Human and Mouse Transcriptome Profiling Identifies Cross-Species Homology in Pulmonary and Lymph Node Mononuclear Phagocytes

Sonia M. Leach[1,2,10], **Sophie L. Gibbings**[3,10], **Anita D. Tewari**[4], **Shaikh M. Atif**[5], **Brian Vestal**[1,2], **Thomas Danhorn**[1], **William J. Janssen**[6,7], **Tor D. Wager**[8], **Claudia V. Jakubzick**[3,4,9,11,*]

[1]Center for Genes, Environment, and Health, National Jewish Health, Denver, CO 80206, USA

[2]Department of Biomedical Research, National Jewish Health, Denver, CO 80206, USA

[3]Department of Pediatrics, National Jewish Health, Denver, CO 80206, USA

[4]Department of Microbiology and Immunology, Dartmouth College, Hanover, NH 03756, USA

[5]Department of Medicine, Division of Asthma, Allergy, and Clinical Immunology, University of Colorado, Denver, CO 80045, USA

[6]Department of Medicine, National Jewish Health, Denver, CO 80206, USA

[7]Division of Pulmonary Sciences and Critical Care, University of Colorado, Denver, CO 80045, USA

[8]Department of Psychology and Neuroscience, University of Colorado, Boulder, CO 80309, USA

[9]Department of Immunology, University of Colorado, Denver Anschutz Campus, Denver, CO 80045, USA

[10]These authors contributed equally

[11]Lead Contact

SUMMARY

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. Although murine MPs are well characterized, their pulmonary and lymph node (LN) human homologs remain poorly understood. To address this gap, we have created a gene expression

*Correspondence: claudia.jakubzick@dartmouth.edu.

compendium across 24 distinct human and murine lung and LN MPs, along with human blood and murine spleen MPs, to serve as validation datasets. In-depth RNA sequencing identifies corresponding human-mouse MP subtypes and determines marker genes shared and divergent across species. Unexpectedly, only 13%–23% of the top 1,000 marker genes(i.e., genes not shared across species-specific MP subtypes) overlap in corresponding human-mouse MP counterparts. Lastly, CD88 in both species helps distinguish monocytes/macrophages from DCs. Our cross-species expression compendium serves as a resource for future translational studies to investigate beforehand whether pursuing specific MP subtypes or genes will prove fruitful.

Graphical Abstract
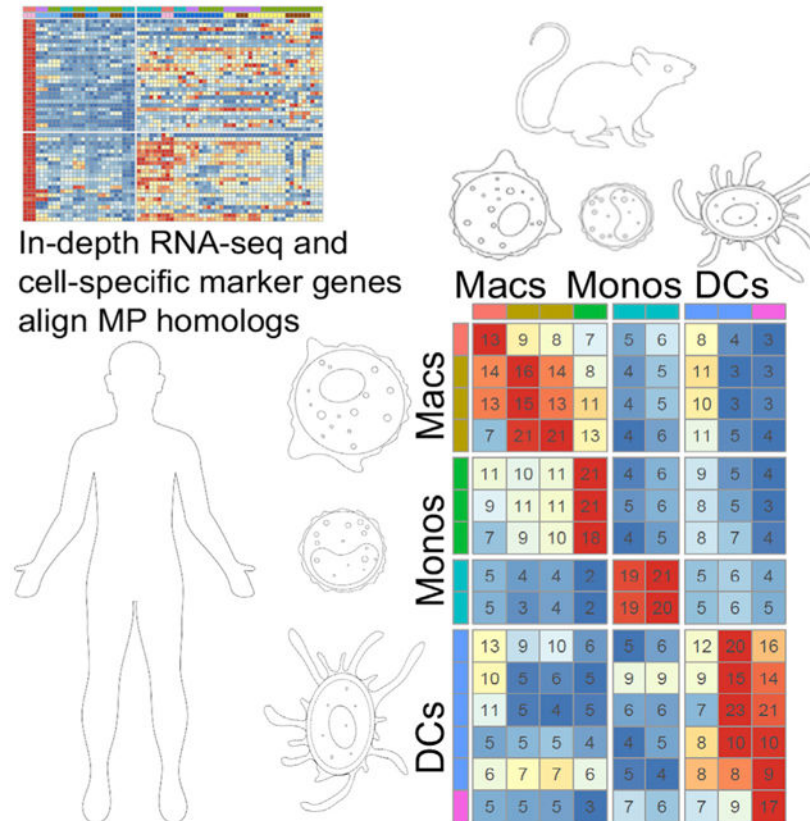
In Brief

Understanding the homology of mononuclear phagocytes between humans and mice is essential for translating murine studies into human medicine. Leach et al. use in-depth RNA sequencing to define highly discriminatory marker genes for 15 human and 9 mouse MPs that align cross-species homologs and define conserved and divergent marker genes.

