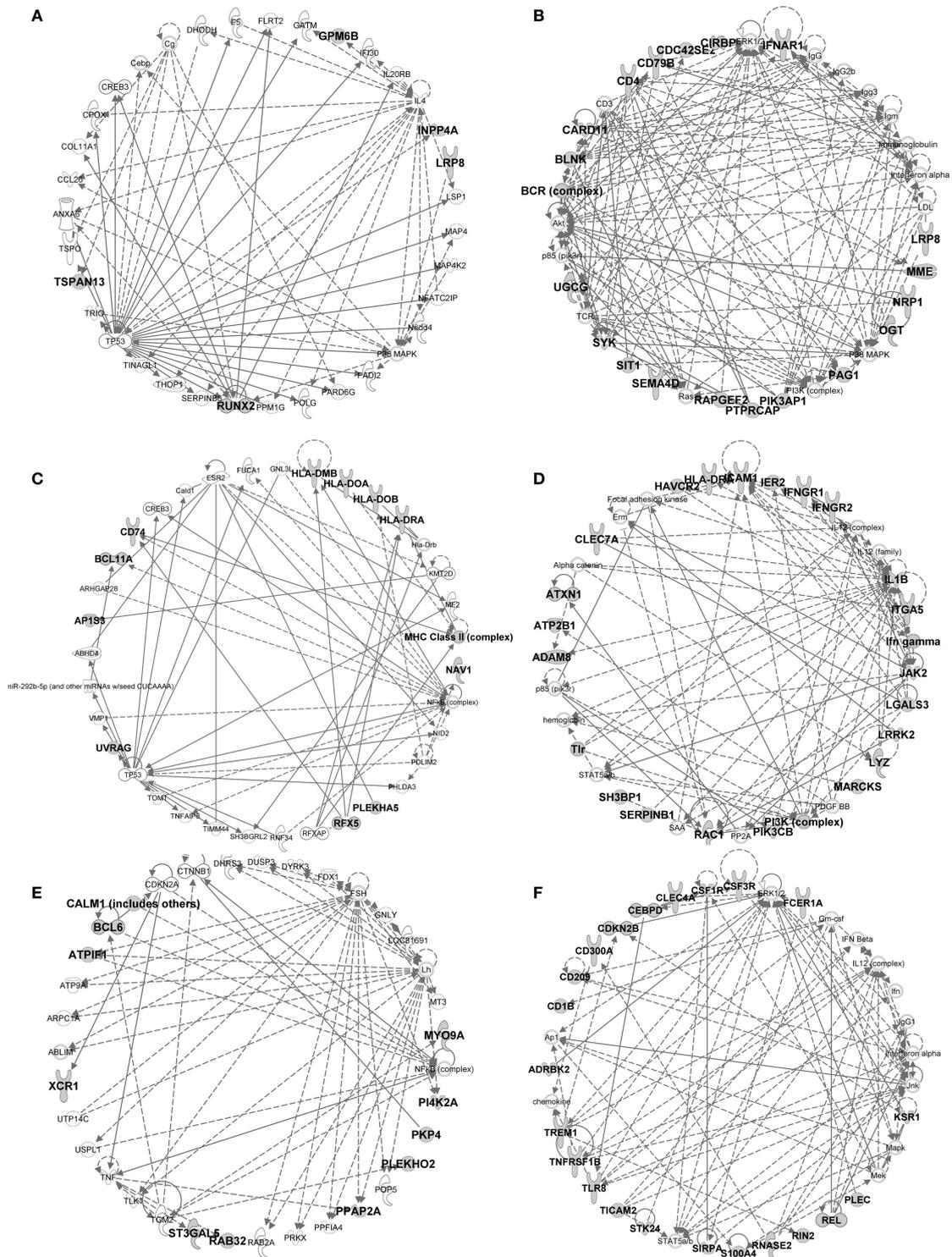
FIGURE 9 | IPA gene interaction networks of the conserved signatures in subsets of the DC lineage. The conserved signatures of subsets of the DC lineage were analyzed in Ingenuity Pathway Analysis which generates networks based on the connectivity of the genes in each signature (in boldface) but also on their connectivity with genes not belonging to the signature (in plain characters). The identified networks are displayed as graphs showing the molecular relationships between genes/gene products. Genes are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). The edges can represent direct (continuous) or indirect (dashed) relationships between nodes. Selected networks generated by IPA and covering parts of conserved cell-specific signatures are displayed: **(A)** pDC signature network, **(B)** pDC vs. cDC signature network, **(C)** cDC vs. Mo/MP signature network, **(D)** cDC vs. pDC signature network, **(E)** cDC1 vs. (pDC and cDC2) signature network, **(F)** cDC2 vs. (pDC and cDC1) signature network.
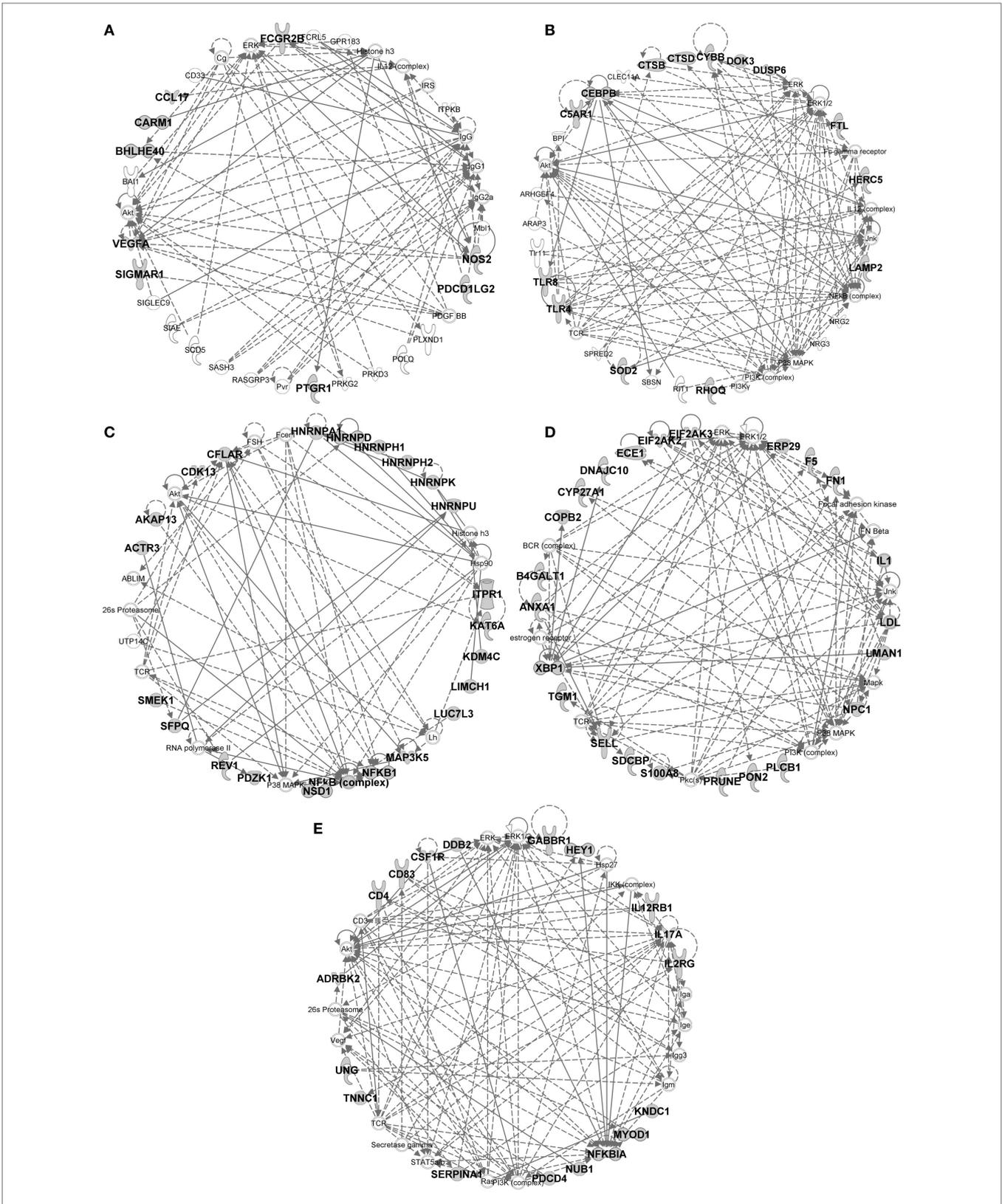
FIGURE 10 | IPA gene interaction networks of the conserved signatures in subsets of the monocytic lineage. The conserved signature of subsets of the monocytic lineage were analyzed in Ingenuity Pathway Analysis as in

Figure 9. The selected networks displayed are: **(A)** MoDC signature network, **(B)** Mo/MP vs. cDC signature network, **(C)** Mo/MP vs. MoDC signature network, **(D)** cMo vs. ncMo signature network, **(E)** ncMo vs. cMo signature network.

