Human and mouse transcriptome profiling identifies cross-species homology in pulmonary and lymph node mononuclear phagocytes

Sonia M. Leach[1,2*], Sophie L. Gibbings[3*], Anita D. Tewari[3], Shaikh M. Atif[3,8], Brian Vestal[1,2], Thomas Danhorn[1], William J. Janssen[6,7] Tor D. Wager[9] and Claudia V. Jakubzick[3,4,5]

[1]Center for Genes, Environment and Health, National Jewish Health, Denver, CO, [2]Department of Biomedical Research, National Jewish, Denver, CO, [3]Department of Pediatrics, National Jewish Health, Denver, CO, [4]Integrated Department of Immunology, National Jewish Health and UC Denver Anschutz Campus, Denver, CO, [5]Department of Microbiology and Immunology, Dartmouth College, Hanover, NH [6]Department of Medicine National Jewish Health, [7]Division of Pulmonary Sciences and Critical Care, University of Colorado, Denver, CO, [8]Department of Medicine, Division of Asthma, Allergy and Clinical Immunology, University of Colorado, Denver, CO, [9]Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, *Authors contributed equally

Running title: Homologs of Human and Mouse Mononuclear Phagocytes

Correspondence:
Claudia Jakubzick, PhD
Geisel School of Medicine at Dartmouth
Department of Microbiology and Immunology
628W Borwell
One Medical Center Drive
Lebanon, NH 03756
Phone: 303-898-3791
claudia.jakubzick@dartmouth.edu

Keywords: monocytes, macrophages, dendritic cells, pulmonary, lung

Abstract

The mononuclear phagocyte (MP) system consists of macrophages, monocytes, and dendritic cells (DCs). MP subtypes play distinct functional roles in steady-state and inflammatory conditions. Though murine MPs are well-characterized, their human homologs remain poorly understood. To address this gap, we created a gene expression compendium across 15-distinct human and 9-distinct murine MPs from lung, LN, blood, and spleen. This study identified corresponding human-mouse MP subtypes and determined marker genes shared or not between species counterparts. Unexpectedly, at the gene expression level, only 13-23% of the top 1000 marker genes overlapped in corresponding human-mouse MP counterparts, indicating a need for caution when translating mouse studies to human gene targets and functions. Lastly, CD88 was useful in both species to distinguish macrophage/monocytes from DCs. Our cross-species expression compendium serves as a resource for future translational studies to investigate beforehand whether pursuing specific MP subtypes, or gene will prove fruitful.

Introduction

Mice are the most commonly used organisms to study human diseases; they breed rapidly and can be genetically modified, accelerating the pace of discovery. A number of similarities with humans suggest that treatments developed in mice may apply to humans. As an example, both species have an immune system that contain mononuclear phagocytes (MPs), granulocytes, innate lymphoid cells, and lymphocytes [1]. Immune checkpoint blockades, a Nobel prize-winning discovery, were first observed in the immune system of mice and then translated for therapeutic use in humans [2]; and the development of GM-CSF therapy for pulmonary alveolar proteinosis (PAP) was influenced by observations in mice, where the deficiency of GM-CSF results in PAP [3]. Even with these positive translational outcomes, many failures in drug development have highlighted the profound functional differences between humans and mice [4].

Translational efforts have been challenging across multiple diseases, and there are numerous causes that contribute to these challenges. One key obstacle is the inability to identify functional cellular and molecular homologs in humans and mice. Often the targets identified in mice do not function similarly in humans. Conversely, targets identified in humans are frequently reverse translated into the study of inappropriate murine targets based on naïve assumptions about homology. Therefore, understanding cross-species homologies is crucial.

Assuming humans have equivalent MP subtypes to those of mice, understanding of how murine MPs interact with each other to maintain homeostasis and resolve inflammation could help decipher the mechanisms that go awry in human diseases, such as the ones described above. Murine MPs have been extensively characterized and demonstrate a clear division of labor during innate and adaptive immunity [5, 6, 7, 8, 9, 10, 11, 12]. Murine pulmonary MPs consist of alveolar macrophages (AMs), tissue trafficking monocytes, dendritic cell (DCs) subtypes [12, 13], and three interstitial macrophages (IMs): two IMs, IM1-(CD206^{hi}MHCII^{+}) and IM2-(CD206^{hi}MHCII^{lo}), display classical macrophage characteristics, whereas the third IM, IM3-(CD206^{lo}MHCII^{+}), although displaying macrophage properties, has a higher rate of turnover and expresses pro-inflammatory, monocytic and DC genes [7]. Very little is known about IMs, in humans how to identify them,

and in both species, humans and mice, how they functionally contribute to homeostasis, inflammation, or disease. Most notably, with the exception of AMs, all other MP subtypes described above reside in multiple organs [14, 15, 16, 17].

Functionally, circulating monocytes traditionally have been viewed as precursors to tissue-resident macrophages. However, now we know that monocytes continuously traffic through non-lymphoid and lymphoid tissue, where they survey the environment, and unless there is a macrophage niche to fill, steady-state monocytes do not differentiate into self-renewing, tissue resident macrophages [18, 19]. During inflammation, monocytes can differentiate into inflammatory and resolving macrophages, which display distinct properties from resident macrophages [10, 20, 21, 22]. In lymphoid tissue, even though LN monocytes are highly present, their role in adaptive immunity is less defined than DCs [19, 23].

DCs are potent antigen-presenting cells that link innate and adaptive immunity. In the periphery, DCs can acquire pathogens, traffic through lymphatic vessels to draining LNs, and present exogenous antigens to cognate T cells, precipitating adaptive immunity [11, 24, 25, 26, 27]. In mice, there are two main DC types referred to as Batf3^{+} DC1 and Irf4^{+} DC2. These DCs express distinct transcriptional factors that regulate their development (i.e., which they are named after), antigen acquisition, processing and presentation capabilities. All in all, although many of the functional cell types and properties of pulmonary MPs have been defined in mice, an assumption made is that similar functional cell types exist in the human lung.

To date, MP cross-species studies have been performed for human blood-mouse spleen and human-mouse heart [28, 29, 30, 31], and although a number of studies have examined the cell surface marker expression and performed single-cell or bulk RNA sequencing (RNA-seq) for human or murine pulmonary MPs [20, 32, 33, 34, 35, 36, 37, 38, 39], a transcriptional alignment of the human-murine pulmonary AMs, IMs, DCs, monocytes, and lymph node (LN) monocytes and DCs is still lacking. Here we created an expression compendium by RNA-seq to establish the non-diseased transcriptional profile of 15 distinct human and 9 distinct murine MPs, defined by our group and others [32, 33, 35, 36]. The focus is on pulmonary MPs, especially the understudied LN MPs, though as controls to validate the analytical quality of cross-species pulmonary MP analysis, human blood and

murine spleen MPs were used. We developed a strategy to determine the cross-species counterpart for each MP subtype, comparing their global expression profiles as well as contrasting sets of MP subtype marker genes; and demonstrated how the compendium and subtype analysis can be used to develop new cell surface markers to distinguish macrophage/monocytes versus DC lineages.

Results

Identification of human and murine MP populations

We isolated 15 distinct human MPs from four tissue locations: bronchoalveolar space, lung interstitium, and lung-LNs from non-diseased donors, and blood from healthy donors (Fig. 1A and Fig. S1). Note, not every donor provided all 15 MP cell types. MP subtype labels were putatively assigned based on extensively published cell surface markers considered to selectively identify macrophages or DCs [32, 33, 34, 35, 36, 39, 40, 41]. From each human lung, we isolated seven unique MP populations, including four distinct macrophage populations (AM, IM CD36, IM HLADR, and IM CD1c$^+$), two DC subsets (DC CD206$^-$ and DC CD206$^+$), and CD14$^+$ monocytes. The LNs contributed CD14$^+$ LN-monocytes and two LN-DC populations (DC CD1a^{hi} and DC CD1a^{lo}). Lastly, well-defined blood (BL) MPs included two monocyte populations (Mo CD14$^+$CD16$^-$ and Mo CD14$^+$CD16$^+$) and three classical CD11c$^+$HLADR$^+$ DCs (DC Clec9a$^+$, DC CD1c$^+$, and DC CD1c$^-$) [28, 36, 42]. The total number of human MP samples sorted and analyzed by RNA-seq was 66, with between three-to-six replicates per MP subtype.

Nine distinct murine MPs were isolated from four tissue locations, including the bronchoalveolar space, lung tissue, lung-draining LN and spleen (Fig. 1B and Fig. S2). We isolated four lung macrophage populations (AM, IM1, IM2 and IM3) and two LN migratory DCs (DC CD103, DC1 and DC CD11b, DC2), distinguished from lymphoid resident DCs by high MHCII expression. Two splenic DC populations (DC CD8$^+$, DC1 and DC CD11b$^+$CD4$^+$, DC2) and Ly6C^{hi} monocytes were sorted. Since mice lack detectable DCs in circulation, splenic DCs were included in this study as the comparators to human blood DCs. The spleen lacks afferent lymphatics exclusively representing a source of resident, lymphoid DCs, in contrast to tissue experienced, migratory DCs. The total number of murine MP samples sorted and analyzed by RNA-seq was 27, with three replicates per MP subtype.

Confirming human MP subtypes based on defined transcriptional profiles

To support our designations of sorted human MPs as either macrophages, monocytes or DCs, we examined the expression of classical macrophage/monocyte and

DC genes (Fig. 2) [43, 44, 45]. As anticipated, most of the macrophage/monocyte genes such as *MAFB*, Fc gamma receptors, lysosomal proteases and associated membrane proteins were most strongly expressed in human MP subtypes putatively labeled as macrophages or monocytes (Fig. 2) [44, 45, 46]. Conversely, classical DC genes such as *HLA-DR*, *CD1C, CCR7*, *CD40*, *CD80*, *LAMP3*, *IRF4*, *IRF8*, *FLT3* and *ZBTB46* were more strongly expressed in human pulmonary and LN MPs, putatively labeled as DCs (Fig. 2) [44, 45]. In fact, pulmonary and LN-DCs showed higher relative expression of more classical DC genes than blood DCs, which are less mature in the blood, where they lack the expression of *CCR7*. CCR7 is a key DC signature gene. DCs require CCR7 for both migration from tissue to LNs via afferent lymphatics and T cell zone localization [47]. Altogether, based on cell surface markers and transcriptional data, we were able to confirm the appropriate labeling of human pulmonary and LN MPs.

Combined human and murine MP data grouped subtypes by AM, IM, mono or DC lineage

To illustrate the relative correspondence between human and murine MP subtypes, we aligned human and murine MP expression data, borrowing methods originally developed for single-cell sequencing (scRNA-seq) to account for species-specific effects on gene expression [48]. Canonical correlation analysis (CCA), followed by alignment and t-Distributed Stochastic Neighbor Embedding (t-SNE) were applied to quantile-normalized human and murine MP data (Fig. S3-E4) [49]. The analysis used 885 genes shared between human and mouse from among the top 2000 most highly variable autosomal genes in either species, capturing similarities among global gene expression profiles without knowledge of MP subtype identity.

The combined human-mouse t-SNE plots illustrate that the species effect has been removed and the tissue of origin is well distributed across the clusters (Fig. 3A). The cell types show distinct clusters, with AMs and monocytes forming tighter clusters than IMs and DCs (Fig. 3A). To facilitate comparisons, samples were further grouped into six groups, labeled with an indication of their lineage (Fig. 3B). Group-1 (AM) contains AMs, contributed exclusively from BAL (Fig. 3B). Group-3 (Mono) is a pure cluster of monocytes across species, sampled from lung, blood and spleen. The IM (Group-2 IM) group separates from AMs, as previously observed in mice [7]. Thus, even despite AM and IM

shared expression of macrophage core signature genes, their overall gene expression is strongly influenced by their local environment, alveolus versus interstitium [7, 50]. Group-2 (IMs) also includes CD14$^+$ monocytes from the LN (Fig. 3B). This is not surprising since IM and LN-monocytes represent monocyte-derived populations that have matured in the lung environment [18, 51].

The DCs are distributed across three distinct groups: Group-4 (LN-DCs), Group-5 (DC2) and Group-6 (DC1) (Fig. 3B). LN-DCs in Group-4 form a distinct cluster, clearly separated from the other DCs, suggesting a maturation or functional program distinct from blood or splenic DCs. Group-5 (DC2) contains the human lung/blood DCs and murine splenic CD4$^+$ DCs, suggesting an overall DC2 (Irf4-dependent) lineage. Surprisingly, Group-5 also contains human lung CD1c$^+$ IMs and murine lung IM3s. These subtypes were isolated using classical IM cell surface markers, such as CD64, CD36, and MerTK, yet comparison of their global expression patterns suggests these two IM subtypes also exhibit some DC expression programs. Interestingly, there is another example of this in the skin, where a monocyte-derived Langerhans cell has macrophage and DC-like function (i.e., life-long self-renewal and CCR7 migration to LNs), and gene expression for macrophage and DC-specific transcription factors, *mafb* and *zbtb46* [44]. Lastly, Group-6 (DC1) contains human blood Clec9a$^+$ DCs and murine splenic CD8$^+$ DCs, the expected counterparts of the DC1 (Batf3-dependent) lineage (Fig. 3B) [52]. Overall, the unbiased computational analysis of global expression data demonstrates that across species the majority of MP subtypes group as expected by AM, IM, mono or DC lineage.

Marker gene analysis identifies the closest cross-species homologue per MP subtype

We developed a strategy to determine the maximally corresponding MP subtype between human and mouse. First, we determined a set of marker genes for each MP subtype within each species. Then, compared these marker gene sets across species to find the best counterpart. Marker genes are identified as highly and preferentially expressed within a subtype over all other subtypes within a species. Hence, 15 marker gene sets for human and 9 marker gene sets for mouse were selected from a dataset consisting of 11,959 genes across 93 samples. Each gene is ranked for each subtype by a score that multiples two factors: 1) measuring the level of expression in a given subtype

relative to the median expression over all subtypes; and 2) measuring how differentially expressed the gene is in that subtype over all other subtypes [53]. Thus, genes with scores earlier in the ranking for a subtype are most highly expressed and most differentially expressed for that subtype.