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. For example,

both species have an immune system that contains mononuclear phagocytes (MPs), granulocytes, innate lymphoid cells, and lymphocytes (Masopust et al., 2017). 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 (Alt-mann, 2018); and the development of granulocyte-macrophage colony-stimulating factor (GM-CSF) therapy for pulmonary alveolar proteinosis (PAP) was influenced by observations in mice, in which the deficiency of GM-CSF results in PAP (Trapnell et al., 2009).

In spite of these encouraging examples, numerous failures in drug development across multiple diseases make it clear that ro dent-to-human translation remains an important challenge (Begley and Ellis, 2012). 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 naive assumptions about homology. Therefore, assessing the transcriptome alignment of cross-species human-mouse MPs is crucial.

If equivalent classes of human and murine MP subtypes and their transcriptional markers can be established, studying how murine MPs interact with one another to maintain homeostasis and resolve inflammation could help us to decipher the mechanisms that go awry in human diseases. Murine MPs have been extensively characterized and demonstrate a clear division of labor during innate and adaptive immunity (Guilliams et al., 2013; Desch et al., 2014; Gibbings et al., 2017; Atif et al., 2015, 2018b; Kim et al., 2014; McCubbrey et al., 2018; Tussiwand et al., 2015). Pulmonary murine MPs consist of alveolar macrophages (AMs), tissue trafficking monocytes, dendritic cell (DC) subtypes (Tussiwand et al., 2015; Murphy, 2013) and three interstitial macrophages (IMs). Two IMs, IM1 ($CD206^{hi}MHCII^{+}$) and IM2 ($CD206^{hi}MHCII^{lo}$), display classical macrophage characteristics. The third IM, IM3 ($CD206^{lo}MHCII^{+}$), displays macrophage properties, but has a higher rate of turnover and expresses pro-inflammatory, monocytic, and DC genes (Gibbings et al., 2017; Schyns et al., 2019; Chakarov et al., 2019). Little is known about the functional contribution of IMs to homeostasis, inflammation, or disease, partly due to difficulties in their identification (Ural et al., 2020; Lim et al., 2018). Improving our knowledge of the IM transcriptome is important since they appear to maintain their transcriptional signature across multiple organs, as do DCs (Gibbings et al., 2017; Tamoutounour et al., 2012,, 2013; Epelman et al., 2014; Plantinga et al., 2013).

Circulating monocytes were traditionally viewed as precursors to tissue-resident macrophages. However, we now know that monocytes continuously traffic through tissue, where they survey the environment, and unless there is a macrophage niche to fill, they do not differentiate into tissue resident macrophages in steady state (Guilliams and Scott, 2017; Bain et al., 2016; Jakubzick et al., 2013; Jakubzick et al., 2017). During inflammation, monocytes can differentiate into inflammatory and resolving macrophages, which display distinct properties from resident macrophages (Mould et al., 2017, 2019; McCubbrey et al., 2018; Misharin et al., 2017). Even though monocytes are abundant in the lymph nodes (LNs) of humans and mice, their role in adaptive immunity is less defined than DCs (Jakubzick et al., 2017; Larson et al., 2016; Bosteels et al., 2020).