FIGURE 11 | IPA analysis of the conserved cell type gene signatures: upstream regulators (A) and biological functions and canonical pathways (B). A promoter sequence analysis of the conserved cell type gene signatures performed using IPA is displayed as a heatmap of the p-value [upstream regulators, **(A)**]. A biological function and canonical pathway analysis of the conserved cell type gene signatures performed using IPA is displayed as a heatmap of the p-value **(B)**. Selected upstream regulators and functions and pathways (*), in **(A,B)**, respectively, were classified using hierarchical clustering based on the average linkage metrics. Enrichments were considered significant when supported by at least three genes and by a p-value ≤0.05.

the WDFY family member 4 (*WDFY4*) whose polymorphism is associated to autoimmune diseases (62), and two intracellular trafficking proteins, a formin-binding protein (*FNBP1*) (63) and Sorting Nexin-22 (*SNX22*) (64) (**Table 1**). The conserved cDC1 vs. (pDC and cDC2) relative signature provides a longer list of genes belonging to an interaction network that includes *BCL6*, a transcriptional repressor that was recently found involved in the specification of cDC1 (17) as well as *XCR1* and *CALM1* (**Figure 9E**). IPA did not retrieve significant annotations for the cDC1 absolute or relative gene signatures. This emphasizes how little is currently known on the molecular regulation of the functions specific to cDC1s, such as cross-presentation. Future studies investigating in mouse cDC1s the functional role of the genes identified here as being part of the conserved cDC1 signatures will advance our understanding of the functions of these cells and their molecular regulation.

The absolute cDC2 conserved signature was empty. Many genes of the relative cDC2 vs. (pDC and cDC1) signature belong to a network that includes *SIRPα* (*CD172A*), a selective marker of cDC2 within the DC lineage (44), together with *CSFR1*, *TREM1*, *CLEC4A* (also known as *DCIR*), *CD1B*, and *RELB* which is known to control mouse cDC2 differentiation (65) (**Figure 9F**). A second network includes *ITGAM* (*CD11b*), a marker used to identify mouse cDC2, *CLEC6A*, and *IL1B* (Figure 7 in Supplementary Material). SPI1 ($p = 10^{-9.8}$), CEBPD ($10^{-6.2}$), and CEBPB ($10^{-3.6}$) are predicted upstream regulators, as well as CSF2 ($10^{-3.8}$), STAT3 ($10^{-4.4}$), and IFNG (10^{-4}) which were already enriched in the cDC vs. pDC signature (**Figure 11A**). This cDC2 relative signature also includes *IFI30*, also known as *GILT*, a lysosomal thiol reductase important in MHC class II and class I antigen processing (66, 67) (**Table 1**). This relative signature is enriched for "function of

antigen-presenting cells" ($p = 10^{-6.2}$), "inflammatory response" ($p = 10^{-6.9}$), and for the pathways "TREM1 Signaling" ($10^{-2.5}$), "Toll-like receptor canonical signaling" ($10^{-2.5}$), and "role of pattern recognition receptors in recognition of bacteria and viruses" ($p = 10^{-3.9}$) (**Figure 11B**). Other genes were uncovered that may be important regulators of the function of cDC2s or which product could be used to identify or target these cells, including the genes coding for plasma membrane proteins such as glycoprotein *CD300A*, the sialic binding lectin *SIGLEC8*, and the paired immunoglobin-like type 2 receptor *PILRA*. This conserved relative signature shows that within the DC lineage throughout species, cDC2 express specific networks of genes related to pathogen sensing, antigen presentation, IL-1β production, and inflammation.

In the conserved signatures corresponding to the monocytic lineage, the absolute MoDC and relative MoDC vs. Mo/MP signatures are enriched for "inflammation of organ" ($p = 10^{-6.9}$ and $p = 10^{-3.3}$), "function of antigen-presenting cells" ($p = 10^{-5.1}$ and $p = 10^{-3.6}$), and "migration of antigen-presenting cells" ($p = 10^{-3.3}$ and $p = 10^{-4.3}$) (**Figure 11B**) and encompasses *NOS2*, *CCL17*, *VEGFA*, and *FCGR2B* that map to a common major network (**Figure 10A**). IL-4 ($p = 10^{-5.7}$) and IL5 ($p = 10^{-5.9}$) are predicted regulators (**Figure 11A**). Among other genes of interest that had not yet been associated to MoDC are the triose phosphate isomerase *TPI1*, the NADH dehydrogenase flavoprotein *NDUFV2*, the aldolase *ALDOA*, and the *CD200R1* gene that encodes for an inhibitory cell surface receptor of MP functions (68) (**Table 1**). The relative MoDC vs. Mo/MP signature encompasses additional genes that participate in "migration of cells" ($p = 10^{-2.3}$, with *S1PR3*, *CCL17*, and *SLC2A1*), "bacterial infection" ($p = 10^{-3.2}$, with *CD1B*, *CD209*, and *FCGR2B*), and "synthesis of nitric oxide" ($10^{-2.5}$, with *PLAU*, *IL1R2*, and *NOS2*) (**Figure 11B** and Data Sheet S4 in Supplementary Material). The conserved MoDC signatures indicate a dominant association of this subset to inflammation, as well as to DC functional properties when compared to Mo/MP across species.

Most of the genes in the Mo/MP vs. cDC conserved signature had been previously identified as overexpressed in murine MP, such as *TLR4*, *CEBPB*, *C5AR1*, and *SOD2* (9, 15) (**Table 1** and **Figure 10B**). A significant proportion of the genes within this signature are related to "inflammation of organ" ($p = 10^{-2.7}$), "production of reactive oxygen species" ($p = 10^{-4.4}$), "synthesis of nitric oxide" ($p = 10^{-3.9}$), "bacterial infection" ($p = 10^{-4.6}$), "role of pattern recognition receptor in recognition of bacteria and viruses" ($p = 10^{-3.3}$), and "acute phase response signaling" ($p = 10^{-2.9}$) (**Figure 11B**). Putative upstream regulators are NPC1 ($p = 10^{-6.8}$), APOE ($p = 10^{-6.8}$), IFNG ($p = 10^{-6.4}$), SPI1 ($p = 10^{-5.2}$), and NFkB ($p = 10^{-5.1}$) (**Figure 11A**). Additional proteins are potential transcriptional regulators of importance in Mo/MP, such as the RNA-binding protein RBMS1 and the cell cycle progression factor CCPG1. The Mo/MP vs. MoDC signature includes a gene network centered on *NFkB* and *MAP3K5* (**Figure 10C**). Overall, the conserved Mo/MP relative signatures support the association of Mo/MP to inflammation and oxidative stress across species.