Each of the 15 human and 9 murine subtypes is tested against the remaining subtypes within their species. As a result, core signature genes at the level of macrophages, monocytes or DCs will not be considered highly ranked, since those genes are shared among many other subtypes at that same level, and thus receive a low score for differential expression. By considering each subtype separately and not grouping them at a higher level, we take advantage of the granularity at which the subtypes were isolated to find the maximally corresponding homologue based on unique genes specific to that subtype.

To determine the closest cross-species homologs, we developed a method based on a technique known as Correspondence-at-the-Top (CAT) plots [54]. CAT plots have previously been used to compare expression results across experimental platforms since they are agnostic to differences in underlying data distributions. In a CAT plot, two ranked marker lists are compared to determine the percentage of overlap of progressively larger lists, such as percentage overlap among the top genes (i.e., the top-5 genes, the top-10 genes, etc) (Fig. 4A). The marker list for each MP subtype is used as the reference and compared to each MP subtype marker list in the opposing species (examples illustrated in Fig. 4A, all human and mouse MPs in Fig. S5-S6). As seen in the CAT plots, the relationships between the curves generally stabilize after the top 1000 genes (vertical dashed line).

For any given top list size, we prefer a pair of marker lists with strong overlap at the top of the list and continued high overlap as the top list size increases, compared to a pair that only shows strong overlap later in the rankings. Thus, we choose to summarize the correspondence of each subtype pair across species as the mean percentage overlap of the corresponding marker gene lists for the pair across all top list sizes from 1 to 1000 (Fig. 4B). In this manner, pairs that agree quickly in the ranking and maintain agreement up to list size 1000 are distinguished from those that only achieve agreement later in the ranking. Of note, the clustering of MP subtypes seen in the previous t-SNE plot analysis

of global expression is echoed in this CAT plot analysis of marker gene sets. Overall, the average percentage of overlap between any pair across species for rank lists of size 1 to 1000 ranges between 3% and 23% (Fig. 4B). The closest cross-species homologue for a given MP subtype is chosen as that subtype in the opposing species with the highest mean percentage overlap, which ranges from 13% to 23%.

Finally, as another form of validation for our computational analysis, we analyzed cell types derived from our human blood and splenic MPs, and compared the populations with several recent studies that have profiled human blood MPs [55, 56] and splenic MPs [57] using single-cell RNA-seq (scRNA-seq) (Fig. S7). Based on the correspondence, we confirmed our cell types aligned with theirs, and conclude that our bulk-RNA-seq data is as good as the scRNA-seq data at aligning cell populations, with the advantage of higher-depth sequencing and analysis of comparison with more cell types than previously published.

Human MPs correspond to expected mouse MPs except CD1c$^+$ IMs and LN-monocytes

Most of the human MPs align best with a single murine cell type. For example, AMs correspond most highly across species, with a 13% mean overlap, especially early in the ranking among the top 200 marker genes (Fig. 4A and 4B, Group-1 AM). Human blood and lung monocytes correspond directly to murine splenic monocytes with 21% mean overlap (Group-3 Mono). Human blood Clec9a$^+$ DCs are most similar to murine splenic CD8$^+$ DCs, with 17% mean overlap (Group-6 DC1); unexpectedly, these Clec9a$^+$ DCs are not strongly aligned with murine LN CD103$^+$ DCs (7% mean overlap), as would be hypothesized based upon the shared Batf3-dependent lineage of CD8$^+$ splenic DCs and CD103$^+$ LN-DCs in the mouse. Instead LN-DCs are highly correspondent across species (Group-4 LN-DCs) suggesting that DC maturation state or tissue imprinting has a greater effect on gene expression than was previously appreciated.

On the other hand, there are a few human MPs that share homology with more than one mouse MP subtype. Human IM HLADR and IM CD36 are most similar to murine IM1 and IM2, with 13-16% mean overlap (Fig. S5 and Fig. 4B, Group-2 IM), demonstrating a homology across IMs, as expected. However, human IM CD1c is most similar to murine splenic CD4$^+$ DC (20% mean overlap) and less similar to its expected counterpart murine

IM3 (12% mean overlap) (Fig. S5 and Fig. 4B, Group-5 DC2). Note, however, inversely the murine IM3 is most similar to human IM CD1c, as anticipated (Fig. S5 and Fig. 4B).

In the LN, human LN-monocytes correspond highly with murine IM1 and IM2 (21% mean overlap) instead of splenic monocytes (13% mean overlap, 3rd highest) (Fig. 4A and Fig. 4B, Group-2 IM). This was evident at the t-SNE level (Fig. 3B), reflecting the shared characteristics of differentiated monocytes.

Murine MPs match expected human MPs, yet exhibit evidence of dual expression programs

We also performed the inverse correspondence assessment, taking each mouse MP and identifying the most similar human MP (Fig. S6 and Fig. 4). Murine AMs (Fig. 4B, Group-1 AM), are highly similar to human AMs (13% mean overlap), especially when comparing up to the top 200 genes (Fig. 4A) Although, among the top 200 to 1000 genes, human IM CD1c and IM HLADR (Group-2 IM) begin to overlap increasingly more (13-14% mean overlap), suggesting murine AMs share characteristics with human AMs and IMs. Murine LN-DCs and splenic monocytes were most similar to human LN-DCs and lung/blood monocytes (Group-4 LN-DCs ~20% mean overlap and Group-3 mono ~20% mean overlap, respectively), as anticipated.

Interestingly, where human blood Clec9a$^+$ DCs align most strongly with their expected counterpart mouse splenic CD8$^+$ DCs, the inverse is slightly different depending on the total number of *top genes* analyzed (Fig. 4, Group-6 DC1). Averaging the percent overlap over lists of size 1 to 1000, murine CD8$^+$ DCs correlate best with the human lung CD206$^-$ DCs (Group-5 DC2, 21% mean overlap) as opposed to Clec9a$^+$ DCs (17% mean overlap). Similar to the results for murine AMs, murine splenic CD8$^+$ DCs do match their expected homologue, Clec9a$^+$ DCs, best when considering only the top 50 ranked genes (Fig. 4A). This suggests that mouse splenic CD8$^+$ DCs may represent an intermediate state between human blood Clec9a$^+$ DCs and lung CD206$^-$ DCs, which are both relatively immature compared to tissue-experienced, migratory LN-DCs.

Murine IM1 and IM2 correspond best to human LN-monocytes, with 21% mean overlap, followed by human HLADR and CD36 IMs, with ~15% mean overlap (Fig. 4B Group-2 IM and Fig. S6). Murine IM3 corresponds highly with human LN-monocytes as

well as IM HLADR (11% mean overlap) (Fig. 4B Group-2 IM and Fig. S6), though IM3 matches most strongly with human IM CD1c (12% mean overlap) (Fig. 4B Group-5 DC2).

Overall, CAT plot analysis of marker gene sets highlights cross-species correspondences and defines the best overall counterpart for human and murine MPs within our MP subtypes (Fig. 4B). Where the predicted counterparts do not match expectations, the analysis gave evidence that the marker genes of the expected homologue did appear early in the ranking (i.e., at the top-50 marker genes) suggesting multiple MP expression programs exist in some subtypes when 1000 top marker genes are analyzed.

Conserved and distinct genes expressed in well-defined human-mouse homologs

Expression data for the top marker gene sets were examined in the best corresponding human-mouse MP pairs. The functional properties of genes are outlined in Table SII. We considered three MP cell types as well-defined human-mouse homologs: AMs, monocytes and DC1 (Fig. 5). For figures 5-7, the top 50 marker genes are shown in the reference MP subtype (Fig. S8-S31), and any of the 50 genes shared among the top 1000 marker genes in each of the top three closest matching subtypes in the other species are annotated based on relative closeness with *, ** or ***.

As anticipated, signature genes that define these cell types are present such as *FABP4*, *MARCO and PPARg* for AMs, *S100A8*, *S100A9,* and *CD14* for monocytes (Group-3 mono, Fig. 5A-5B), *XCR1, CADM1,* and *Clec9a*, along with a recently identified gene, *WDFY4*, required for cross-presentation for human blood Clec9a^{+} DCs and murine splenic CD8^{+} DCs (Group-6 DC1, Fig. 5C) [58]. Although we outlined genes known to be part of these human-murine homologs, it is important to note that many genes were not shared. For example, differences in gene expression for human-mouse AMs were *Serping1, CES1, MME, IL17RB, and C1q (not shown)*, which are solely expressed in human AMs and not mouse AMs; whereas *SPP1*, *MMP8,* and *CARD11* are solely expressed in mouse AMs and not human AMs (Fig. 5). Human blood Clec9a^{+} DCs and mouse splenic CD8^{+} DCs differentially expressed *IDO1* or *IDO*, in addition to classical murine DC1 genes such as *CD207* (langerin) and *CCR7,* which are not expressed in human blood DC1 (Fig. 5).

Conserved and distinct genes expressed in less-defined human-mouse MPs

Since this method of analysis supports previously identified human-mouse MP homologs, we next assessed less-defined homologs: pulmonary IMs, LN-monocytes, and pulmonary/LN-DCs. Human IM CD36 and IM HLADR were best matched with murine IM1, closely followed by murine IM2 (Group-2 IM). The two murine IMs have a classical IM signature and shared gene expression for *FOLR2, SEPP1, LYVE1, APOE, C1q (not shown), CCL7,* and *CCL2* (Fig. 6A). An example gene not shared is *SPP1,* highly expressed in human IMs (and murine AMs), but not murine IMs (Fig. 6A).

Inversely, murine IM1 and IM2 matched best with human LN-monocytes followed by human IM HLADR and IM CD36 (Fig. 6B). Unlike human blood and lung monocytes that correspond with murine splenic monocytes (Group-3 Mono), human LN-monocytes (Group-2 IM) reciprocally align best with murine IM1 and IM2, followed by splenic monocytes (Fig. 7A). Human LN-monocytes share classical murine monocyte genes such as *S100A8*, *S100A9* and *CD14* (not shown), as well as mature IM genes such as *LYVE1*, *FOLR2*, *MMP9*, *F13A1*, *C1Q*, *MAFB, MAF, TIMD4,* and *STAB1* (Fig. 7A). Interestingly, murine IM3 matched best with human IM CD1c expressing inflammatory monocyte genes *IL1b, C1qb, C1qc,* and *F13a1* (Fig. 6B, Group-5 DC2). However, human IM CD1c matched best with murine splenic CD4+ DCs (Fig. 6A), based on their shared non-classical DC genes i.e., cytoskeletal genes, chemokine receptor 6, and cell cycle genes, *CD1K1* and *CDC20* (Fig. 6A).

Similar to the alignment of monocytes and their implied state of maturation, human lung DCs align best with murine splenic $CD4^+$ DCs compared to LN-DCs (Fig. 7B and Fig. 3 and 5B, Group-5 DC2), expressing genes such as *CCR6*, *CD80*, *CD207*, and *AXL*. Consequently, human LN-DCs align best with murine LN-DCs (Fig. 3 and 5B, Group-4 LN-DCs) and expressed classical DC genes such as *ZEB1, CCL22, CCL17,* and *CCR7*, but not the human classical DC gene, *LAMP3* (Fig. 7). Overall, we highlighted some shared and non-shared genes among the top signature genes that align human-murine MP homologs.

A cell surface marker that allows a clear distinction between macrophage/monocytes and DCs in human and mouse

The rich expression compendium created in this study can be mined to identify new cell surface markers to isolate MP subtypes. In mice, the co-expression of F4/80, Ly6C, MerTK, and CD64 has helped distinguish macrophage/monocytes from DCs [32, 39, 50, 59]. However, humans do not have an F4/80 or Ly6C homolog, and we find the expression of MerTK to be variable, perhaps due to the lack of a reliable antibody.

Statistical analysis of our RNA-seq data from human macrophage/monocytes and DCs resulted in identification of CD88 as a viable cell surface marker that can differentiate macrophage/monocytes from DCs in humans, as has been elegantly illustrated in mice [60]. To validate this cross-species marker, human lung and LN MPs were plotted as CD1c versus CD88 to distinguish DCs from macrophage/monocytes (Fig. 8). Interestingly, we observed that human LN monocyte maturation based on increased HLADR expression correlated with the upregulation of CD11c and CD206 (Fig. 8), as observed in mice [15, 18, 61]. All in all, of the many commercially available antibodies against cell surface markers for macrophages, monocytes and DCs, we observed that the use of CD88, as well as CD64 (data not shown) [36, 50], assists in clarifying the monocyte/macrophage lineage from the DC lineage in human and mouse.

Discussion

The aim of this study was to investigate the homology between human and murine mononuclear phagocytes in order to aid in the translation of animal studies to human medicine and improve the experimental design of animal models to mimic human disease. Towards this goal, we created a gene expression compendium using RNA-seq analysis of MP subtypes from human and mouse, and a combination of computational methods to determine the closest cross-species counterparts. First, we identified marker gene signatures for each MP cell type and then compared these signatures against the MPs across species to identify homologous cell types. Overall, this dataset allowed us to explore the correspondence between human and murine MPs as well as to discover potential signature markers shared in common (or not) across human and murine MP subtypes.

The efficacy of our analysis to identify MP homologs was demonstrated by its *unbiased prediction of known homologs.* As expected, 1) human and murine alveolar macrophages (Group 1); 2) human blood and murine splenic monocytes (Group 3); and 3) human blood Clec9a^{+} DCs and murine splenic CD8^{+} DCs (Group 6) aligned best with each other, respectively, compared to all other MPs across species [28, 62]. These three pairs provided a useful reference for assessing the confidence with which our strategy could identify less defined MP subtype homologs.