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 (Vermaelen et al., 2001; Jakubzick et al., 2006, 2008; Desch et al., 2011; Atif et al., 2018b). In mice, there are two main DC types, $Irf8^+Batf3^+$ DC1 and $Irf4^+$ DC2. These DCs express distinct transcriptional factors that regulate their development, antigen acquisition, processing, and presentation capabilities. 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 comparisons have been performed for human blood (BL) to mouse spleen and human heart to mouse heart (Ingersoll et al., 2010; Bajpai et al., 2018; Bachem et al., 2010; Jongbloed et al., 2010). 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 (Yu et al., 2016; Yu and Tighe, 2018; Gibbings and Jakubzick, 2018a, 2018b; Desch et al., 2016; Mould et al., 2017; Reyfman et al., 2019; Bharat et al., 2016; Misharin et al., 2013), a transcriptional alignment of the human-murine pulmonary AMs, IMs, DCs, monocytes, and LN monocytes and DCs is still lacking.

Here, we created a gene expression compendium by in-depth RNA-seq to establish the non-diseased transcriptional profile of 15 distinct human and 9 distinct murine MPs, defined by our group and others (Desch et al., 2016; Gibbings and Jakubzick, 2018a; Yu et al., 2016; Yu and Tighe, 2018). The primary focus is pulmonary MPs, especially the understudied LN MPs. However, we also include human BL and murine spleen MPs as control tissues to validate the analytical quality of cross-species pulmonary MP analysis, as homology has been more extensively characterized. Interestingly, the overlapping percentage of the top 1,000 marker genes defined for the homologous pairs, human BL and murine splenic MPs, were similar to those observed for lung and LN MP cross-species counterpart pairs. Overall, we(1) developed a strategy to determine the cross-species counterpart for each MP subtype, (2) compared MP subtypes' global expression profiles, (3) contrasted specific sets of MP subtype marker genes, and (4) 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 4 tissue locations: bronchoalveolar space, lung interstitium and draining LNs from non-diseased donors, and BL from healthy donors. Ten different MP populations were sorted from each lung donor, and 5 different populations were sorted from each BL donor (Figures 1A and S1). Putative MP subtype labels were assigned based on (1) analyses of >100 human lungs examined in our laboratory for given MP populations consistently observed regardless of age and gender and (2) well-established cell surface markers considered to selectively identify macrophages or DCs (Figures 1A, 1C, and S1) (Gibbings and Jakubzick, 2018a, 2018b; Desch et al., 2016; Bharat et al., 2016;

Misharin et al., 2013; Yu et al., 2016; Yu and Tighe, 2018; Guilliams et al., 2016). For each human lung, we isolated 7 MP populations: 4 distinct macrophage populations (AM, IM CD36, IM HLADR, and IM CD1c[+]), 2 CD1c[+]CD36[−] DC subsets (DC CD206[−] and DC CD206[+]), and CD14[+] monocytes. Peribronchial LNs contributed CD14[+] LN monocytes and 2 LN DC populations (DC CD1a[hi] and DC CD1a[lo]). Lastly, we sorted well-defined BL MPs including 2 monocyte populations (Mo CD14[+]CD16[−] and Mo CD14[+]CD16[+]) and 3 classical CD11c[+]HLADR[+] DCs (DC Clec9a[+], DC CD1c[+], and DC CD1c[−]) (Desch et al., 2016; Ingersoll et al., 2010; Huysamen et al., 2008). In total, we sorted and analyzed 66 human MP samples by RNA-seq, with 3–6 replicates per MP subtype.

In the mouse, we isolated 9 distinct murine MPs from 4 tissue locations: bronchoalveolar space, lung tissue and draining LNs, and spleen (Figures 1B and S2). We isolated 4 lung macrophage populations (AM, IM1, IM2, and IM3) and 2 LN migratory DCs (CD103[+] DC1 and CD11b[+] DC2), distinguished from lymphoid resident DCs by high major histocompatibility complex class II (MHC class II) expression (Figure 1B). In addition, 2 splenic DC populations (CD8[+] DC1 and 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 BL DCs (Haniffa et al., 2012; Bachem et al., 2010; Schlitzer et al., 2013; Reynolds and Haniffa, 2015; Ingersoll et al., 2010). 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 3 replicates per MP subtype.