The conserved comparative signature of cMo vs. ncMo retrieved genes belonging to a network with *IL1*, fibronectin (*FN*), *S100A8*, and *XBP1* (**Figure 10D**), the latter being proposed as an upstream regulator ($10^{-2.5}$) together with *NPC1* ($p = 10^{-4}$) (**Figure 11A**), and is strongly associated to "inflammatory response" ($p = 10^{-8.4}$) (**Figure 11B**). The reciprocal ncMo vs. cMo signature includes a gene network with *IL17A*, *CSFR1*, *NFKBIA*, and *serpinA1* (**Figure 10E**), and is significantly associated to the "glucocorticoid receptor signaling pathway" ($p = 10^{-2.5}$) and to some extent to the "inflammation of organ" ($p = 10^{-2.6}$) (**Figure 11B**). These relative signatures indicate that cMo have a conserved gene program directed to strong inflammation, whereas ncMo, a poorly understood subset, might be exquisitely regulated by glucocorticoids as suggested in the literature (69, 70).

Altogether, the mononuclear phagocyte system from distantly related mammals is composed of a diversity of subsets that belong to the DC or to the Mo/MP lineage and express discriminating gene signatures involved in distinct regulatory networks and biological functions conserved through mammalian evolution. In most instances, the subset signatures also point to several unexpected genes and upstream regulators that are likely to be important in the subset biology since their selective expression pattern across subsets of mononuclear cells is conserved across species.

Phylogenetic Evidences for the Existence of a Gene Repertoire for Mononuclear Phagocyte Subsets in Birds and Bony Fishes

The existence of orthologous genes of the conserved mononuclear phagocyte subset signatures in reptile/birds, fishes, and agnathans, would indicate that the genetic equipment for mononuclear phagocyte subset diversity is available in vertebrate species distant from mammals. In the case of birds and reptiles, it remains unknown whether they have pDC, cDC1, and cDC2 subsets homologous to mammals. An orthology analysis of selected genes from conserved subset signatures revealed that most genes possess a unique ortholog in birds and reptiles, with conserved synteny with human, for instance *XCR1*, *BATF3*, *RUNX2*, *TSPAN13*, and *CSF1R* (**Table 2**). Furthermore, these same genes also possess one or more orthologs in fish. Multiple orthologs in fish are often due to the whole genome duplication that occurred during the evolution of teleosts, and to further local duplications. Importantly, fish co-orthologs of mononuclear phagocyte subset genes are generally supported by conserved synteny. Genes duplicated in fish may have been subjected to sub-functionalization, as it is the case for many immune genes duplicated in this group of vertebrates; however, some markers have a unique counterpart in fish genomes (like *BATF3*, *RFX5*, and *CIITA*), with copy loss possibly due to detrimental effects of duplication. The case of *MHC class II* is particular: although fish *MHC class II* genes are not always considered as true orthologs of human *MHC class II* genes, their sequences show the hallmarks of *bona fide* class II antigen-presenting receptors and they likely have similar functions. For c-type lectin-like (*CLEC*) molecules, no true orthologs can be identified in fish nor in birds/reptiles, as each branch of vertebrates – even each group of mammals – shows its own set of expanded *CLEC* genes. Altogether these data show that a repertoire of conserved genes for mononuclear phagocyte subsets exists in bony fishes and reptiles, which constitutes a list of candidates for relevant markers.

TABLE 2 | Search for the existence of orthologs in reptile/birds, fishes, and agnathans for selected genes of the conserved mononuclear phagocyte subset signatures.

Cell subset gene signatures	Gene	Reptiles/birds		Fishes		Agnathans (lamprey)	
		Orthologs	Conserved synteny	Orthologs	Conserved synteny	Orthologs	Conserved synteny
cDC vs. Mo/MP	HLA-DR	–	–	–	–	–	–
(MHC-related	HLA-DM	?	–	–	–	–	–
molecules)	HLA-DO	–	–	–	–	–	–
	CD74	+ (1)	Yes	+ (Multiple)	Yes	–	–
	CIITA	+ (1)	Yes	+ (1)	Yes	–	–
cDC vs. Mo/MP	NAV1	+ (1)	Yes	+ (Multiple)	Yes	–	–
	RFX5	+ (1)	Yes	+ (1)	Yes	–	–
	BCL11A	+ (1)	Yes	+ (Multiple)	Yes	–	–
MoMP vs. cDC	CEBPB	+ (1)	Yes	+ (1)	Yes	–	–
	C5AR1	+ (1)[a]	Yes[b]	+ (1)[c]		–	–
	SOD2	+ (1)	Yes	+ (1)	Yes	+ (1)	?
	APOE	–	–	+ (Multiple)	Yes	–	–
	TLR4	+ (1)	Yes	+[d]	Unclear	–	–
cDC1	XCR1	+ (1)	Yes	+ (Multiple)	Yes[e]	–	–
	FLT3	+ (1)	Yes	+ (1)	Yes	–	–
cDC vs. pDC	BATF3	+ (1)	Yes	+ (1)	Yes	–	–
	ARHGAP22	+ (1)	Yes	+ (Multiple)	Yes	+ (1)	?
	CLEC7A	–	–	–[f]	–	–	–
B cells	CD79B	+ (1)	Yes	+ (1)	Loose[g]	–	–
	PAX5	+ (1)	Yes	+ (1)	Yes	(+)[h]	?
	CD19	–	–	–	–	–	–
pDC	RUNX2	+ (1)	Yes	+ (1)[i]	Yes	+ (1)	?
	TSPAN13	+ (1)	Yes	+ (Multiple)	Yes	+ (2)	? (for both)
cDC2 vs. (cDC1	IFI30/GILT	+ (1)	Yes	+ (Multiple)	Yes	+ (1)	?
and pDC)	CSF1R	+ (1)	Yes	+ (Multiple)	Yes	–[j]	–
	SIRPA	?[k]	–	–	–	–	–
	TREM1	–[l]	–	–	–	–	–
	CLEC4A	–	–	–[f]	–	–	–
	CLEC6A	–	–	–[f]	–	–	–
More or less	CLEC9A	–	–	–[f]	–	–	–
cDC1-specific	CLNK	+ (Turkey)	Yes	+ (1)	Yes	+ (1)[m]	?

[a]Birds have one co-ortholog of human C5RA1 and C5RA2; [b]only in the lizard Anolis, not in available bird genomes; [c]fish generally have one co-ortholog of human C5RA1 and C5RA2; [d]only in some species: zebrafish, catfish, and salmonids; [e]see Ref. (36); [f]for all CLEC, no true ortholog, each deep branch of vertebrates has its own set of expanded CLEC; [g]the neighborhood is not conserved but zebrafish CD79B is close to Arhgap27 and Plekhm1 that are on the same human chromosome (chr17) as CD79B but at 20 megabases; CD79B genes often are not annotated in fish genomes. In zebrafish, CD79B is ENSDARG00000088902; [h]a lamprey gene ortholog to PAX5 has been identified and was selectively expressed in lamprey VLRB+ cells which resemble B lymphocytes (55); however, this gene is not identified in the current publicly available assembly of the lamprey genome; [i]duplicated in zebrafish and cavefish; [j]a lamprey gene is a co-ortholog to all vertebrate CSF1R, PDGFR, KIT, FLT3, etc.; [k]in birds species, several genes are co-orthologs of all mammalian SIRPs including SIRPA; [l]bird TREM-like genes are more closely related to TREM2 rather than to TREM1; [m]co-ortholog of CLNK, BLNK, and other related genes.