Also reassuring is the preserved general distinction between MP lineages in both human and mouse. For example, in Figure 3 there appears to be a diagonal axis that separates monocyte/macrophage lineage cells from dendritic cells. Groups 1, 2, and 3 contain exclusively macrophages and monocytes and whilst groups 4, 5, and 6 contain DCs (Fig. 3, Fig. 4B), with two exceptions. The first exception is human LN monocytes in the IM-enriched Group 2, which as previously stated represents a more matured monocyte expressing many classical monocyte and differentiated macrophage genes. The second exception is human lung IM CD1c^{+}, which aligns more closely with murine splenic CD4^{+} DCs than its expected counterpart mouse lung IM3, although, conversely, mouse IM3 aligns best with lung IM CD1c^{+} (Fig. 4B).

After broadly dividing macrophage/monocytes from DC lineages, we observe a pattern of maturation that supports the concept that immature cells in circulation undergo differentiation first in the lung tissue and further in either the air space in the case of alveolar macrophages, or in the lymph node in the case of LN DCs and monocytes. Based on previous observations in humans and mice, monocytes isolated from lung tissue with no to low CD206 expression represent cells present in the capillaries of the lung or recently migrated into the tissue [18, 34, 35, 36]. Murine splenic monocytes have similar gene expression profiles as human blood monocytes. Similarly, monocytes newly recruited into tissue experience minimal changes in their gene profile, with some strong exceptions such as genes known to be associated with transmigration and early maturation (i.e., elevated HLADR/MHCII and CD206) [18]. The human blood and lung monocyte observation and mouse correlation are also reassuring in that it demonstrates a negligible effect on the transcriptome of cell extraction procedures from blood versus tissue digest.

Interestingly, human blood DCs and splenic DCs clustered away from human and mouse LN DCs, perhaps due to tissue imprinting. One such distinction is the expression of CCR7, which is a hallmark feature of DCs. Since human blood DCs lack the expression of CCR7, this suggests that they are still immature and require additional endothelial-tissue interaction to up-regulate the expression of CCR7, a receptor required for lymphatic migration and localization in the T cell zone of LNs [63]. Another observation is the expression of CD62L, a transmigration molecule for circulating monocytes and DCs [18]. CD62L is expressed at high levels in human blood DCs, but not in tissue DCs, similar observations were made in mouse for circulating preDCs [64]. Another alternative, to the concept of a human blood DC giving rise to human tissue DCs would be instead that they have a shared precursor cell giving rise to DCs in different compartments (i.e., blood versus tissue DCs). Hence, a human blood DC may not necessarily give rise to a tissue DC, no more than a murine lymphoid DC gives rise to a migratory, tissue DC, where the latter expresses more MHCII and co-stimulatory molecules at steady state.

What was surprising was the relatively small effect of DC lineage on subtype correlation. In mice, there is a strong distinction between DC1 (Batf3-dependent) and DC2 (Irf4-dependent) lineages, as seen for $CD8^+$ versus $CD4^+$ splenic DCs. The existence of two DC lineages in humans is supported by the alignment of blood $Clec9a^+$ DC with

murine $CD8^+$ splenic DC (Group 6), compared with $CD4^+$ splenic DCs. However, our ability to separate human lung or LN DCs into a given DC lineage appears to be masked by the stronger effects of tissue imprinting, as evidence by the distinct clustering of the LN DC (Group 4) from lung DCs (Group 5) and human blood/splenic DC1 (Group 6). In addition, we originally separated LN DCs based on CD1a expression because we continuously observed $CD1a^{hi}$ and $CD1a^{lo}$ $CD1c^+$ lung-draining LN DCs in over 50 individuals, making the assumption that they might be distinct cell types. However, it is now clear that although the sorted human LN DCs are distinct at the protein level, these populations are not transcriptional distinct. Further investigations on how to subset $DEC205^+CD1c^+$ lung-draining LN DCs using different cell surface markers is needed. More importantly, human LN DCs are understudied and even though we were unable to decipher more than one LN DC cluster, there is a wealth of genetic information provided in this datum. For instance, CCL17 and CCL22 expression is restricted to DC populations in lung and LNs, and not expressed on human blood DCs [65, 66, 67, 68]. This is also true in mice, where these chemokines distinguish DCs from other immune cell types and are important for T cell recruitment and interactions.

All in all, marker gene sets were defined for each of the 15 human and 9 mouse MP subtypes. To qualify as a marker gene for a given MP subtype, a gene had to be expressed primarily in just that one single subtype. Expressed genes shared across multiple closely related subtypes would be disregarded as a consequence of this analysis. An alternative approach would be to detect the broader differences between monocytes/macrophages versus DC lineages, which would require a different analysis set. In our preliminary investigation, we identified CD88 expression to be highly predictive of macrophage/monocytes compared to DCs in both human and mouse. Fortunately, there are commercially available antibodies against CD88 (and CD64, shown previously) that work well for separation of monocytes/macrophages from DCs in human and mouse [18, 36, 50, 60].

To note, we attempted several times to scRNA-seq human lung and LN MPs and observed that tissue-embedded cells (IMs and DCs) start the process of death shortly after sorting, unlike luminal cells (AMs and monocytes) (unpublished data). Hence, IM/DC subtypes originally viable and present at the sorting stage were not observed at the 10X

genomics scRNA-seq analytical stage. It is unclear why this occurs. One hypothesis is that the delicate nature of cells residing in tertiary space, requiring continuous survival signals from neighboring cells, making it challenging to scRNA-seq interstitial and LN MPs. A difference between 10X genomics scRNA-seq and bulk-RNA-seq is that the latter sorts directly into RNA-extraction buffer, not requiring minutes-to-hours of cell survival post-sort. However, we are positive that we, and others, will, or have, overcome this obstacle.

In conclusion, this study provides a platform to aid in the design of future translational studies at the MP subtype level and at the gene level. Awareness of cross-species homology, or lack thereof, can save a great amount of time and expense in medical research.

Materials and Methods

Human lung mononuclear phagocyte isolation

We received de-identified human lungs that were not used for organ-transplantation from the International Institute for the Advancement of Medicine (Edison, NJ, USA) and the University of Colorado Donor Alliance. We selected donors without a history of chronic lung disease and with reasonable lung function with a P_aO_2/F_iO_2 ratio of >225, a clinical history and X-ray that did not indicate infection, and limited time on a ventilator. Lungs from 7 individuals were processed for RNA sequencing; the mean donor age was 35.8 years (SD 24.7 years), 3 and 4 donors were female and male respectively. We also noted the cause of death, medical history and time of death. 4/7 donors were non-smokers (defined as never having smoked) 2 were light smokers (less than a pack a year) and 1 was a former smoker. Additionally, blood was drawn from 3 healthy, non-smoking volunteers, all female with a mean age of 30 years (SD 4.6). 10 different mononuclear phagocyte (MP) populations were sorted from each lung donor, and 5 different populations were sorted from each blood donor, as outlined in Figure 1. After data quality control analysis, this resulted in 73 total human MP samples. The Committee for the Protection of Human Subjects at National Jewish Health approved this research.

BAL, lung, lymph node, and PBMC cell preparation

Lungs were removed en bloc in the operating room and included the trachea, LNs and pulmonary vessels. Pulmonary arteries were perfused in the operating room with cold histidine-tryptophan-ketoglutarate (HTK) solution to preserve endothelial cell function and prevent intravascular clot formation. The lungs were submerged in HTK, and immediately shipped on ice. All lungs were processed within 24 hours of removal. The lungs were visually inspected for lesions or masses and were eliminated from the study if grossly abnormal. Peribronchial LNs were identified and removed. Procedure described in detail elsewhere (1, 2).

Bronchoalveolar lavage (*BAL*) was performed on the right middle lobe or lingula by completely filling the lobe three times with PBS and 5mM EDTA and then 3 times with PBS alone. After each instillation, lavage fluid was drained from the lung, collected and

pooled. *Lung* tissue digestion was performed by inflating the lobe with 4U/mL elastase solution in PBS containing Ca^{2+}, Mg^{2+}, HEPES and dextrose. Airways were clamped to prevent leakage of enzyme and the tissue was incubated at 37°C for 40 minutes. Manual disruption of the tissue was performed, firstly with scissors and secondly with a food processor in buffer containing fetal bovine serum. The resulting tissue homogenate was filtered through a 100µm nylon filter membrane to create a single-cell suspension. Density gradient centrifugation using 30% and 65% Percoll (Vendor) was used to deplete red blood cells and dead cells. *Lymph nodes* were identified and dissected from lung tissue, pooled, minced and enzymatically digested with 2.5 mg/mL collagenase D (Roche/Sigma) for 30 minutes at 37°C. Digested tissue was collected and filtered through a 100µm nylon filter membrane to create a single cell suspension

Peripheral blood mononuclear cells (PBMC) isolation, blood was obtained by venipuncture from healthy adults as per IRB approved protocol.

Murine lung, LLN, and spleen MP isolation

Transcriptome data from murine mononuclear phagocytes can be accessed from the NCBI Gene Expression Omnibus via accession numbers GSE132911 (monocyte and DCs) and GSE94135 (AMs and IMs).
Briefly, Naïve female C57Bl/6 mice were obtained from Charles River Laboratories. Mice were used at 6–10 weeks of age, housed in a specific pathogen-free environment at National Jewish Health (Denver, CO), and used in accordance with protocols approved by the Institutional Animal Care and Use Committee. Mice were euthanized by CO_2 asphyxiation and lungs were perfused with cold PBS via the heart. Lungs, LLN or spleen were then removed and subjected to enzymatic dissociation with Collagenase D, cells of interest were enriched using Miltenyi microbeads and then FACS sorted as shown in Figure 3. Each sample represents pooled cells from 5 mice; different cohorts of mice were

used to isolate lungs, LLNs and spleens. RNA was isolated using the RNeasy Mini Kit (Qiagen).

Mononuclear phagocyte enrichment

Cells of interest were positively enriched prior to FACS sorting using Miltenyi microbeads according to manufacturer's protocols. Blood monocytes and dendritic cells were enriched from PBMCs using biotin-conjugated anti-CD141 and anti-CD1c in combination with anti-biotin microbeads. Enrichment of lung and lymph node cells was achieved using PE conjugated anti-CD1c and biotin-conjugated anti-CD11c antibodies in combination with anti-PE and anti-biotin microbeads as indicated in figures.

FACS sorting and flow cytometry

BAL, enriched blood, lung and lymph node cells were re-suspended in PBS with 2mM EDTA, 0.1% BSA and 10% pooled human serum, and stained with fluorescently conjugated antibodies as described in Table SI. Dead cells were excluded by uptake of 4′,6-diamidino-2-phenylindole (DAPI). Cells were either sorted using a FACS Aria Fusion (BD) for RNAseq analysis or analyzed using an LSRII or LSR Fortessa (BD).

Human RNA-seq data

Isolated mononuclear phagocytes were immediately re-suspended in RLT buffer for RNA extraction using Qiashredder columns and the RNeasy micro kit (Qiagen) per manufacturer's instructions. Isolated total RNA was processed for next-generation sequencing (NGS) library construction as developed in the NJH Genomics Facility for analysis with a HiSeq 2500 (San Diego, CA, USA). A Clontech SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing (Mountain View, CA, USA) and Nextera XT (San Diego, CA, USA) kit were used. Briefly, library construction started from isolation of total RNA species, followed by SMARTer 1st strand cDNA synthesis, full length dscDNA amplification by LD-PCR, followed by purification and validation. After that, the samples were taken to the Nextera XT protocol where the sample is simultaneously fragmented and tagged with adapters followed by a limited cycle PCR that adds indexes. Once

validated, the barcoded-pooled libraries were sequenced using 1x50bp chemistry on the HiSeq 2500 as routinely performed by the NJH Genomics Facility.

RNA-Seq data generation

For human samples, FASTQ files were generated with the Illumina bcl2fastq converter (version 2.17). Nextera TruSight adapters were trimmed, and degenerate bases at the 3'-end, as well as highly degenerate reads were removed using skewer (version 0.2.2), which by default removes reads with a remaining length of less than 18 nt. For the mouse samples, basecalling, barcode demultiplexing and adapter trimming was performed using the Ion Torrent Suite software, and reads less than 30 nt in length were discarded. The quality of the reads was assessed using FastQC (version 0.11.5) http://www.bioinformatics.babraham.ac.uk/projects/fastqc]. The preprocessed sequence reads were mapped to the genome of the respective species (canonical chromosomes of release hg 19 for human and mm10 for mouse, respectively) using STAR (3). Reads mapping uniquely to each gene of the Ensembl 75 annotation (http://feb2014.archive.ensembl.org/) of the respective species were quantified with the featureCounts program from the subread software package (version 1.5.2, using parameters -s 0 -O --fracOverlap 0.5 (4).

Alignment and visualization of human-murine MP data

Human and murine quantile-normalized gene-expression data were aligned and jointly visualized using the Seurat pipeline, originally developed for single-cell sequencing and previously demonstrated to integrate human and murine single-cell data. The standard pipeline was applied here to the human and murine MP data using the Seurat 2.3.2 package for R version 3.4.2 (5). Briefly, the data were first normalized using the default methods in Seurat to yield natural-log transformed data. Then a set of 885 highly variable genes (hvg) was identified as shared among the 2000 most highly variable human genes and 2000 most highly variable mouse genes. The data were then scaled using the standard Z-score scaling before further dimensionality reduction. Canonical Correlation Analysis (CCA) was applied to the scaled hvg set to identify a lower dimensional space

of CC vectors that maximized the shared correlational structure across human and mouse data sets. Using the MetageneBicorPlot function to display the correlation strength for each CC vector, the optimal number of 12 CC vectors was identified for use in the dynamic time warping alignment step. The resulting single, aligned, low-dimensional space representing the integrated human and mouse data was visualized in 2-D using t-Distributed Stochastic Neighbor Embedding (t-SNE). (Figure 4, Supplementary E4).