Confirming Human MP Subtypes Based on Defined Transcriptional Profiles

To validate our designations of sorted human MPs as either macrophages, monocytes, or DCs, we examined the expression of classical macrophage/monocyte and DC genes (Figure 2) (Scott et al., 2014; Wu et al., 2016; Satpathy et al., 2012). As anticipated, most of the macrophage/monocyte genes such as *MAFB*, Fcg receptors, lysosomal proteases, and associated membrane proteins were most strongly expressed in human MP subtypes putatively labeled as macrophages or monocytes (Figure 2) (Tussiwand et al., 2012; Satpathy et al., 2012; Wu et al., 2016). Conversely, classical DC genes such as *HLA-DR*, *CD1C*, *CCR7*, *CD40*, *CD80*, *LAMP3*, *IRF4*, *IRF8*, *FLT3*, and *ZBTB46* were strongly expressed in human pulmonary and LN MPs, putatively labeled as DCs (Figure 2) (Wu et al., 2016; Satpa thy et al., 2012). In fact, pulmonary and LN DCs showed a higher relative expression of classical DC genes than did BL DCs. For example, Figure 2 shows a high expression of *CCR7*, a key DC signature gene, in pulmonary and LN DCs. DCs require CCR7 for both migration from tissue to LNs via afferent lymphatics and T cell zone localization (Förster et al., 1999). Overall, 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, Monocyte, or DC Lineage

A cross-species dataset of RNA-seq reads was obtained for all autosomal genes with a 1:1 homology between human and mouse. Sex chromosome genes were omitted to avoid variation based on donor sex. 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 RNA-seq (scRNA-seq) to account for species-specific effects on gene expression (Butler et al., 2018). 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 (Figures S3 and S4) (Bolstad et al., 2003). The unbiased analysis used 885 genes shared between human and mouse from among the top 2,000 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 cell populations with similar profiles included a mix of human and mouse cells and multiple tissues of origin, indicating that species and tissue type were not the main sources of variation in gene expression (Figure 3A). Crucially, the transcriptional profiles were instead clustered by MP cell type, with AMs, monocytes, and DCs forming tight clusters, followed by IMs (Figure 3A). To facilitate comparisons, samples were further grouped into six groups, labeled with an indication of their lineage (Figure 3B). Group 1 contains AMs from BAL. The IM group (group 2) separates from AMs, as previously observed in mice (Gibbings et al., 2017). Group 3 (monocyte [Mono]) is a pure cluster of monocytes across species, sampled from lung, BL, and spleen. Despite the fact that AM and IM share the expression of a number of macrophage core signature genes, their overall gene expression is strongly influenced by their local environment, alveolus versus interstitium (Gautier et al., 2012; Gibbings et al., 2017). Group 2 also includes CD14$^+$ monocytes from the LN. This likely reflects their shared origin as monocyte-derived populations that matured in the lung environment.

The DCs are distributed across three distinct groups: group 4 LN DCs, group 5 DC2, and group 6 DC1 (Figure 3B). LN DCs in group 4 form a distinct cluster, clearly separated from the other DCs, suggesting a maturation or functional program distinct from BL or splenic DCs. Group 5 DC2 contains the human lung/blood DCs and murine splenic CD4$^+$ DCs, suggesting an overall DC2 (inter-feron regulatory factor 4 [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 that this IM subtype in humans and mice also exhibit some DC programming. This draws a parallel to the Langerhans cell in skin, which has both macrophage and DC-like functions and expression of macrophage and DC-specific transcription factors, *Mafb* and *Zbtb46* (Wu et al., 2016). Lastly, group 6 DC1 contains human BL Clec9a$^+$ DCs and murine splenic CD8$^+$ DCs, the expected human-to-mouse counterparts of the Irf8, Batf3-dependent DC1 lineage (Hildner et al., 2008). Overall, the unbiased computational analysis of global gene expression data demonstrates that across species, the majority of MP subtypes group by cell type—AM, IM, Mono, or DC lineage—across species.

Marker Gene Analysis Identifies the Closest Cross-Species Homology per MP Subtype

We develop a novel, robust strategy to determine the maximally corresponding MP subtype between human and mouse. Other methods to establish homologous MP populations have

relied on global correlation (Travaglini et al., 2020) or marker gene set enrichment (Chakarov et al., 2019), which can be affected by the set of genes used in the calculations. For instance, genes with low expression levels and housekeeping genes can bias global correlations so that cell types appear similar, failing to discriminate (Figure S5A). Strong correlations across multiple cell types can occur even when using only highly variable genes and may disagree with the global correlations (Figure S5B). Marker gene sets often overlap for closely related cell types, especially when not accounting for the level of uniqueness of expression for that cell type. Thus, enrichment analysis based on intersections of unranked marker gene lists can obscure the dominant functional homolog (Figure S6A). If the closest homolog is chosen as the cell type with the largest overlapping marker set, the closest homolog using the top 50 marker genes may be different from the closest homolog using the top 1,000 (Figure S6B). Ideally, the best homolog would be the cell type that continues to have the largest overlap, even as the size of the (ordered) marker set increases. Our approach captures this intuition to determine the maximally corresponding MP subtype between species.