The presence of *BATF3* and *XCR1* are hints at possible existence of cDC1 in these species, as *BATF3* specifically controls cDC1 development in mice (71) and *XCR1* expression is strictly associated to cDC1 in several mammals (34, 36, 40, 41). In contrast, the lamprey does not have identified orthologs for many of the genes selected from the transcriptomic fingerprints of the subsets of mammalian mononuclear phagocytes (**Table 2**). Agnathans, including lampreys and myxines, harbor three adaptive immune cell types, each expressing a specific class of variable lymphocyte receptors, VLRC, VLRA, and VLRB, and showing transcriptomic and functional commonalities with gnathostome γδ T lymphocytes, αβ T lymphocytes, and B lymphocytes, respectively (55, 72). However, it is uncertain whether or not the activation of agnathan lymphocytes requires APCs, and if so, to which extent these cells could resemble gnathostome APCs (72). Contrary to the situation in birds and fishes, our observations do not support the existence in the lamprey of gene sets similar to those defining the transcriptomic fingerprints of the mononuclear phagocytes

of mammals. Although incomplete assembly and annotation of the genome of the lamprey do not allow drawing definitive conclusions, our observations are consistent with the lack in agnathans of MHC functional homologs and of the particular proteasome machinery used by mammalian APCs for antigen processing (72). Altogether, this phylogenetic study shows that the repertoire of key genes characterizing the diversity of the mononuclear phagocytes in mammals were already present in the common ancestor of tetrapods and fishes but might be largely absent in agnathans.

Discussion

Our computational transcriptomic meta-analysis indicates that the complex organization of the mononuclear phagocyte system shows conservation throughout distantly related mammals, a finding that appears to extend to chicken, a non-mammalian vertebrate. In the present work, by using GSEA and hierarchical

clustering for unbiased pan-genomic analysis of the molecular identity of immune cell subsets across four vertebrate species, we convincingly established the existence of strong homologies between these cell types across mammals, beyond the already known existence of B cells in all species. Specifically, we could align across mammals cDC1, cDC2, pDC, MoDC, Mo/MP, and cMo vs. ncMo. In addition, we found that many of the genes that we showed to be selectively expressed in distinct mononuclear phagocyte subsets in mammals have existing orthologs in bony fishes while this appears not to be the case in lamprey. Thus, our study suggests that conserved mononuclear phagocyte subsets might exist in all gnathostomes but not in agnathans. However, this hypothesis will require to be tested experimentally, by re-examining the presence of orthologous genes in lamprey upon completion of the genome assembly and its annotation, by identifying and studying candidate mononuclear phagocyte subsets in bony fishes, and by determining whether similar cells exist in sharks, rays, and lamprey. For example, orthologous genes of the conserved mononuclear phagocyte signatures (**Table 2**) could be targeted by the CRISPR/Cas9 technology with a reporter gene marker in order to identify and characterize mononuclear phagocyte subsets in bony fishes (73), with for certain genes the need to test several putative orthologs in fish due to genome duplication.

The two methodologies that we used to assess subset homologies across species, i.e., hierarchical clustering and GSEA, display complementary functionalities. Hierarchical clustering on filtered, centered, reduced, and aggregated datasets has the advantage of integrating all samples together into a single analysis and of providing a global overview of the homologies between cell subsets of various species (15, 17, 18, 74, 75). However, the integration of distinct datasets requires a cross-normalization procedure which consists in a rather profound mathematical transformation of the data. The normalization procedure artificially increases the variance for genes with only small differences in their initial signal intensities between the different cell types studied. Conversely, it comparatively decreases the variance for genes with high differences in their initial signal intensities between the different cell types studied. To limit the biases that this normalization introduces, it is thus necessary to select only the orthologous genes that vary strongly in their expression across the cell types examined within each species. Another corollary is that this analysis can only be applied to genes that have known orthologs in all species. If one ortholog is missing in only one species, the gene must be removed from the analysis. Hence, this method should be used with caution, only under conditions where dataset normalization does not yield too strong biases in gene-expression profiles. It is also not appropriate when the structures of the different datasets are too different (i.e., the number and potential identities of cell types vary too much across datasets), because the dynamic ranges of gene expression between datasets are not expected to be the same and should therefore not be forced to similarity. Even under conditions where the experimental design is favorable to the use of hierarchical clustering, GSEA ensures of the robustness of interpretation. GSEA has been used by us and others to perform cross-species comparisons (5, 19, 29,

42, 76–78). GSEA notably displays advantages and drawbacks distinct from those of hierarchical clustering. First, it is easier to perform GSEA since dedicated ready-to-use stand-alone programs are available which do not require bio-informatics expertise. Second, GSEA is more sensitive, notably to detect overlaps of common functions/gene networks between cell populations or cellular contaminations, as exemplified with sheep *pDC enriched in human and mouse B cell fingerprints. This higher sensitivity is linked to (i) the fact that GSEA can detect coordinate regulation of gene modules (geneset-based approach) and thus does not rely on the strong regulation of few single genes (single gene-based approach), (ii) the fact that GSEA, when applied to multiple species, takes into account all genes that have orthologous counterparts in the considered species and is not restricted only to highly variable genes. Third, GSEA can perform cross-platform comparison without any cross-normalization thus without any supplementary artificial manipulation of the expression data. Finally, it can be performed on multiple datasets, even if their structures are different. However, GSEA presents the limitation of performing pairwise comparisons whose results can be integrated and visualized with our Bubble GUM software, but it nevertheless does not provide a global trans-species overview of subset homology. Overall, in order to increase confidence in the interpretation of the results, it is important to combine both approaches and verify that they both lead to consistent conclusions.

Our subset assignment methodology demonstrates similarity or proximity between subsets across species but not strict identity. Besides possible intrinsic transcriptomic differences between species, one of the reasons that explain this limitation is the process of subset identification itself, which makes use of different surface markers. Whenever possible, similar marker combinations were used such as CADM1 and CD172 that are known to be conserved markers across human, mouse, and sheep cDC subsets (42). However, mAb anti-CD11c did not exist for the initial gating in pig and the mAbs in the exclusion pool were not the same in pig and sheep. Moreover, existing marker combinations are not always specific and can lead to cross-contamination between different cell subsets. Indeed, the GSEA of the sheep *cDC2 revealed that they may have been contaminated by pDC, despite our attempt to avoid this problem through exclusion of CD45RB-expressing cells. It remains possible that pDC expressing minimal levels of CD45RB were still present in the sorted *cDC2 population, and not in the sheep cDC1 subset. However, since sheep *cDC2 were found in the correct cDC branch of the hierarchical clustering, their contamination by pDC is likely to have been limited. Similarly, it is likely that the sorted sheep *pDC include residual B cells, explaining the enrichment for the human B cell fingerprint at a level above expectation: indeed after exclusion of B cells with a pan-B cell marker, sheep *pDC were selected with a mAb directed to CD45RB, which may react with residual B cells that have escaped the pan-B cell exclusion. Yet, sheep *pDC still cluster with other species pDC, separately from B cells. In the case of pig, pDC were selected using markers not expressed by B cells and they displayed an enrichment for B cell fingerprints at a level encountered in GSEA analyses of mouse pDC (Figure 4 in Supplementary Material). Finally, our approach was able to

demonstrate that *a priori* assignment of subset identity based on the expression of a few membrane markers could be wrong, like in the case of the pig *cDC2. Moreover, our approach had the power to properly re-assign cell subset identity, demonstrating that pig *cDC2 were actually homologous to mouse and human ncMo. Another laboratory analyzed the transcriptome of similar pig cells sorted as CD14^{low} CD163^{high} cells, but they could not assign them to classical nor to non-classical human Mo, due to differences in bio-informatics approaches in this study (79) and in ours.

Our study will help improving in the near future the toolbox available in each species for rigorous and consistent phenotypic identification of cell subsets, thanks to our identification of novel, conserved, and specific, combinations of surface markers for each cell subset, which should allow generating more appropriate staining reagents. For instance, fluorescently labeled recombinant XCL1 could theoretically be used in any species to rigorously identify and sort cDC1 (38, 41). In addition, cell surface proteins encoded by genes shown here to be selectively expressed in a conserved manner in specific subsets of mononuclear phagocytes represent new candidate markers to refine and homogenize phenotypic identification of these cells across species, such as LRP8, TSPAN13, NRP1, and SLC30A5 for pDC, FCGR2B, and CD200R1 for MoDC, SIGLEC8 and IGSF6 for cDC2, and CSF1R, TLR4, and C5AR1 for Mo/MP (**Table 3**). However, these potential new markers for subset identification need to be validated at the protein level.