Identifying human-mouse MP analogs

The corresponding MP subtype between species could be found by minimizing distances between sample groups using the quantile-normalized gene expression vectors, or distances in the lower dimensional projection defined by the aligned CC vectors. However, with so few replicates per sample group and a difference in gene variance structure across species, both these methods were found to be very sensitive to outliers and data scaling (data not shown). To compute an unbiased correspondence between human and murine MP, the data were analyzed instead based on overlap of marker gene sets identified within each species. The set of marker genes for a reference MP subtype in one species was ranked in decreasing order of marker score and compared to the similarly ranked marker genes in the subtypes of the other species. The proportion of genes in common between the reference marker ranked list and each candidate cell type ranked list was plotted against the list size as a Correspondence-at-the-Top (CAT) plot (6). (Figure 5A, Supplementary E5 and E6). Due to the ranking criterion, higher agreement at smaller list sizes, i.e., the top of the list, indicates two subtypes agree strongly on genes most highly and preferentially expressed in a single subtype per species. The cross-species analog was chosen as that which maximized the mean percentage agreement across all list sizes from 1 to 1000 (Figure 5B).

Identifying cell-type specific marker genes

Marker genes were identified as those significantly over-expressed primarily in one MP subtype versus all others within a species. Over-expression was quantified using the quantile normalized VST counts. The median expression was computed for each gene in

each subtype within a species and then scaled by the median of those expression values across all samples within the species. Scaling by the median expression helps distinguish between genes highly expressed primarily in a minority of subtypes from genes like housekeeping genes that are highly expressed in all subtypes.

Significance of over-expression in one MP subtype was determined by calculating a Z-score for each gene for each subtype of interest as follows. RNA-Seq counts from genes that had a human-mouse equivalent were filtered to include only genes with at least 3 samples having more than one read count per million aligned reads. Next, the counts were transformed to continuous values using the VST from the DESeq2 R package. For each gene, a linear mixed model using subtype for the fixed effects and a random intercept for subjects was fit using the lmerTest R package version 3.1-0 (7). An overall F-test, using Satterthwaite's degrees of freedom method, to determine which genes showed any evidence for differences in expression between the subtypes was performed, and then these p-values were adjusted using the Benjamini-Hochberg method (8). For genes with adjusted p-values ≤ 0.05 for the overall F-test, a meta-analysis approach was used to evaluate potential signature genes for each different MP subtype. For the subtype of interest, all pairwise comparisons were done using linear contrasts and t-tests (again with Satterthwaite's degrees of freedom) using the linear mixed model described above. The raw p-values were then combined using the Z-score method where p_i was used if the effect size for that pairwise comparison was positive (i.e. the expression was higher in the subtype of interest) or $\max(p_i, 1-p_i)$ was used if the expression was lower in the subtype of interest. The same process was repeated for the mouse samples. A higher positive Z-score indicates that the gene is expressed higher in the subtype of interest than the other subtypes, on average. A final marker score for each candidate gene per subtype was computed as the product of the median scaled expression data and the Z-score using that subtype as the subtype of interest. In this way, genes that were both highly and preferentially expressed in a single subtype would be ranked more highly by the marker score. Note, this strategy discounts genes like housekeeping genes that may be highly expressed in all subtypes but not preferentially in a given subtype.

List of Supplementary Materials List your supplementary items, including titles for each of the supplementary figures, tables, data files, movies or other files.

References

1. Masopust, D., Sivula, C.P. & Jameson, S.C. Of Mice, Dirty Mice, and Men: Using Mice To Understand Human Immunology. *J Immunol* **199**, 383-388 (2017).

2. Altmann, D.M. A Nobel Prize-worthy pursuit: cancer immunology and harnessing immunity to tumour neoantigens. *Immunology* **155**, 283-284 (2018).

3. Trapnell, B.C., Carey, B.C., Uchida, K. & Suzuki, T. Pulmonary alveolar proteinosis, a primary immunodeficiency of impaired GM-CSF stimulation of macrophages. *Curr Opin Immunol* **21**, 514-521 (2009).

4. Begley, C.G. & Ellis, L.M. Drug development: Raise standards for preclinical cancer research. *Nature* **483**, 531-533 (2012).

5. Guilliams, M., Lambrecht, B.N. & Hammad, H. Division of labor between lung dendritic cells and macrophages in the defense against pulmonary infections. *Mucosal Immunol* **6**, 464-473 (2013).

6. Desch, A.N. *et al.* Dendritic cell subsets require cis-activation for cytotoxic CD8 T-cell induction. *Nat Commun* **5**, 4674 (2014).

7. Gibbings, S.L. *et al.* Three Unique Interstitial Macrophages in the Murine Lung at Steady State. *Am J Respir Cell Mol Biol* **57**, 66-76 (2017).

8. Atif, S.M. *et al.* Cutting Edge: Roles for Batf3-Dependent APCs in the Rejection of Minor Histocompatibility Antigen-Mismatched Grafts. *J Immunol* **195**, 46-50 (2015).

9. Kim, T.S., Gorski, S.A., Hahn, S., Murphy, K.M. & Braciale, T.J. Distinct dendritic cell subsets dictate the fate decision between effector and memory CD8(+) T cell differentiation by a CD24-dependent mechanism. *Immunity* **40**, 400-413 (2014).

10. McCubbrey, A.L. *et al.* Deletion of c-FLIP from CD11b(hi) Macrophages Prevents Development of Bleomycin-induced Lung Fibrosis. *Am J Respir Cell Mol Biol* **58**, 66-78 (2018).

11. Atif, S.M. *et al.* Immune Surveillance by Natural IgM is Required For Early Neoantigen Recognition and Initiation of Adaptive Immunity. *Am J Respir Cell Mol Biol* (2018).

12. Tussiwand, R. *et al.* Klf4 expression in conventional dendritic cells is required for T helper 2 cell responses. *Immunity* **42**, 916-928 (2015).

13. Murphy, K.M. Transcriptional control of dendritic cell development. *Adv Immunol* **120**, 239-267 (2013).

14. Tamoutounour, S. *et al.* Origins and functional specialization of macrophages and of conventional and monocyte-derived dendritic cells in mouse skin. *Immunity* **39**, 925-938 (2013).

15. Tamoutounour, S. *et al.* CD64 distinguishes macrophages from dendritic cells in the gut and reveals the Th1-inducing role of mesenteric lymph node macrophages during colitis. *Eur J Immunol* **42**, 3150-3166 (2012).

16. Epelman, S. *et al.* Embryonic and adult-derived resident cardiac macrophages are maintained through distinct mechanisms at steady state and during inflammation. *Immunity* **40**, 91-104 (2014).

17. Plantinga, M. *et al.* Conventional and monocyte-derived CD11b(+) dendritic cells initiate and maintain T helper 2 cell-mediated immunity to house dust mite allergen. *Immunity* **38**, 322-335 (2013).

18. Jakubzick, C. *et al.* Minimal differentiation of classical monocytes as they survey steady-state tissues and transport antigen to lymph nodes. *Immunity* **39**, 599-610 (2013).

19. Jakubzick, C.V., Randolph, G.J. & Henson, P.M. Monocyte differentiation and antigen-presenting functions. *Nat Rev Immunol* **17**, 349-362 (2017).

20. Mould, K.J. *et al.* Cell Origin Dictates Programming of Resident versus Recruited Macrophages during Acute Lung Injury. *Am J Respir Cell Mol Biol* **57**, 294-306 (2017).

21. Mould, K.J., Jackson, N.D., Henson, P.M., Seibold, M. & Janssen, W.J. Single cell RNA sequencing identifies unique inflammatory airspace macrophage subsets. *JCI Insight* **4** (2019).

22. Misharin, A.V. *et al.* Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J Exp Med* **214**, 2387-2404 (2017).

23. Larson, S.R. *et al.* Ly6C(+) monocyte efferocytosis and cross-presentation of cell-associated antigens. *Cell Death Differ* **23**, 997-1003 (2016).

24. Vermaelen, K.Y., Carro-Muino, I., Lambrecht, B.N. & Pauwels, R.A. Specific migratory dendritic cells rapidly transport antigen from the airways to the thoracic lymph nodes. *J Exp Med* **193**, 51-60 (2001).

25. Jakubzick, C., Tacke, F., Llodra, J., van Rooijen, N. & Randolph, G.J. Modulation of dendritic cell trafficking to and from the airways. *J Immunol* **176**, 3578-3584 (2006).

26. Jakubzick, C., Helft, J., Kaplan, T.J. & Randolph, G.J. Optimization of methods to study pulmonary dendritic cell migration reveals distinct capacities of DC subsets to acquire soluble versus particulate antigen. *J Immunol Methods* **337**, 121-131 (2008).

27. Desch, A.N. *et al.* CD103+ pulmonary dendritic cells preferentially acquire and present apoptotic cell-associated antigen. *J Exp Med* **208**, 1789-1797 (2011).

28. Ingersoll, M.A. *et al.* Comparison of gene expression profiles between human and mouse monocyte subsets. *Blood* **115**, e10-19 (2010).

29. Bajpai, G. *et al.* The human heart contains distinct macrophage subsets with divergent origins and functions. *Nat Med* **24**, 1234-1245 (2018).

30. Bachem, 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* **207**, 1273-1281 (2010).

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

32. Yu, Y.A. *et al.* Flow Cytometric Analysis of Myeloid Cells in Human Blood, Bronchoalveolar Lavage, and Lung Tissues. *Am J Respir Cell Mol Biol* (2015).

33. Yu, Y.A. & Tighe, R.M. Isolation and Characterization of Human Lung Myeloid Cells. *Methods Mol Biol* **1809**, 111-119 (2018).

34. Gibbings, S.L. & Jakubzick, C.V. Isolation and Characterization of Mononuclear Phagocytes in the Mouse Lung and Lymph Nodes. *Methods Mol Biol* **1809**, 33-44 (2018).

35. Gibbings, S.L. & Jakubzick, C.V. A Consistent Method to Identify and Isolate Mononuclear Phagocytes from Human Lung and Lymph Nodes. *Methods Mol Biol* **1799**, 381-395 (2018).

36. Desch, A.N. *et al.* Flow Cytometric Analysis of Mononuclear Phagocytes in Non-diseased Human Lung and Lung-draining Lymph Nodes. *Am J Respir Crit Care Med* (2015).

37. Reyfman, P.A. *et al.* Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* (2018).

38. Bharat, A. *et al.* Flow Cytometry Reveals Similarities Between Lung Macrophages in Humans and Mice. *Am J Respir Cell Mol Biol* **54**, 147-149 (2016).

39. Misharin, A.V., Morales-Nebreda, L., Mutlu, G.M., Budinger, G.R. & Perlman, H. Flow cytometric analysis of macrophages and dendritic cell subsets in the mouse lung. *Am J Respir Cell Mol Biol* **49**, 503-510 (2013).

40. Bharat, A. *et al.* Flow Cytometry Reveals Similarities Between Lung Macrophages in Humans and Mice. *Am J Respir Cell Mol Biol* (2015).

41. Guilliams, M. *et al.* Unsupervised High-Dimensional Analysis Aligns Dendritic Cells across Tissues and Species. *Immunity* **45**, 669-684 (2016).

42. Huysamen, C., Willment, J.A., Dennehy, K.M. & Brown, G.D. CLEC9A is a novel activation C-type lectin-like receptor expressed on BDCA3+ dendritic cells and a subset of monocytes. *J Biol Chem* **283**, 16693-16701 (2008).

43. Scott, C.L., Henri, S. & Guilliams, M. Mononuclear phagocytes of the intestine, the skin, and the lung. *Immunol Rev* **262**, 9-24 (2014).

44. Wu, X. *et al.* Mafb lineage tracing to distinguish macrophages from other immune lineages reveals dual identity of Langerhans cells. *J Exp Med* **213**, 2553-2565 (2016).

45. Satpathy, A.T. *et al.* Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages. *J Exp Med* **209**, 1135-1152 (2012).

46. Tussiwand, R. *et al.* Compensatory dendritic cell development mediated by BATF-IRF interactions. *Nature* **490**, 502-507 (2012).

47. Forster, R. *et al.* CCR7 coordinates the primary immune response by establishing functional microenvironments in secondary lymphoid organs. *Cell* **99**, 23-33 (1999).

48. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).

49. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193 (2003).

50. Gautier, E.L. *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat Immunol* **13**, 1118-1128 (2012).

51. Swirski, F.K. *et al.* Identification of splenic reservoir monocytes and their deployment to inflammatory sites. *Science* **325**, 612-616 (2009).

52. Hildner, K. *et al.* Batf3 deficiency reveals a critical role for CD8alpha+ dendritic cells in cytotoxic T cell immunity. *Science* **322**, 1097-1100 (2008).

53. Rosenthal, R. Combining Results of Independent Studies. *Psychological Bulletin* **85**, 185-193 (1978).

54. Irizarry, R.A. *et al.* Multiple-laboratory comparison of microarray platforms. *Nat Methods* **2**, 345-350 (2005).

55. Villani, A.C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356** (2017).

56. Dutertre, C.A. *et al.* Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity* **51**, 573-589 e578 (2019).

57. Brown, C.C. *et al.* Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell* **179**, 846-863 e824 (2019).

58. Theisen, D.J. *et al.* WDFY4 is required for cross-presentation in response to viral and tumor antigens. *Science* **362**, 694-699 (2018).

59. Atif, S.M., Gibbings, S.L. & Jakubzick, C.V. Isolation and Identification of Interstitial Macrophages from the Lungs Using Different Digestion Enzymes and Staining Strategies. *Methods Mol Biol* **1784**, 69-76 (2018).