Our approach prioritizes potential marker genes for each MP subtype within each species. Marker genes should be highly and preferentially expressed within a single subtype over all other subtypes within a species. In total, 15 marker gene sets for human and 9 marker gene sets for mouse are evaluated from a dataset consisting of 11,959 genes across 93 samples. Each gene is ranked for each subtype by a score that multiplies two factors: (1) the level of expression in a given subtype relative to the median expression over all subtypes and (2) how differentially expressed the gene is in that subtype over all other subtypes (Rosenthal, 1978).

To achieve a high score, marker genes for each of the 15 human and 9 murine subtypes must be differentially expressed with respect to each remaining subtype within their species, in contrast to other marker gene identification methods that consider the subtype versus the union of all of the samples from the other subtypes. Note that 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 the multiple related subtypes, and thus receive a low score for differential expression. In this way, we take advantage of the granularity at which the subtypes were isolated to find the maximally corresponding homolog based on marker genes essentially unique to that subtype.

Once marker genes are ranked, we can examine the degree of overlap of marker sets between species to determine the closest cross-species match. Our method takes into account the level of uniqueness of expression for that cell type and prioritizes cell types with strong overlap regardless of the size of the marker set considered. We develop a method based on correspondence-at-the-top (CAT) plots, previously used to compare expression results across experimental platforms (Irizarry et al., 2005). CAT plots visualize the intersection percentage of two ranked lists as the list size progressively increases (Figure 4A). Using an MP subtype in one species as the reference, each MP subtype in the opposing species is compared to the reference to determine the percentage of overlap as the marker gene set size increases—in other words, comparing the top 5 genes, the top 10 genes, and so forth (examples illustrated in Figure 4A; all human and mouse MPs in Figures S7 and S8). As

seen in the CAT plots, generally the curves stabilize and one cell type dominates as the best cross-species match within the top 1,000 genes (vertical dashed line).

One option for choosing the closest match between species would be the cell type with the maximum match at the top 1,000 genes. However, marker sets are ordered by degree of unique expression for the cell type, so maximizing at any one list size would not distinguish between a pair that agrees quickly in the ranking and maintains agreement up to that list size (i.e., strong, robust agreement among the most unique marker genes) from a pair that achieves high agreement only later in the ranking (i.e., includes mainly marker genes appearing in multiple cell types or more lowly expressed genes). To be robust to marker set size and determine which subtype quickly and consistently dominates, we instead chose 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 1,000, referred to henceforth as mean CAT overlap (Figure 4B).

Overall, the mean CAT overlap ranges between 2% and 23%. The closest cross-species homolog for a given MP subtype is then chosen as that subtype in the opposing species with the highest mean CAT overlap, which ranges from 13% to 23%. A potential factor contributing to the low percentages is the fact that our ranked marker genes are selected to be those expressed exclusively in a single cell type. Core signature genes appearing in multiple cell types are discarded. Including these would substantially increase the overall percentage overlap but would be less meaningful because these genes are not diagnostic of one cell type. Of note, the clustering of MP subtypes seen in the previous t-SNE plot analysis of global expression is echoed in the CAT plot analysis of marker gene sets (Figures 3B and 4B). This suggests that despite the low overlap by our strict marker gene definition, the same biological correspondences are found as in the unbiased analysis.

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. AMs correspond across species, with a 13% mean CAT overlap (Figures 4A and 4B, group 1 AM). Human BL and lung monocytes correspond directly to murine splenic monocytes with 21% mean CAT overlap (group 3 Mono). Human BL Clec9a$^+$ DCs are most similar to murine splenic CD8$^+$ DCs, with 17% mean CAT overlap (group 6 DC1); unexpectedly, these Clec9a$^+$ DCs are not strongly aligned with murine LN CD103$^+$ DCs (7% mean CAT overlap), as would be hypothesized based upon the shared Irf8, 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 may have a greater effect on gene expression.