The subset-specific signatures that are conserved throughout distant mammals included variable number of genes that were sometimes far lower than the numbers of genes in the human/ mouse common signatures. There are several explanations to this finding. There is a contribution of the very high stringency of the "Min (test) vs. Max (ref)" ≥1x method that we used to establish the signatures, since any gene which was not consistently found overexpressed in all the replicates of all the species was excluded. As an example, the gene *DNAJC7*, identified as specific of pDC in our previous work (15) was removed from the human pDC signature because its "Min (test) vs. Max (ref)" ratio was equal to 0.933, due to a single lower human pDC replicate compared to a single replicate found with a higher signal in human MoDC. There is also a contribution of incomplete mapping of the genome of some of the species studied, leading to an underestimation of the number of orthologous genes that could be queried across all species. For example, *POU2F2*, more highly expressed in human and murine B cells as compared to many other immune cells, has not been mapped yet to the pig genome while it has been mapped to the genome of more distant species such as the spotted gar with a 1-to-1 orthology relationship. Another prominent cause is linked to technical limitations of the microarray approach, such as lack of ProbeSets against certain genes in certain species. This is notably the case for the gene *CLEC9A*, known to be specific of cDC1 but for which no ProbeSet exists in the human Affymetrix HG U133 plus2 gene chip. Sometimes, low signal-to-noise ratio for certain ProbeSets can also be responsible for the loss of putative interesting signature genes, such as *ZNF521* (*Zfp521* in mouse) found to be highly specific of pDC in mouse and human while the pig and sheep orthologous ProbeSet remains at the background level whatever the cell type considered. Recent technological advances now allow performing high throughput RNA sequencing at single cell levels with high sensitivity and processivity, which could solve most of the above issues; indeed, all expressed genes should be detected without any bias and analysis at the single cell level should alleviate any issue of cross-contamination between cell types. Therefore, the generation of gene-expression data for many individual cells of the same type should increase statistical power to define genes co-expressed at the single cell level and defining cell type-specific transcriptomic modules (22). Single cell gene-expression profiling recently allowed the unbiased and *de novo* identification of the different cell types of spleen (80) and central nervous system (81, 82) via the description of their molecular identity, starting from the bulk population of all the cells that could be extracted from the organ, without any prior enrichment procedure, based on the use of potentially confounding phenotypic marker combinations. However, this strategy is

TABLE 3 | Proposition of marker combination for oligo-phenotyping of mononuclear phagocytic cell subsets across species.

Exclusion	Anti-CD3, anti-NK cells, and anti-B cells, if available[a]						
Targeted cell population	pDC	cDC1	cDC2	MoDC	cMo	ncMo	MP
Combination of known markers[b]	FLT3$^+$ SIRPα^{lo} MHC-II^{lo}	FLT3^{hi} SIRPα^{lo} MHC-II$^+$ CD11c$^+$ CADM1^{hi}	FLT3$^+$ SIRPα$^+$ MHC-II$^+$ CD11c$^+$ CADM1^{lo}	FLT3$^-$ SIRPα$^+$ MHC-II$^+$ CD11c$^+$	FLT3$^-$ SIRPα$^+$	FLT3$^-$ SIRPα$^+$	FLT3$^-$ SIRPα$^+$
New additional candidates[c]	LRP8$^+$ TSPAN13^{hi} SLC30A5$^+$ NRP1$^+$	XCR1$^+$	SIGLEC8$^+$ IGSF6$^+$	FCGR2B^{hi} CD200R1$^+$	CSF1R^{int} CCR1$^+$ C19ORF59$^+$	CSF1R^{hi} CD83$^+$	CSF1R$^+$ TLR4^{hi} C5AR1$^+$

[a] Exclusion with anti-CD3, anti-NK cells, and anti-B cell markers is desirable when appropriate tools are available.

[b] A combination of known markers including FLT3, MHC-II, CD11c, SIRPα, and CADM1 allows a first step of identification of subset candidates but is at risk of contamination by sister cell types, or may be incomplete due to non-availability of one of the marker. FLT3 labeling may be performed by using recombinant His-tag FLT3L generated for the relevant species as recently proposed in a review (21).

[c] New additional candidate markers for refinement of subset identification are derived from the identification of genes encoding cell surface molecules from the conserved cell subset gene signatures.

still extremely difficult to apply to species which genome has not yet been completely assembled, as well as to very rare cell types recovered upon prior phenotype-based enrichment. Moreover, to obtain information of sufficient completeness on functionally important genes for which few mRNA are expressed per cell, it is necessary to sequence at a sufficient depth of about one million reads per cell, which today still represents a very high cost when multiplied by the number of individual cells and conditions. Finally, the interpretation of the RNA-seq data on single cells is still largely based on the transcriptomic/molecular identity of cell types that are deduced from microarray analysis of purified cell pools. Hence, our work constitutes a major advancement in the field and is a necessary step before an eventual, later, refinement of the definition of cell subsets and their associated molecular signatures using single cell RNA-seq. The canonical gene-expression signatures that we generated can be used to distinguish and identify cell subsets in other vertebrate species. The cDC1 signature and the cDC2 vs. cDC1 signatures could be evaluated in chicken cDC sorted as single cells to determine whether this population includes only cDC1, as suggested by the trans-vertebrate hierarchical clustering, or a mixture of cDC1 and cDC2.

The conservation of gene signatures and interacting gene networks in homologous cell subsets throughout evolution is likely to bear strong biological meaning. Indeed, many genes of the conserved signatures were already known for their functions in these cells, validating the biological relevance of our signatures. In several instances, the same functional annotations were enriched in distinct subset signatures, but the genes responsible for the enrichments differed. For example, the genes responsible for the enrichment of the pathway "role of pattern recognition receptor in recognition of bacteria and viruses" were *TLR4*, *TLR8*, and *C5AR1* for the Mo/MP vs. cDC signature, *TLR2*, *CLEC7A*, *IL1B*, and *PIK3CB* for the cDC vs. pDC signature, and *TLR8*, *CLEC6A*, *DDX58*, *OAS2*, and *IL1B* for the cDC2 vs. (pDC and cDC1) signature. This analysis shows that cDC and MP express different sets of pattern recognition receptors for detection of viruses and bacteria, and that, within DC, cDC2 are also equipped differently from cDC1 and pDC for sensing of viruses and bacteria. These observations extended to other mammalian species the previous reports that human and mouse cDC2 are preferentially equipped with PRR targeting bacteria or involved in cytosolic sensing of viral infection (83, 84), and that TLR4 is very weakly expressed on pDC and cDC as compared to Mo/MP (83, 85). Similarly, different subset signatures were all enriched for "inflammatory response," "inflammation of organs," and "bacterial infection" but due to different genes. Altogether, this analysis indicates that different mononuclear phagocyte subsets express distinct and specific gene-expression modules which can sometimes contribute in a complementary way to the same general biological process in a conserved manner throughout evolution. Within the conserved gene-expression programs in mononuclear phagocyte subsets, we identified novel candidate genes and putative upstream regulators which likely contribute to the control of the ontogeny or functions of the corresponding cell type. For instance, the *FNBP1* and *SNX22* encoded proteins may be involved

in the specific intracellular trafficking properties promoting antigen cross-presentation by cDC1, *ARHGAP22* and *NAV1* could modulate the organization of the cytoskeleton of cDC to control their mobility or antigen presentation functions, and the transcription regulators *BCL11A* and *MSL2A* may control specific gene networks in cDC. *BLC11A* is known to be key in murine pDC development (50) but it may have a specific role in cDC homeostasis, as inferred from a previous study (86). Our study thus opens the way for deciphering the sets of genes encoding functional cellular modules and their specifying transcription factors in subsets of mononuclear cells, in order to further improve and connect together the molecular and functional definitions of these cell types across species (22, 23).