60. Nakano, H. *et al.* Complement receptor C5aR1/CD88 and dipeptidyl peptidase-4/CD26 define distinct hematopoietic lineages of dendritic cells. *J Immunol* **194**, 3808-3819 (2015).

61. Langlet, C. *et al.* CD64 expression distinguishes monocyte-derived and conventional dendritic cells and reveals their distinct role during intramuscular immunization. *J Immunol* **188**, 1751-1760 (2012).

62. Caminschi, I. *et al.* The dendritic cell subtype-restricted C-type lectin Clec9A is a target for vaccine enhancement. *Blood* **112**, 3264-3273 (2008).

63. Randolph, G.J., Angeli, V. & Swartz, M.A. Dendritic-cell trafficking to lymph nodes through lymphatic vessels. *Nat Rev Immunol* **5**, 617-628 (2005).

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

65. Alferink, J. *et al.* Compartmentalized production of CCL17 in vivo: strong inducibility in peripheral dendritic cells contrasts selective absence from the spleen. *J Exp Med* **197**, 585-599 (2003).

66. Stutte, S. *et al.* Requirement of CCL17 for CCR7- and CXCR4-dependent migration of cutaneous dendritic cells. *Proc Natl Acad Sci U S A* **107**, 8736-8741 (2010).

67. Rapp, M. *et al.* CCL22 controls immunity by promoting regulatory T cell communication with dendritic cells in lymph nodes. *J Exp Med* **216**, 1170-1181 (2019).

68. Vulcano, M. *et al.* Dendritic cells as a major source of macrophage-derived chemokine/CCL22 in vitro and in vivo. *Eur J Immunol* **31**, 812-822 (2001).

Acknowledgments

This manuscript is dedicated to the lung donors and their families that enabled this research. Thanks to Donna Bratton, Peter Henson and Elizabeth Castel for project discussions and technical support isolating human myeloid cells. Funding: NIH R01 HL135001 and R01 HL115334.

Authors contributions

SML, TDW CVJ prepared the manuscript and designed the experiments. SLG, AT and SMA executed sorting, flow cytometry, and IHC experiments. SML, TD and BV executed the bioinformatics and biostatistics data. All authors provided intellectual input and critical feedback.

Conflict of Interest Disclosures

The authors declare no competing interests.

Figures

Figure 1. Gating strategy to sort human and mouse mononuclear phagocytes for RNA isolation. A) Cell suspensions were magnetically enriched, as indicated, from bronchoalveolar lavage (BAL), lung digest, lung-draining lymph nodes (LLN), or blood (BL). Fifteen different populations of mononuclear phagocytes (MPs, labeled in red) were FACS sorted, as shown, after gating to select Live; Single; $CD45^+$; $Lineage^-$ (Lin) cells. Side scatter properties (SSC) were used to separate AMs-SSC^{hi} from other MPs-SSC^{int}. BAL AMs were identified by co-expression of CD43 and HLA-DR and lack of CD14. Pulmonary Interstitial macrophages (IM) were defined as $CD206^+$ and $CD14^+$ and separated into three populations based on their differential expression of CD36, HLA-DR and CD1c. Pulmonary monocytes were taken from the $CD206^-$ fraction (Lu Mo). In lung and LLN, dendritic cells (DC) were characterized as $CD14^-$, $CD36^-$ cells that express CD1c and were separated into two populations in each organ based on the differential expression of CD206 and CD1a. In the LLN, DEC205 was also used to define DCs. Similarly, in blood CD14 and CD16 expression was used to separate monocytes (Mo) from DCs. BL DC subpopulations were defined on the basis of Clec9a or CD1c expression. **B)** Cell suspensions were magnetically enriched as indicated from bronchoalveolar lavage (BAL), lung digest, lung-draining lymph nodes (LLN), or spleen. Nine different populations of mononuclear phagocytes (MPs, labeled in red) were FACS sorted as shown, after gating to select Live; Single; $CD45^+$; $Lineage^-$ (Lin) cells. Alveolar macrophages (AM) co-express CD11c and Siglec F in the BAL. In the lung, interstitial macrophages (IM) are defined by expression of MerTK, CD64 and CD11b, and can be divided into three subpopulations based on the differential expression of CD206, MHCII and CD11c. In the LLN and spleen DCs co-express CD11c and MHCII separated by differential expression of either CD103 or CD8 (DC1) and CD11b (DC2). Monocytes are distinguished from other splenic MPs by expression of CD11b and Ly6C.

Figure 2. Confirmation of human MP subtypes based on classical gene signatures. Expression of signature DC and macrophage transcripts from RNA-seq are shown as normalized values. Data first undergo a variance stabilizing transformation on read

counts, where the donor identity has been regressed out to mitigate any donor effect. The data for each gene are then min-max scaled into the range [0,1] to emphasize which MP subtypes have the minimum or maximum expression for a given gene. Three replicates are shown for each subtype.

Figure 3. Visualization of MP subtypes across species.

Quantile-normalized VST subject-regressed data from 3-6 donors per subtype for 15 human and 9 murine MP subtypes was subjected to the Seurat v2.0 pipeline to visualize human and mouse data within a common two-dimensional space. A) Samples in the resulting 2-D projection are colored by species, tissue, or cell type. B) Samples are grouped into six clusters based on later shared marker gene analysis (Figure 5). Shape indicates cell type, size indicates species and color indicates group assignment. MP subtypes are denoted by the letters A through X. Note each group contains related subtypes across species.

Figure 4. Correspondence of MPs across species.

Candidate marker genes were ranked by their score, which multiplies their expression levels scaled relative to the median expression by their marker gene Z-score (see Methods). **A)** For each reference subtype (top row human reference to all mouse MPs, bottom row mouse reference to all human MPs), the proportion of agreement of ranked marker gene lists versus ranked marker gene lists for each candidate subtype in the other species was calculated for progressively larger list sizes and visualized with a Correspondence At The Top (CAT) plot for a select set of MP subtypes. Lines are labeled by letters from Figure 4 indicating the subtype compared to the reference shown in the title of the figure. A vertical line indicates which subtype had the highest mean percentage agreement across all list sizes from 1 to 1000 with respect to the reference. **B)** The mean percentage agreement over list sizes 1 to 1000 is shown for each subtype pair. Rows are scaled by the max and the min to emphasize the highest subtype correspondence for the human subtypes. Rows and columns are ordered to cluster like subtypes. Clusters of MP

subtypes are then annotated by group number and these groupings are used for the coloring in Figure 4B.

Figure 5. Human cell types with a single highest mouse equivalent.

Expression is shown for the first 50 ranked genes using the scaled data as in Figure 6. Three replicates for the three closest subtypes are shown. Genes are annotated by whether they are also in the top 1000 marker genes in the three highest corresponding subtypes with *, ** or ***, respectively. A) Highest correspondence for human and mouse AMs. B) Reciprocal highest correspondence for human Bl monocytes and mouse Spl monocytes. C) Highest correspondence for human Bl DC Clec9a^{+} and mouse Spl CD8^{+}.

Figure 6. Corresponding human and mouse IM subtypes.

Expression is shown for the first 50 ranked genes using the scaled data, as in Figure 6. Three replicates for the reference and the three closest corresponding subtypes are shown. Genes are annotated by whether they are also in the top 1000 marker genes in the three highest corresponding subtypes with *, ** or ***, respectively. A) human IMs. B) mouse IMs.

Figure 7. Corresponding human and mouse monocytes and DC subtypes.

Expression is shown for the first 50 ranked genes using the scaled data as in Figure 6. Three replicates for the reference and the three closest corresponding subtypes are shown. Genes are annotated by whether they are also in the top 1000 marker genes in the three highest corresponding subtypes with *, ** or ***, respectively. A) human monocytes B) human lung DCs C) human and mouse LN-DCs.

Figure 8. CD88 expression distinguishes monocytes and macrophages from DCs in both human and mouse.

Flow cytometry was used to assay CD88 expression on human (left) and mouse (right) MPs from lung (top) and lung lymph node (LN, bottom). All plots were first gated to exclude dead cells, CD45^{-} cells, doublets, lymphocytes, and granulocytes (Lin). In human lung, side scatter and HLA-DR expression is used to identify pulmonary interstitial MPs,

then CD1c and CD88 can be used to separate dendritic cells (DCs) from macrophages and monocytes (Mo-Macs). Similarly, in the human LN, after gating on CD11c$^+$ MPs, CD88 distinguishes Mo-Macs from CD1c$^+$ DCs. The second plot shows that LN-monocytes up-regulate both CD206 and HLA-DR. In the mouse lung, CD206 and CD11b can be used to identify extravascular pulmonary MPs and these similarly show expression of CD88 on Mo-Macs but not DCs. Mouse LN MPs are identified by expression of either CD11c or CD11b.

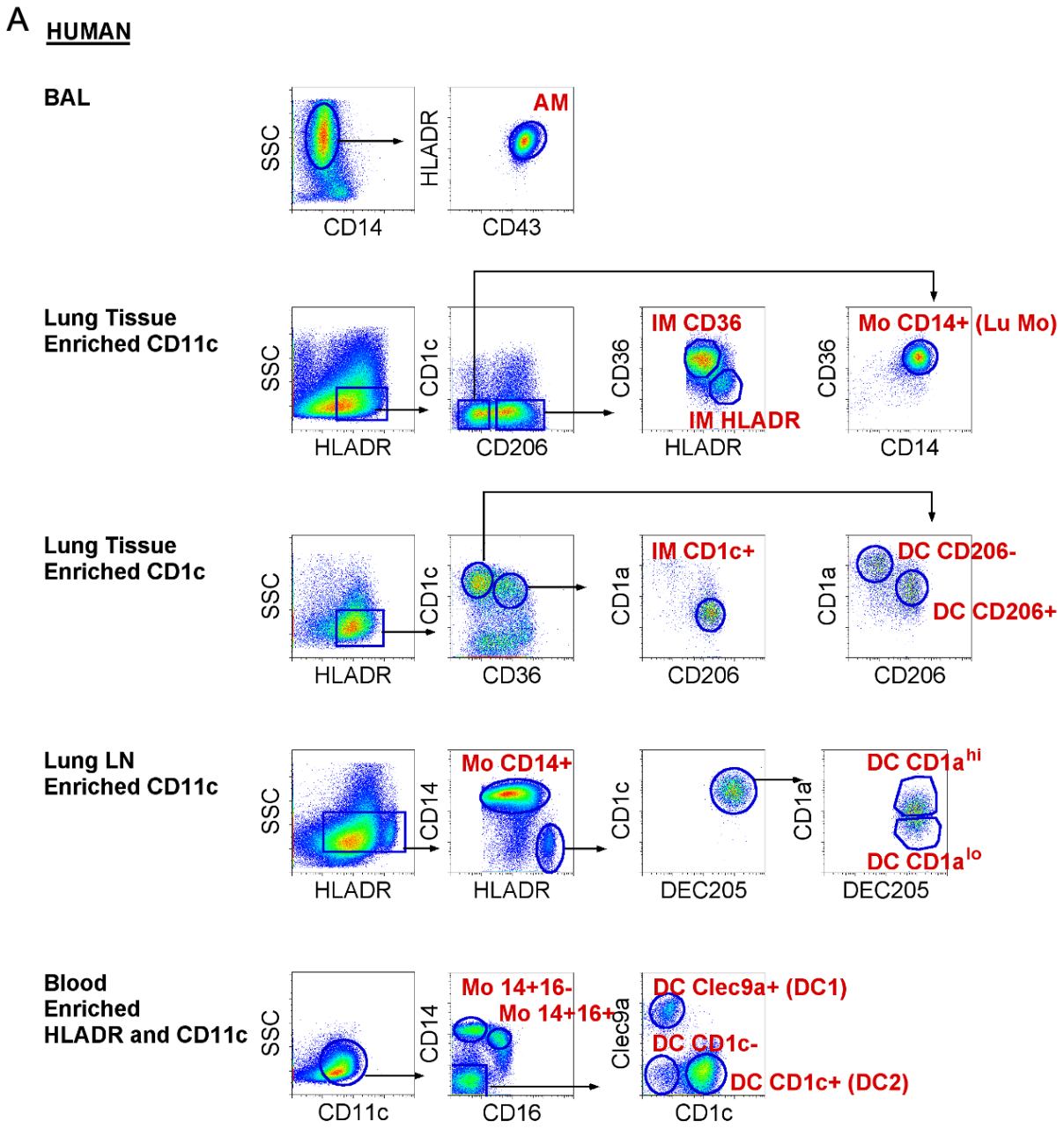

Figure 1A. Gating strategy to sort human and mouse mononuclear phagocytes for RNA isolation.