There are a few human MPs that share homology with more than one mouse MP subtype, however. Human IM HLADR and IM CD36 are the most similar to murine IM1 and IM2, with 13%–16% mean CAT overlap (Figure 4B, Group-2 IM, and Figure S7), demonstrating correspondence across IMs, as expected. However, human IM CD1c is most similar to murine splenic CD4$^+$ DC (20% mean CAT overlap) and less similar to its expected counterpart murine IM3 (12% mean CAT overlap) (Figure 4B, group 5 DC2, and Figure S7).

Note, however, inversely the murine IM3 is most similar to human IM CD1c, as anticipated (Figure 4B and Figure S8).

Human LN monocytes correspond highly with murine IM1 and IM2 (21% mean CAT overlap) instead of splenic monocytes (13% CAT mean overlap, 3rd highest) (Figures 4A and 4B, group 2 IM). This was also evident in the t-SNE plots (Figure 3B), reflecting the shared characteristics of differentiated monocytes in tissue.

Murine MPs Match Expected Human MPs, Yet Exhibit Evidence of Dual Expression Programs

We also assessed the inverse correspondence, taking each mouse MP and identifying the most similar human MP (Figure 4 and Figure S8). Murine AMs (Figure 4B, group 1 AM) are highly similar to human AMs (13% mean CAT overlap), especially when comparing the top 200 genes, thus sharing the most unique AM marker genes (Figure 4A). However, among the top 200–1,000 genes, human IM CD1c and IM HLADR (group 2 IM) begin to overlap increasingly (13%–14% mean CAT overlap), suggesting that murine AMs share dual characteristics with human AMs and IMs. Murine LN DCs and splenic monocytes were most similar to human LN DCs and lung/BL monocytes, respectively (group 4 LN DCs ~20% mean CAT overlap and group 3 mono 21% mean CAT overlap), as anticipated.

Interestingly, where human BL Clec9a$^+$ DCs align most strongly with their expected counterpart mouse splenic CD8$^+$ DCs, the inverse is slightly different, depending on the number of top genes analyzed (Figure 4, group 6 DC1). Averaging the percentage of overlap over lists of size 1–1,000, murine CD8$^+$ DCs correlate best with the human lung CD206$^-$ DCs (group 5 DC2, 21% mean CAT overlap) as opposed to Clec9a$^+$ DCs (17% mean CAT overlap). Similar to the results for murine AMs, murine splenic CD8$^+$ DCs do match their expected homolog, Clec9a$^+$ DCs, best when considering only the first top 1–50 ranked genes (Figure 4A). This suggests that the transcriptional program of mouse splenic DCs shares similarities with both human BL Clec9a DCs and lung CD206 DCs.

Murine IM1 and IM2 correspond best to human LN monocytes, with 21% mean CAT overlap, followed by human HLADR and CD36 IMs, with ~15% mean CAT overlap (Figure 4B, group 2 IM, and Figure S8). Murine IM3 corresponds highly with human LN monocytes as well as IM HLADR (11% mean CAT overlap) (Figure 4B, group 2 IM, and Figure S8), although IM3 matches most strongly with human IM CD1c (12% mean CAT overlap) (Figure 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 (Figure 4B). In cases in which the predicted counterparts do not match expectations, the analysis gave evidence that the marker genes of the expected homolog did appear early in the ranking(i.e., among the top 50 marker genes), suggesting that multiple MP expression programs exist in some subtypes when patterns across a larger set of top marker genes (i.e., 1,000) are analyzed.

Conserved and Distinct Genes Expressed in Well-Defined Human-Mouse Homologs

Expression data for the top marker gene sets are first examined in three well-defined human-mouse homologs: AMs, monocytes, and DC1 (Figure 5). The functional properties of genes mentioned below are outlined in Table S1. Expression of the top 50 marker genes in the reference MP subtype is shown in the opposing species (Figures 5, 6, and 7). Note that the best match suggested visually among expression of just the top 50 genes may differ from the assigned best match maximizing mean CAT overlap computed from the top 1,000 genes, thus motivating the decision to highlight the top 3 closest subtypes of the opposing species.