Conclusion

Our meta-analysis that combines cell sorting and comparative transcriptomic analysis was implemented as a methodology pipeline that could be used by biologists with minimal training in bio-informatics for subsequent extension to other species and to other complex cellular systems. Our study should lead to the identification of homologous mononuclear phagocyte subsets in species other than sheep and pigs, and which are of importance for biomedical investigations, such as bats, rabbits, ferrets, guinea pigs, possibly zebrafishes, and in species of veterinary importance including pets and animals of the food economy. The characterization of mononuclear phagocyte subsets in these species will allow manipulating their immune responses against diseases for the sustainability of our environment.

Author Contributions

Acknowledgments

Supplementary Material

The Supplementary Material for this article can be found online at http://journal.frontiersin.org/article/10.3389/fimmu.2015.00299

References

1. Ziegler-Heitbrock L, Ancuta P, Crowe S, Dalod M, Grau V, Hart DN, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood* (2010) **116**(16):e74–80. doi:10.1182/blood-2010-02-258558

2. Mildner A, Yona S, Jung S. A close encounter of the third kind: monocyte-derived cells. *Adv Immunol* (2013) **120**:69–103. doi:10.1016/B978-0-12-417028-5.00003-X

3. Tamoutounour S, Guilliams M, Montanana Sanchis F, Liu H, Terhorst D, Malosse C, et al. Origins and functional specialization of macrophages and of conventional and monocyte-derived dendritic cells in mouse skin. *Immunity* (2013) **39**(5):925–38. doi:10.1016/j.immuni.2013.10.004

4. Bain CC, Bravo-Blas A, Scott CL, Gomez Perdiguero E, Geissmann F, Henri S, et al. Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. *Nat Immunol* (2014) **15**(10):929–37. doi:10.1038/ni.2967

5. McGovern N, Schlitzer A, Gunawan M, Jardine L, Shin A, Poyner E, et al. Human dermal CD14(+) cells are a transient population of monocyte-derived macrophages. *Immunity* (2014) **41**(3):465–77. doi:10.1016/j.immuni.2014.08.006

6. Molawi K, Wolf Y, Kandalla PK, Favret J, Hagemeyer N, Frenzel K, et al. Progressive replacement of embryo-derived cardiac macrophages with age. *J Exp Med* (2014) **211**(11):2151–8. doi:10.1084/jem.20140639

7. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* (2014) **159**(6):1312–26. doi:10.1016/j.cell.2014.11.018

8. Gosselin D, Link VM, Romanoski CE, Fonseca GJ, Eichenfield DZ, Spann NJ, et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* (2014) **159**(6):1327–40. doi:10.1016/j.cell.2014.11.023

9. Gautier EL, Shay T, Miller J, Greter M, Jakubzick C, Ivanov S, et al. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat Immunol* (2012) **13**(11):1118–28. doi:10.1038/ni.2419

10. Cheong C, Matos I, Choi JH, Dandamudi DB, Shrestha E, Longhi MP, et al. Microbial stimulation fully differentiates monocytes to DC-SIGN/CD209(+) dendritic cells for immune T cell areas. *Cell* (2010) **143**(3):416–29. doi:10.1016/j.cell.2010.09.039

11. Xu Y, Zhan Y, Lew AM, Naik SH, Kershaw MH. Differential development of murine dendritic cells by GM-CSF versus Flt3 ligand has implications for inflammation and trafficking. *J Immunol* (2007) **179**(11):7577–84. doi:10.4049/jimmunol.179.11.7577

12. Mayer CT, Ghorbani P, Nandan A, Dudek M, Arnold-Schrauf C, Hesse C, et al. Selective and efficient generation of functional Batf3-dependent CD103+ dendritic cells from mouse bone marrow. *Blood* (2014) **24**:3081–91. doi:10.1182/blood-2013-12-545772

13. Greter M, Lelios I, Pelczar P, Hoeffel G, Price J, Leboeuf M, et al. Stroma-derived interleukin-34 controls the development and maintenance of Langerhans cells and the maintenance of microglia. *Immunity* (2012) **37**(6):1050–60. doi:10.1016/j.immuni.2012.11.001

14. Schlitzer A, Ginhoux F. Organization of the mouse and human DC network. *Curr Opin Immunol* (2014) **26**:90–9. doi:10.1016/j.coi.2013.11.002

15. Robbins SH, Walzer T, Dembele D, Thibault C, Defays A, Bessou G, et al. Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. *Genome Biol* (2008) **9**(1):R17. doi:10.1186/gb-2008-9-1-r17

16. Haniffa M, Shin A, Bigley V, McGovern N, Teo P, See P, et al. Human tissues contain CD141(hi) cross-presenting dendritic cells with functional homology to mouse CD103(+) nonlymphoid dendritic cells. *Immunity* (2012) **37**(1):60–73. doi:10.1016/j.immuni.2012.04.012

17. Watchmaker PB, Lahl K, Lee M, Baumjohann D, Morton J, Kim SJ, et al. Comparative transcriptional and functional profiling defines conserved programs of intestinal DC differentiation in humans and mice. *Nat Immunol* (2014) **15**(1):98–108. doi:10.1038/ni.2768

18. Cros J, Cagnard N, Woollard K, Patey N, Zhang SY, Senechal B, et al. Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. *Immunity* (2010) **33**(3):375–86. doi:10.1016/j.immuni.2010.08.012

19. Crozat K, Guiton R, Guilliams M, Henri S, Baranek T, Schwartz-Cornil I, et al. Comparative genomics as a tool to reveal functional equivalences between human and mouse dendritic cell subsets. *Immunol Rev* (2010) **234**(1):177–98. doi:10.1111/j.0105-2896.2009.00868.x

20. Ingersoll MA, Spanbroek R, Lottaz C, Gautier EL, Frankenberger M, Hoffmann R, et al. Comparison of gene expression profiles between human and mouse monocyte subsets. *Blood* (2010) **115**(3):e10–9. doi:10.1182/blood-2009-07-235028

21. Summerfield A, Auray G, Ricklin M. Comparative dendritic cell biology of veterinary mammals. *Annu Rev Anim Biosci* (2014) **3**:533–57. doi:10.1146/annurev-animal-022114-111009

22. Achim K, Arendt D. Structural evolution of cell types by step-wise assembly of cellular modules. *Curr Opin Genet Dev* (2014) **27**:102–8. doi:10.1016/j.gde.2014.05.001

23. Arendt D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet* (2008) **9**(11):868–82. doi:10.1038/nrg2416

24. Chevallier N, Berthelemy M, Le Rhun D, Laine V, Levy D, Schwartz-Cornil I. Bovine leukemia virus-induced lymphocytosis and increased cell survival mainly involve the CD11b+ B-lymphocyte subset in sheep. *J Virol* (1998) **72**(5):4413–20.