B

MOUSE

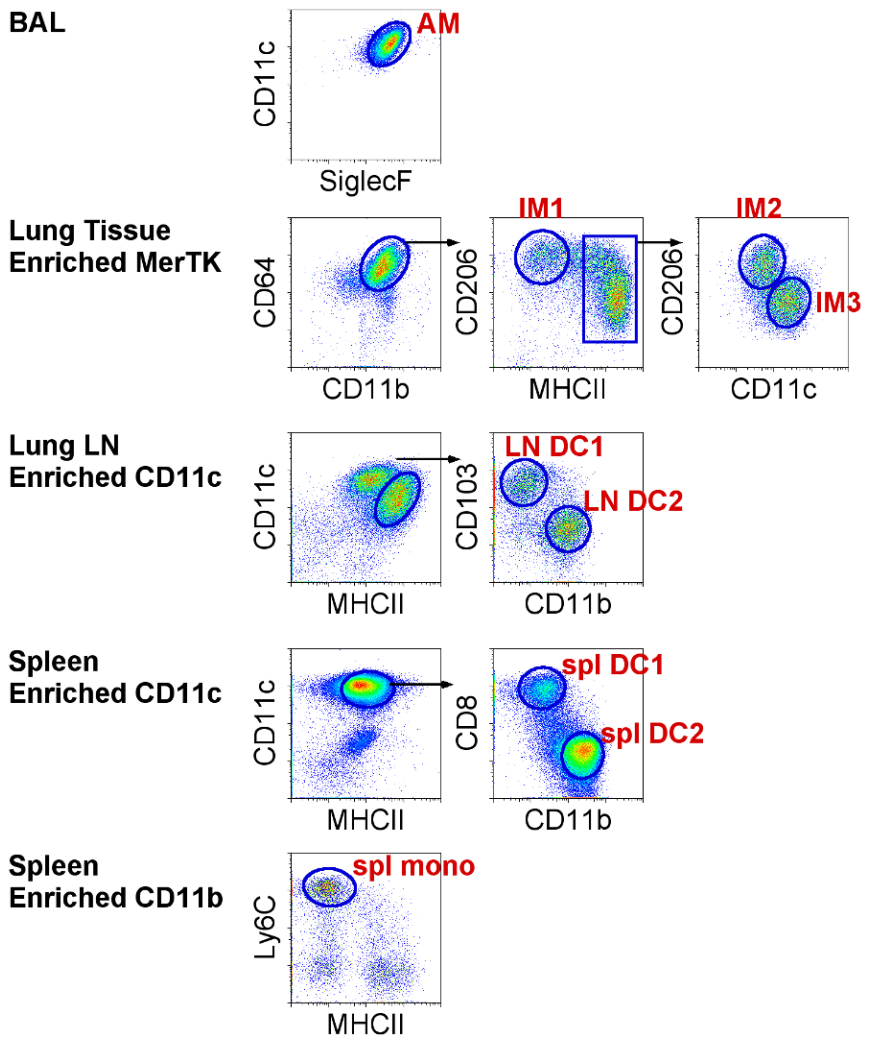

Figure 1B. Gating strategy to sort human and mouse mononuclear phagocytes for RNA isolation.

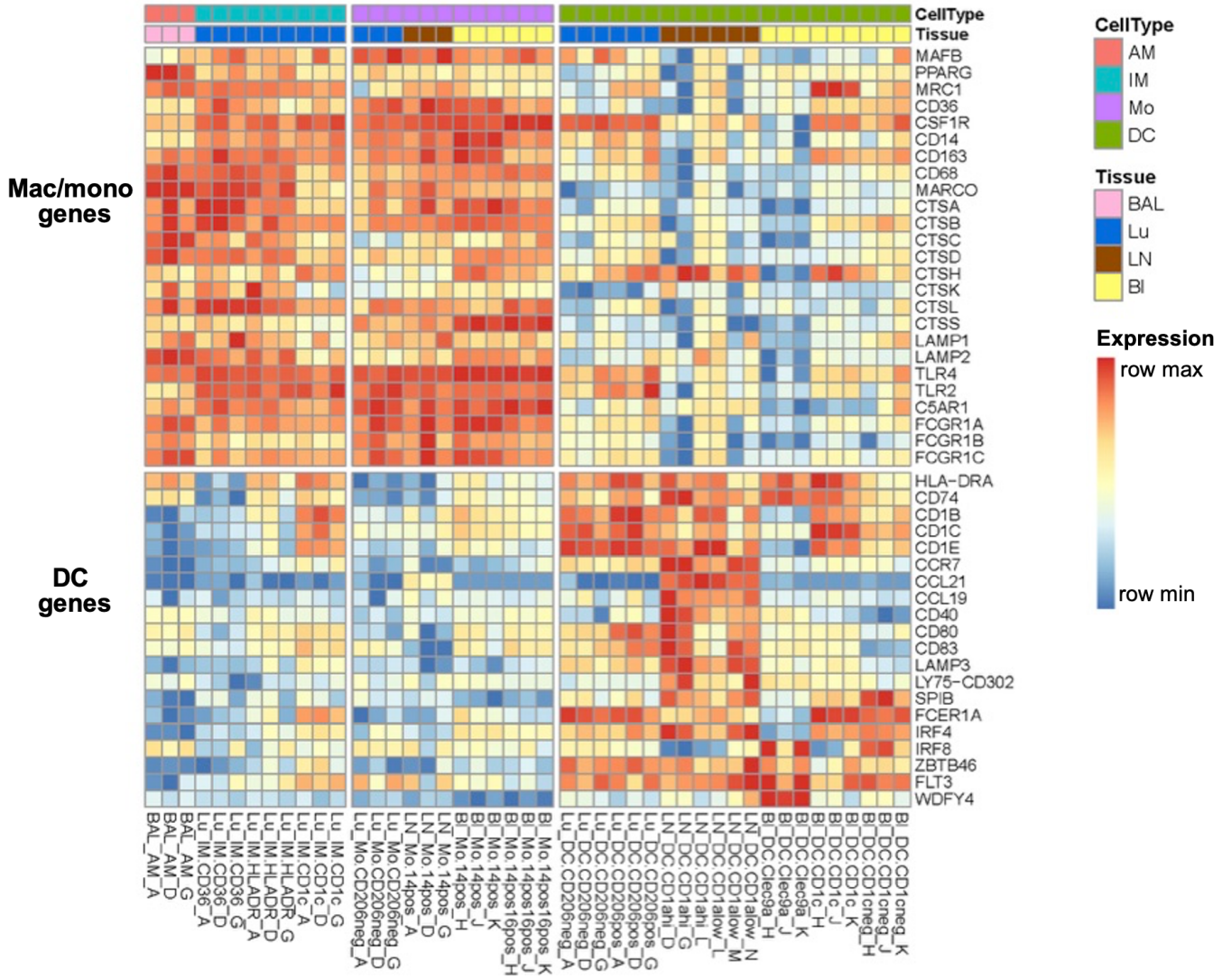

Figure 2. Confirmation of human MP subtypes based on classical gene signatures.

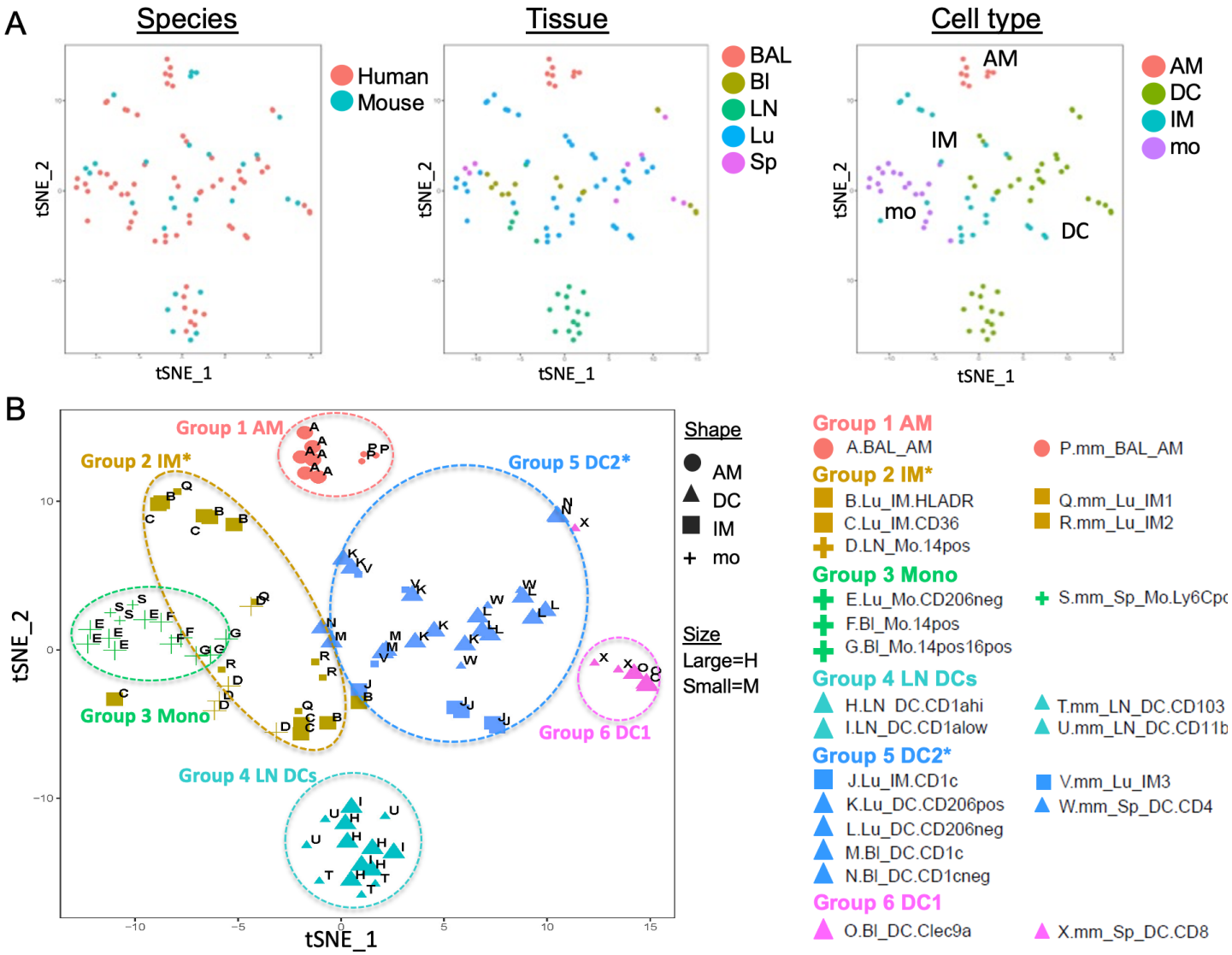

Figure 3. Visualization of MP subtypes across species.

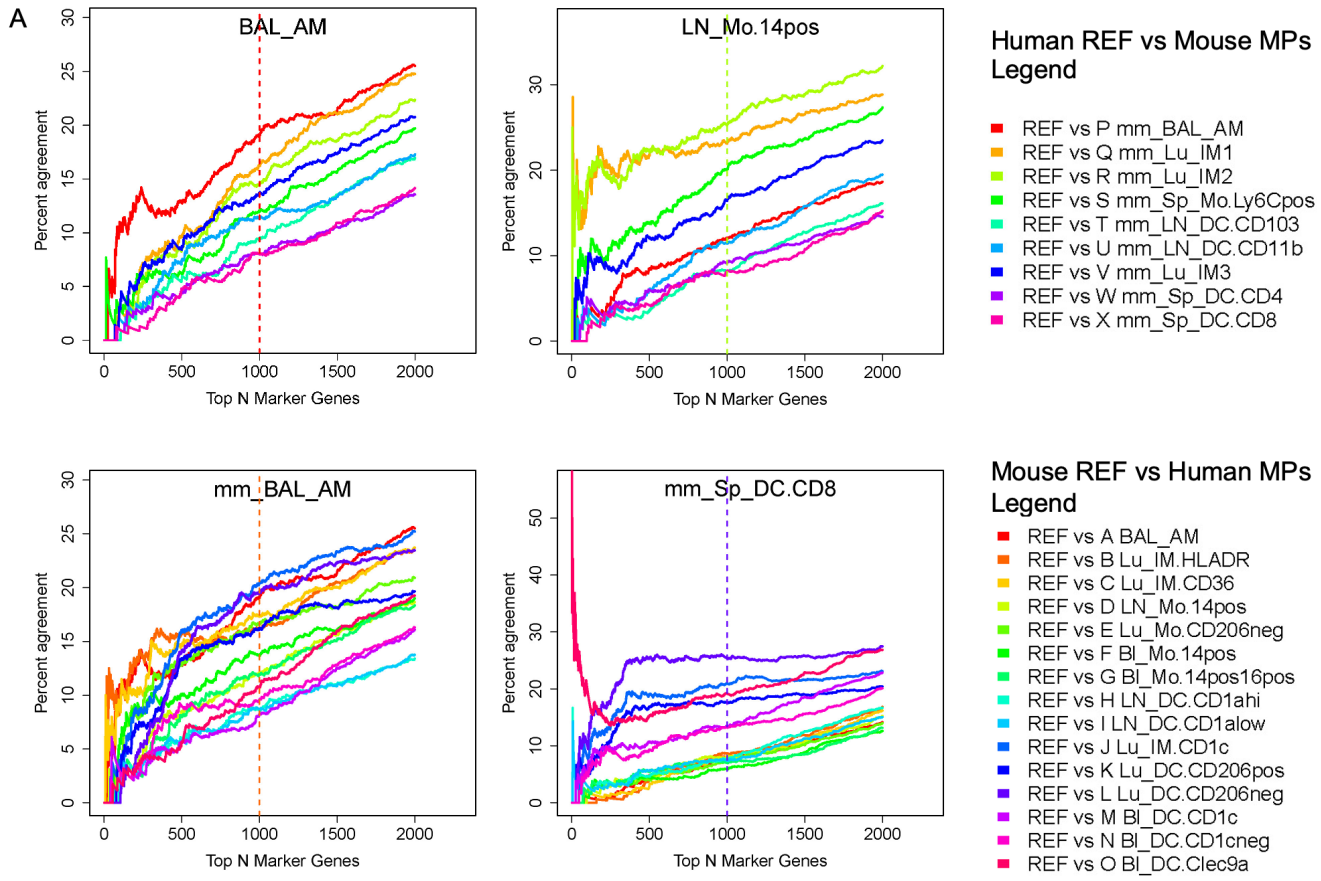

Figure 4A. Correspondence of MPs across species.

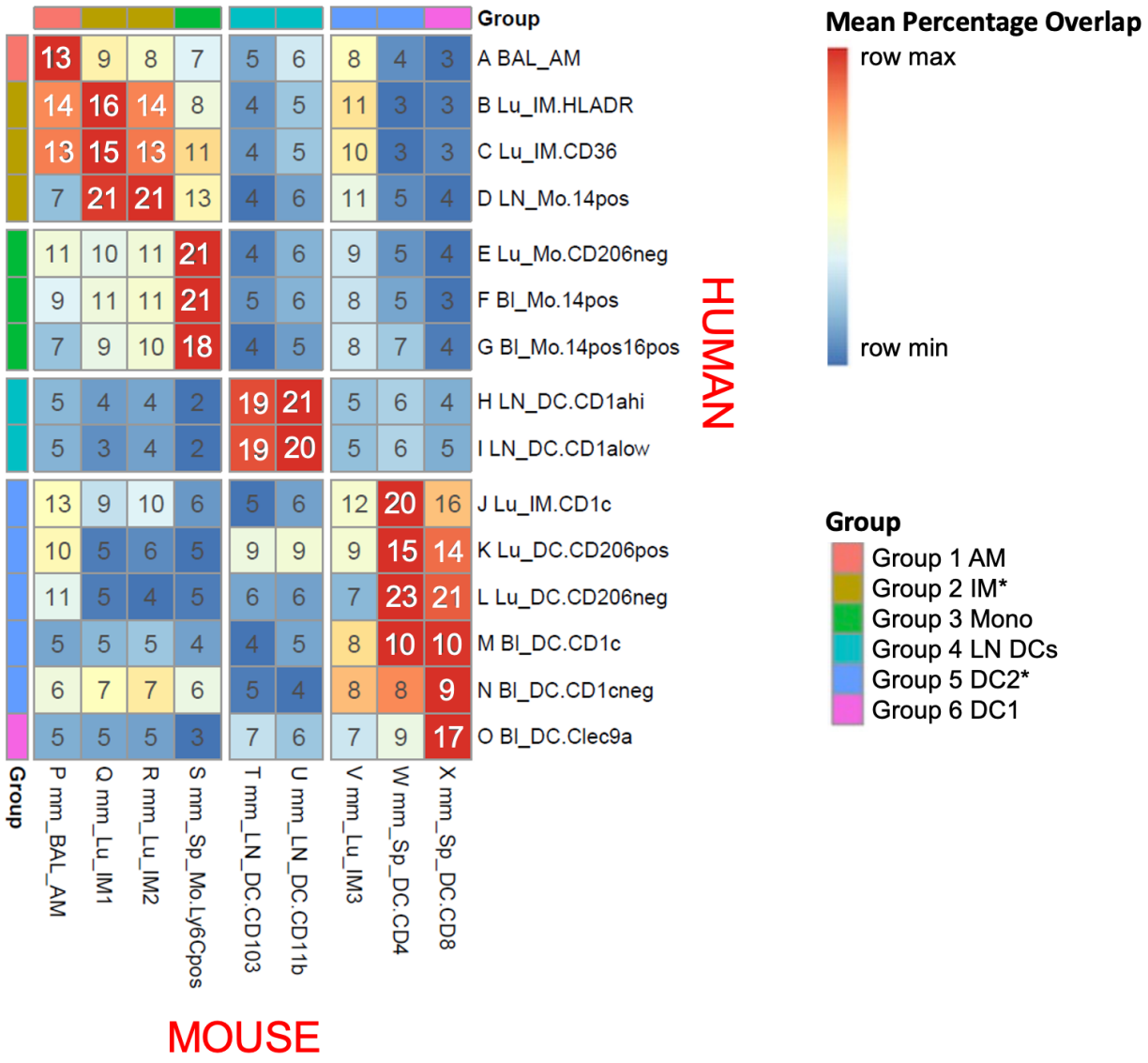

Figure 4B. Correspondence of MPs across species.

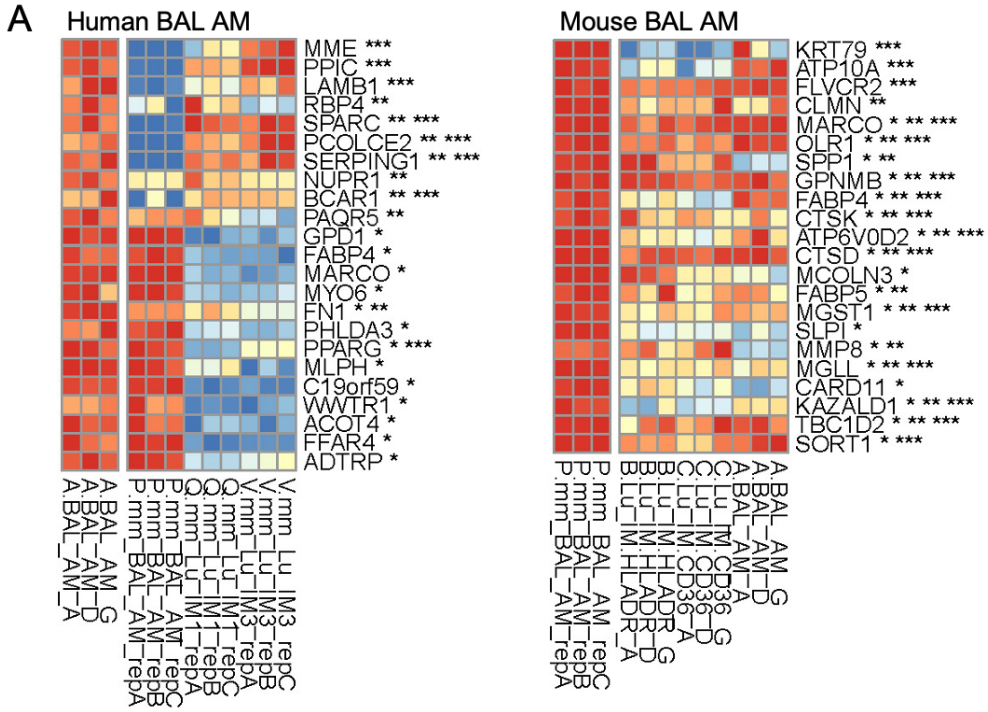

Figure 5A. Human cell types with a single highest mouse equivalent.

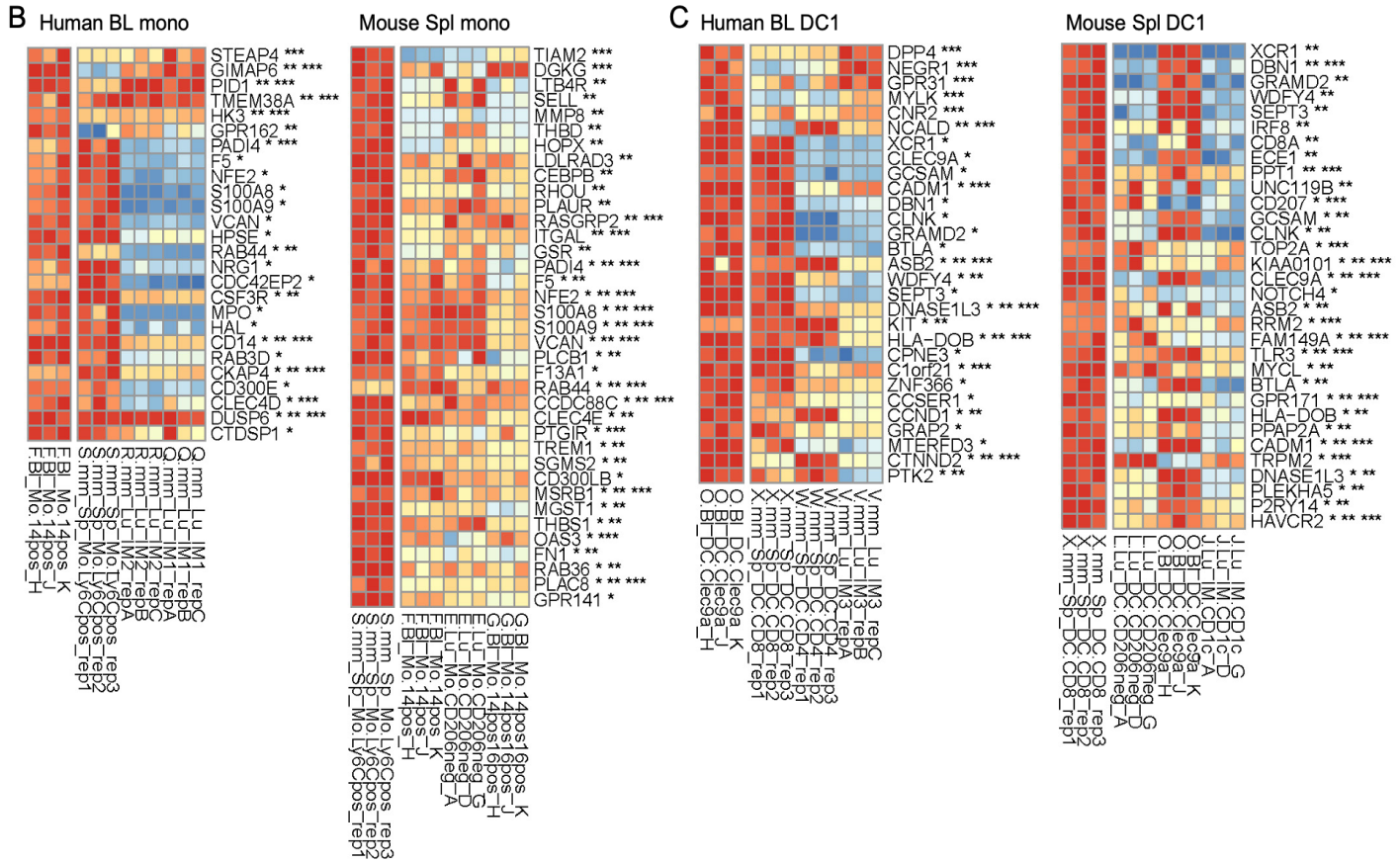

Figure 5B,C. Human cell types with a single highest mouse equivalent.

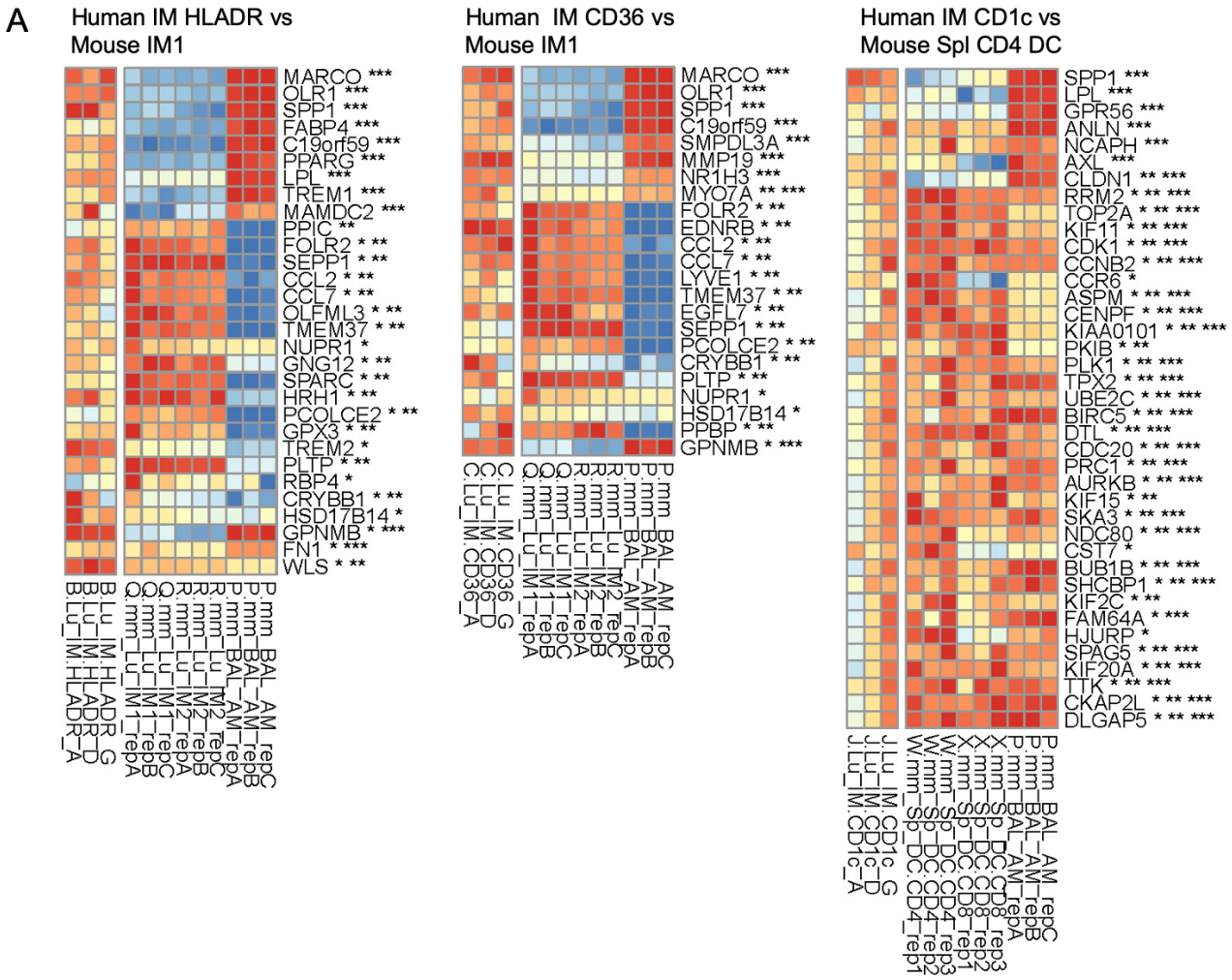

Figure 6A. Corresponding human and mouse IM subtypes.