25. Fiebach AR, Guzylack-Piriou L, Python S, Summerfield A, Ruggli N. Classical swine fever virus N(pro) limits type I interferon induction in plasmacytoid dendritic cells by interacting with interferon regulatory factor 7. *J Virol* (2011) **85**(16):8002–11. doi:10.1128/JVI.00330-11

26. Ruscanu S, Pascale F, Bourge M, Hemati B, Elhmouzi-Younes J, Urien C, et al. The double-stranded RNA bluetongue virus induces type I interferon in plasmacytoid dendritic cells via a MYD88-dependent TLR7/8-independent signaling pathway. *J Virol* (2012) **86**(12):5817–28. doi:10.1128/JVI.06716-11

27. Foulon E, Foucras G. Two populations of ovine bone marrow-derived dendritic cells can be generated with recombinant GM-CSF and separated on CD11b expression. *J Immunol Methods* (2008) **339**(1):1–10. doi:10.1016/j.jim.2008.07.012

28. Ruscanu S, Jouneau L, Urien C, Bourge M, Lecardonnel J, Moroldo M, et al. Dendritic cell subtypes from lymph nodes and blood show contrasted gene expression programs upon bluetongue virus infection. *J Virol* (2013) **87**(16):9333–43. doi:10.1128/JVI.00631-13

29. Marquet F, Vu Manh TP, Maisonnasse P, Elhmouzi-Younes J, Urien C, Bouguyon E, et al. Pig skin includes dendritic cell subsets transcriptomically related to human CD1a and CD14 dendritic cells presenting different migrating behaviors and T cell activation capacities. *J Immunol* (2014) **193**(12):5883–93. doi:10.4049/jimmunol.1303150

30. Casel P, Moreews F, Lagarrigue S, Klopp C. sigReannot: an oligo-set re-annotation pipeline based on similarities with the ensemble transcripts and unigene clusters. *BMC Proc* (2009) **3**(Suppl 4):S3. doi:10.1186/1753-6561-3-S4-S3

31. Vu Manh TP, Marty H, Sibille P, Le Vern Y, Kaspers B, Dalod M, et al. Existence of conventional dendritic cells in *Gallus gallus* revealed by comparative gene expression profiling. *J Immunol* (2014) **192**(10):4510–7. doi:10.4049/jimmunol.1303405

32. Baranek T, Vu Manh TP, Alexandre Y, Maqbool MA, Cabeza JZ, Tomasello E, et al. Differential responses of immune cells to type I interferon contribute to host resistance to viral infection. *Cell Host Microbe* (2012) **12**(4):571–84. doi:10.1016/j.chom.2012.09.002

33. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* (2007) **23**(23):3251–3. doi:10.1093/bioinformatics/btm369

34. Gurka S, Hartung E, Becker M, Kroczek RA. Mouse conventional dendritic cells can be universally classified based on the mutually exclusive expression of XCR1 and SIRPa. *Front Immunol* (2015) **6**:35. doi:10.3389/fimmu.2015.00035

35. Dorner BG, Dorner MB, Zhou X, Opitz C, Mora A, Guttler S, et al. Selective expression of the chemokine receptor XCR1 on cross-presenting dendritic cells determines cooperation with CD8+ T cells. *Immunity* (2009) **31**(5):823–33. doi:10.1016/j.immuni.2009.08.027

36. Crozat K, Guiton R, Contreras V, Feuillet V, Dutertre CA, Ventre E, et al. The XC chemokine receptor 1 is a conserved selective marker of mammalian cells homologous to mouse CD8{alpha}+ dendritic cells. *J Exp Med* (2010) **207**:1283–92. doi:10.1084/jem.20100223

37. Bachem A, Guttler S, Hartung E, Ebstein F, Schaefer M, Tannert A, et al. Superior antigen cross-presentation and XCR1 expression define human CD11c+CD141+ cells as homologues of mouse CD8+ dendritic cells. *J Exp Med* (2010) **207**(6):1273–81. doi:10.1084/jem.20100348

38. Crozat K, Tamoutounour S, Vu Manh TP, Fossum E, Luche H, Ardouin L, et al. Cutting edge: expression of XCR1 defines mouse lymphoid-tissue resident and migratory dendritic cells of the CD8{alpha}+ type. *J Immunol* (2011) **187**(9):4411–5. doi:10.4049/jimmunol.1101717

39. Bachem A, Hartung E, Guttler S, Mora A, Zhou X, Hegemann A, et al. Expression of XCR1 characterizes the Batf3-dependent lineage of dendritic cells capable of antigen cross-presentation. *Front Immunol* (2012) **3**:214. doi:10.3389/fimmu.2012.00214

40. Balan S, Ollion V, Colletti N, Chelbi R, Montanana-Sanchis F, Liu H, et al. Human XCR1+ dendritic cells derived in vitro from CD34+ progenitors closely resemble blood dendritic cells, including their adjuvant responsiveness, contrary to monocyte-derived dendritic cells. *J Immunol* (2014) **193**(4):1622–35. doi:10.4049/jimmunol.1401243

41. Dutertre CA, Jourdain JP, Rancez M, Amraoui S, Fossum E, Bogen B, et al. TLR3-responsive, XCR1+, CD141(BDCA-3)+/CD8alpha+-equivalent dendritic cells uncovered in healthy and simian immunodeficiency virus-infected rhesus macaques. *J Immunol* (2014) **192**(10):4697–708. doi:10.4049/jimmunol.1302448

42. Contreras V, Urien C, Guiton R, Alexandre Y, Vu Manh TP, Andrieu T, et al. Existence of CD8{alpha}-like dendritic cells with a conserved functional specialization and a common molecular signature in distant mammalian species. *J Immunol* (2010) **185**:3313–25. doi:10.4049/jimmunol.1000824

43. Dutertre CA, Wang LF, Ginhoux F. Aligning bona fide dendritic cell populations across species. *Cell Immunol* (2014) **291**(1–2):3–10. doi:10.1016/j.cellimm.2014.08.006

44. Guilliams M, Henri S, Tamoutounour S, Ardouin L, Schwartz-Cornil I, Dalod M, et al. From skin dendritic cells to a simplified classification of human and mouse dendritic cell subsets. *Eur J Immunol* (2010) **40**(8):2089–94. doi:10.1002/eji.201040498

45. Hajjoubi S, Rival-Gervier S, Hayes H, Floriot S, Eggen A, Piumi F, et al. Ruminants genome no longer contains whey acidic protein gene but only a pseudogene. *Gene* (2006) **370**:104–12. doi:10.1016/j.gene.2005.11.025

46. de Jong WW, van Dijk MA, Poux C, Kappe G, van Rheede T, Madsen O. Indels in protein-coding sequences of euarchontoglires constrain the rooting of the eutherian tree. *Mol Phylogenet Evol* (2003) **28**(2):328–40. doi:10.1016/S1055-7903(03)00116-7

47. Douady CJ, Douzery EJ. Molecular estimation of eulipotyphlan divergence times and the evolution of "insectivore". *Mol Phylogenet Evol* (2003) **28**(2):285–96. doi:10.1016/S1055-7903(03)00119-2

48. Pascale F, Contreras V, Bonneau M, Courbet A, Chilmonczyk S, Bevilacqua C, et al. Plasmacytoid dendritic cells migrate in afferent skin lymph. *J Immunol* (2008) **180**(9):5963–72. doi:10.4049/jimmunol.180.9.5963

49. Summerfield A, Guzylack-Piriou L, Schaub A, Carrasco CP, Tache V, Charley B, et al. Porcine peripheral blood dendritic cells and natural interferon-producing cells. *Immunology* (2003) **110**(4):440–9. doi:10.1111/j.1365-2567.2003.01755.x

50. Ippolito GC, Dekker JD, Wang YH, Lee BK, Shaffer AL III, Lin J, et al. Dendritic cell fate is determined by BCL11A. *Proc Natl Acad Sci U S A* (2014) **111**(11):E998–1006. doi:10.1073/pnas.1319228111

51. Cao T, Ueno H, Glaser C, Fay JW, Palucka AK, Banchereau J. Both Langerhans cells and interstitial DC cross-present melanoma antigens and efficiently activate antigen-specific CTL. *Eur J Immunol* (2007) **37**(9):2657–67. doi:10.1002/eji.200636499

52. Cisse B, Caton ML, Lehner M, Maeda T, Scheu S, Locksley R, et al. Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. *Cell* (2008) **135**(1):37–48. doi:10.1016/j.cell.2008.09.016

53. Fredebohm J, Wolf J, Hoheisel JD, Boettcher M. Depletion of RAD17 sensitizes pancreatic cancer cells to gemcitabine. *J Cell Sci* (2013) **126**(Pt 15):3380–9. doi:10.1242/jcs.124768

54. Bloch DB, de la Monte SM, Guigaouri P, Filippov A, Bloch KD. Identification and characterization of a leukocyte-specific component of the nuclear body. *J Biol Chem* (1996) **271**(46):29198–204. doi:10.1074/jbc.271.46.29198