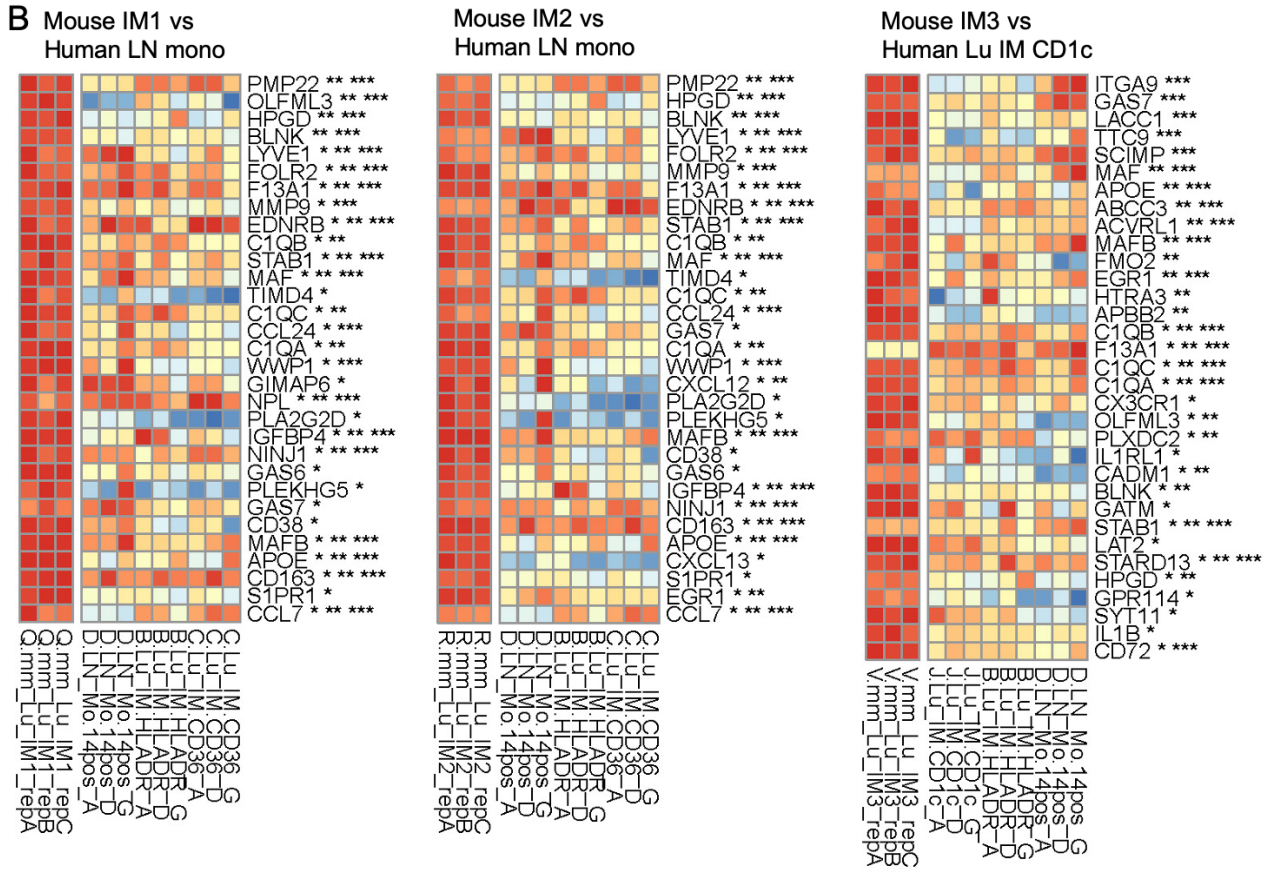

Figure 6B. Corresponding human and mouse IM subtypes.

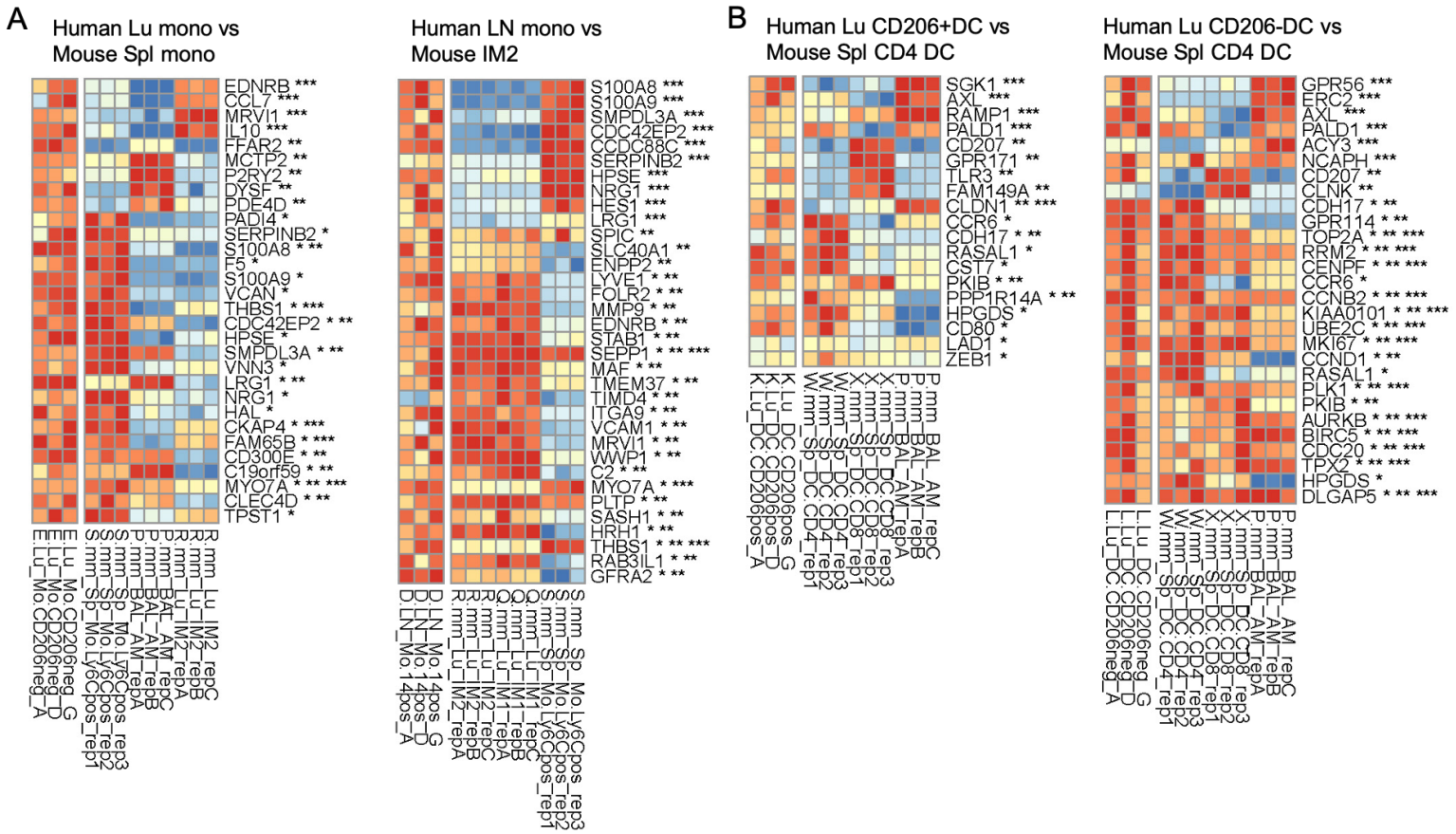

Figure 7A,B. Corresponding human and mouse monocytes and DC subtypes.

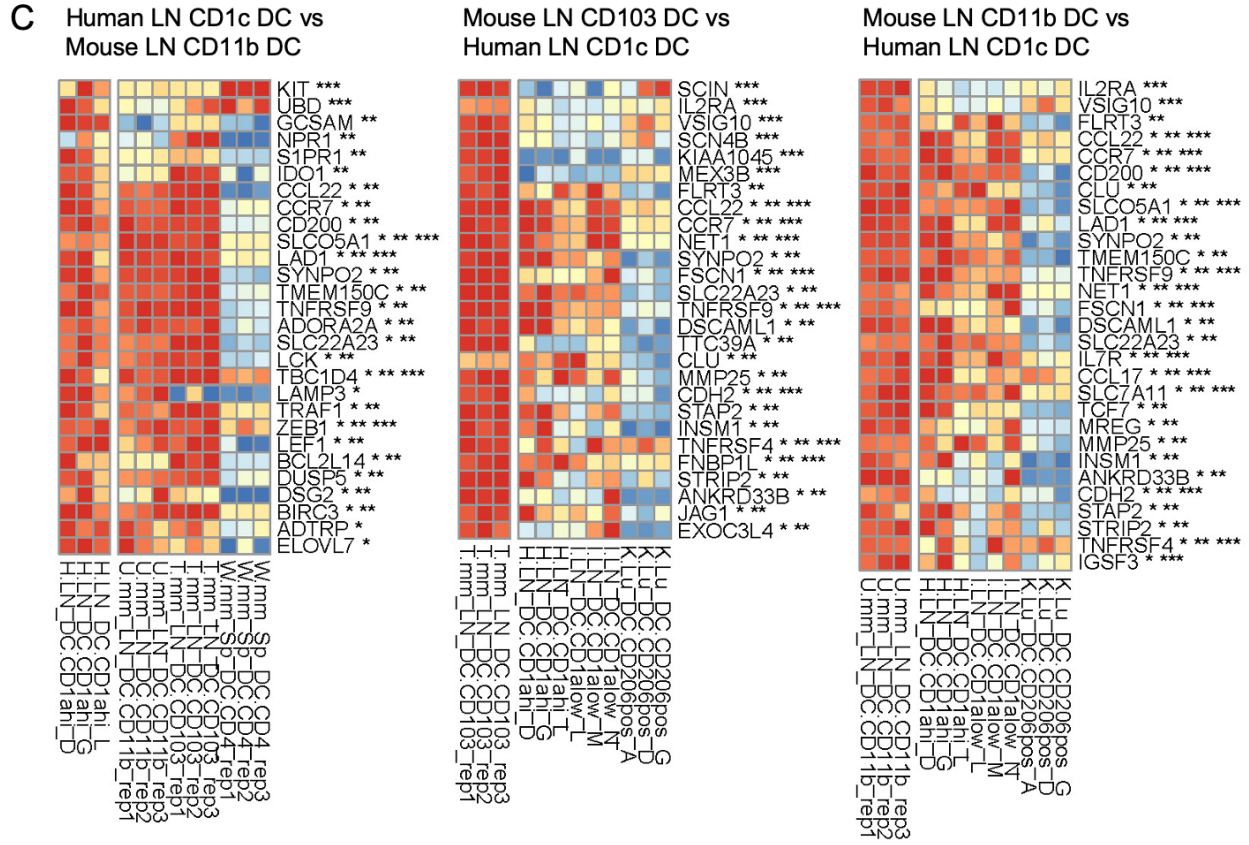

Figure 7C. Corresponding human and mouse monocytes and DC subtypes.

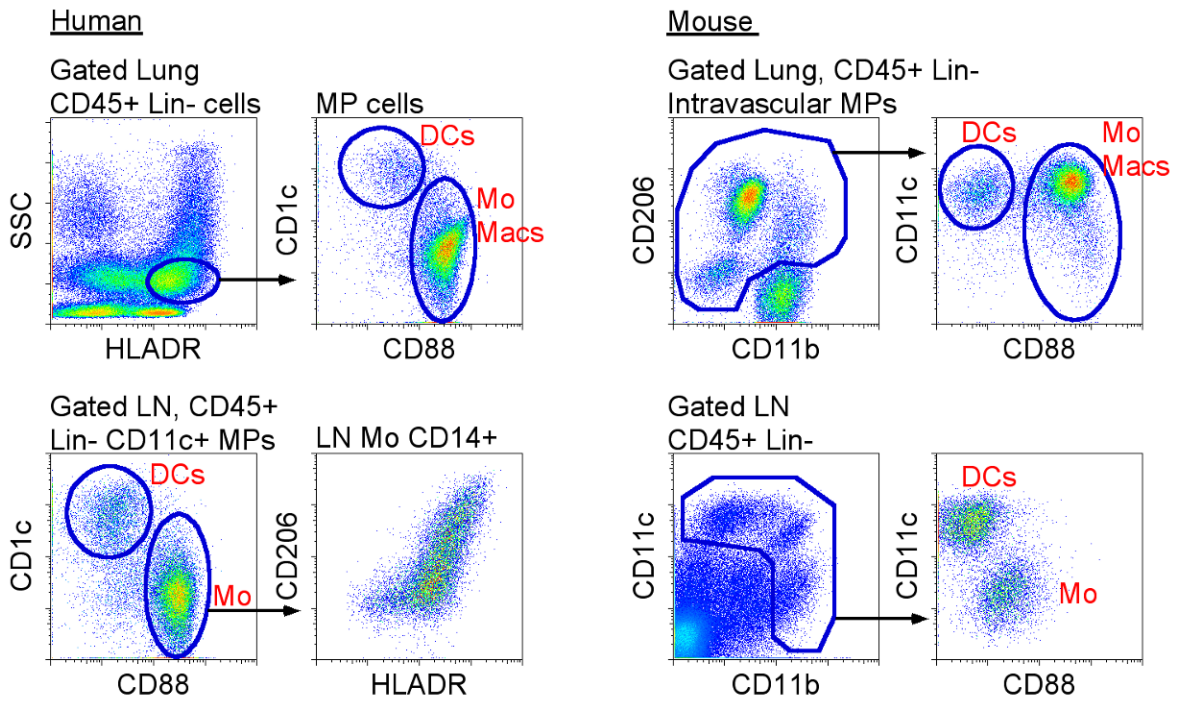

Figure 8. CD88 expression distinguishes monocytes and macrophages from DCs in both human and mouse.