55. Hirano M, Guo P, McCurley N, Schorpp M, Das S, Boehm T, et al. Evolutionary implications of a third lymphocyte lineage in lampreys. *Nature* (2013) **501**(7467):435–8. doi:10.1038/nature12467

56. Sawai CM, Sisirak V, Ghosh HS, Hou EZ, Ceribelli M, Staudt LM, et al. Transcription factor Runx2 controls the development and migration of plasmacytoid dendritic cells. *J Exp Med* (2013) **210**(11):2151–9. doi:10.1084/jem.20130443

57. Iwakoshi NN, Pypaert M, Glimcher LH. The transcription factor XBP-1 is essential for the development and survival of dendritic cells. *J Exp Med* (2007) **204**(10):2267–75. doi:10.1084/jem.20070525

58. Cao W, Zhang L, Rosen DB, Bover L, Watanabe G, Bao M, et al. BDCA2/Fc epsilon RI gamma complex signals through a novel BCR-like pathway in human plasmacytoid dendritic cells. *PLoS Biol* (2007) **5**(10):e248. doi:10.1371/journal.pbio.0050248

59. Haniffa M, Collin M, Ginhoux F. Ontogeny and functional specialization of dendritic cells in human and mouse. *Adv Immunol* (2013) **120**:1–49. doi:10.1016/B978-0-12-417028-5.00001-6

60. Murphy TL, Tussiwand R, Murphy KM. Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. *Nat Rev Immunol* (2013) **13**(7):499–509. doi:10.1038/nri3470

61. Mori M, Saito K, Ohta Y. ARHGAP22 localizes at endosomes and regulates actin cytoskeleton. *PLoS One* (2014) **9**(6):e100271. doi:10.1371/journal.pone.0100271

62. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, Qian XX, et al. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet* (2010) **6**(2):e1000841. doi:10.1371/journal.pgen.1000841

63. Tsujita K, Kondo A, Kurisu S, Hasegawa J, Itoh T, Takenawa T. Antagonistic regulation of F-BAR protein assemblies controls actin polymerization during podosome formation. *J Cell Sci* (2013) **126**(Pt 10):2267–78. doi:10.1242/jcs.122515

64. Worby CA, Dixon JE. Sorting out the cellular functions of sorting nexins. *Nat Rev Mol Cell Biol* (2002) **3**(12):919–31. doi:10.1038/nrm974

65. Shih VF, Davis-Turak J, Macal M, Huang JQ, Ponomarenko J, Kearns JD, et al. Control of RelB during dendritic cell activation integrates canonical and noncanonical NF-kappaB pathways. *Nat Immunol* (2012) **13**(12):1162–70. doi:10.1038/ni.2446

66. Singh R, Cresswell P. Defective cross-presentation of viral antigens in GILT-free mice. *Science* (2010) **328**(5984):1394–8. doi:10.1126/science.1189176

67. West LC, Grotzke JE, Cresswell P. MHC class II-restricted presentation of the major house dust mite allergen Der p 1 Is GILT-dependent: implications for allergic asthma. *PLoS One* (2013) **8**(1):e51343. doi:10.1371/journal.pone.0051343

68. Mihrshahi R, Barclay AN, Brown MH. Essential roles for Dok2 and RasGAP in CD200 receptor-mediated regulation of human myeloid cells. *J Immunol* (2009) **183**(8):4879–86. doi:10.4049/jimmunol.0901531

69. Moniuszko M, Liyanage NP, Doster MN, Parks RW, Grubczak K, Lipinska D, et al. Glucocorticoid treatment at moderate doses of SIV-infected rhesus macaques decreases the frequency of circulating CD14CD16 monocytes but does not alter the tissue virus reservoir. *AIDS Res Hum Retroviruses* (2015) **31**(1):115–26. doi:10.1089/AID.2013.0220

70. Fingerle-Rowson G, Angstwurm M, Andreesen R, Ziegler-Heitbrock HW. Selective depletion of CD14+ CD16+ monocytes by glucocorticoid therapy. *Clin Exp Immunol* (1998) **112**(3):501–6. doi:10.1046/j.1365-2249.1998.00617.x

71. Hildner K, Edelson BT, Purtha WE, Diamond M, Matsushita H, Kohyama M, et al. Batf3 deficiency reveals a critical role for CD8alpha+ dendritic cells in cytotoxic T cell immunity. *Science* (2008) **322**(5904):1097–100. doi:10.1126/science.1164206

72. Kasahara M, Sutoh Y. Two forms of adaptive immunity in vertebrates: similarities and differences. *Adv Immunol* (2014) **122**:59–90. doi:10.1016/B978-0-12-800267-4.00002-X

73. Irion U, Krauss J, Nusslein-Volhard C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development* (2014) **141**(24):4827–30. doi:10.1242/dev.115584

74. Ellwood-Yen K, Graeber TG, Wongvipat J, Iruela-Arispe ML, Zhang J, Matusik R, et al. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell* (2003) **4**(3):223–38. doi:10.1016/S1535-6108(03)00197-1

75. Lee JS, Chu IS, Mikaelyan A, Calvisi DF, Heo J, Reddy JK, et al. Application of comparative functional genomics to identify best-fit mouse models to study human cancer. *Nat Genet* (2004) **36**(12):1306–11. doi:10.1038/ng1481

76. Vu Manh TP, Alexandre Y, Baranek T, Crozat K, Dalod M. Plasmacytoid, conventional, and monocyte-derived dendritic cells undergo a profound and convergent genetic reprogramming during their maturation. *Eur J Immunol* (2013) **43**(7):1706–15. doi:10.1002/eji.201243106

77. Lowes MA, Suarez-Farinas M, Krueger JG. Immunology of psoriasis. *Annu Rev Immunol* (2014) **32**:227–55. doi:10.1146/annurev-immunol-032713-120225

78. Artyomov MN, Munk A, Gorvel L, Korenfeld D, Cella M, Tung T, et al. Modular expression analysis reveals functional conservation between human Langerhans cells and mouse cross-priming dendritic cells. *J Exp Med* (2015) **212**(5):743–57. doi:10.1084/jem.20131675

79. Fairbairn L, Kapetanovic R, Beraldi D, Sester DP, Tuggle CK, Archibald AL, et al. Comparative analysis of monocyte subsets in the pig. *J Immunol* (2013) **190**(12):6389–96. doi:10.4049/jimmunol.1300365

80. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* (2014) **343**(6172):776–9. doi:10.1126/science.1247651

81. Usoskin D, Furlan A, Islam S, Abdo H, Lonnerberg P, Lou D, et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat Neurosci* (2015) **18**(1):145–53. doi:10.1038/nn.3881

82. Zeisel A, Manchado ABM, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* (2015) **347**(6226):1138–42. doi:10.1126/science.aaa1934

83. Crozat K, Vivier E, Dalod M. Crosstalk between components of the innate immune system: promoting anti-microbial defenses and avoiding immunopathologies. *Immunol Rev* (2009) **227**(1):129–49. doi:10.1111/j.1600-065X.2008.00736.x

84. Luber CA, Cox J, Lauterbach H, Fancke B, Selbach M, Tschopp J, et al. Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* (2010) **32**(2):279–89. doi:10.1016/j.immuni.2010.01.013

85. Miller JC, Brown BD, Shay T, Gautier EL, Jojic V, Cohain A, et al. Deciphering the transcriptional network of the dendritic cell lineage. *Nat Immunol* (2012) **13**(9):888–99. doi:10.1038/ni.2370

86. Wu X, Satpathy AT, Kc W, Liu P, Murphy TL, Murphy KM. Bcl11a controls Flt3 expression in early hematopoietic progenitors and is required for pDC development in vivo. *PLoS One* (2013) **8**(5):e64800. doi:10.1371/journal.pone.0064800

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.