As anticipated, signature genes that define these well-defined cell types are present, such as *FABP4*, *MARCO*, and *PPARG* for AMs (Figure 5B, group 1 AM); *S100A8*, *S100A9*, and *CD14* for monocytes (Figure 5C, group 3 Mono); and *XCR1*, *CADM1*, and *CLEC9A*, along with a recently identified gene, *WDFY4*, required for cross-presentation for human BL Clec9a[+] DCs and murine splenic CD8[+] DCs (Figure 5D, group 6 DC1) (Theisen et al., 2018). Although we outline genes known to be shared across human-murine MPs, it is important to note that many genes are not shared. For example, differences in gene expression for human-mouse AMs include *SERPING1*, *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 (Figure 5B). Human and mouse DC1, blood Clec9a[+], and splenic CD8[+] DCs differentially express *IDO1* or *IDO*, *CD207* (langerin), and *CCR7* (Figure 5D).

Conserved and Distinct Genes Expressed in Less-Well-Defined Human-Mouse MPs

Since this method of analysis supports previously identified human-mouse MP homologs, we next assessed marker genes in less-well-defined homologs: pulmonary IMs, LN monocytes, and pulmonary/LN DCs. Human IM CD36 and IM HLADR are best matched with murine IM1, closely followed by murine IM2 (group 2 IM; Figure 4B). The two murine IMs have a classical IM signature and share gene expression for *FOLR2*, *SEPP1*, *LYVE1*, *APOE*, *C1Q* (not shown), *CCL7*, and *CCL2* (Figure 6A). An example gene not shared is *SPP1*, which is highly expressed in human IMs (and murine AMs), but not in murine IMs (Figure 6A).

Inversely, murine IM1 and IM2 match best with human LN monocytes, followed by human IM HLADR and IM CD36 (Figure 6B). Unlike human blood and lung monocytes that correspond with murine splenic monocytes (group 3 Mono), human LN monocytes reciprocally align best with murine IM1 and IM2 (group 2 IM), followed by splenic monocytes (Figure 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* (Figure 7A). Interestingly, murine IM3 matches best with human IM CD1c, where both MPs express inflammatory monocyte genes *IL1B*, *C1QB*, *C1QC*, and *F13A1* (Figure 6B, group 5 DC2). However, human IM CD1c matches best with murine splenic CD4[+] DCs (Figure 6A), based on their shared non-classical DC genes (i.e., cytoskeletal genes, *CCR6*, and cell-cycle genes *CD1K1* and *CDC20*) (Figure 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 (Figure 7B and Figures 3 and 7C, group 5 DC2), expressing genes such as *CCR6*, *CD80*, *CD207*, and *AXL*. Consequently, human LN DCs align best with murine LN DCs (Figures 3 and 7C, group 4 LN DCs) and express classical DC genes such as *ZEB1*, *CCL22*, *CCL17*, and *CCR7*, but not the human classical DC gene, *LAMP3* (Figure 7C). Overall, these in-depth correspondences allow one to investigate genes shared and divergent in both well-defined and less well-defined homologous pairs.

A Cell Surface Marker That Allows a Clear Distinction between Macrophages/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 (EMR1) Ly6C, MerTK, and CD64 has helped distinguish DCs from macrophage and monocytes (Atif et al., 2018a; Yu et al., 2016; Misharin et al., 2013; Gautier et al., 2012). However, human monocytes do not have a Ly6C homolog; monocytes and macrophages express low levels of EMR1 and show variable expression of MerTK (perhaps due to the lack of a reliable antibody).

Analysis of our RNA-seq data identifies CD88 as a viable cell surface marker that can distinguish macrophages and monocytes from DCs in the tissue of humans and mice (Dutertre et al., 2019; Nakano et al., 2015). To validate this cross-species marker, histogram plots demonstrated that human and mouse AMs, IMs, and tissue monocytes express higher levels of CD88 compared to DCs (Figure 7D). All in all, CD88 is a commercially available cell surface marker that helps distinguishes MPs in humans and mice (Desch et al., 2016; Gautier et al., 2012).

DISCUSSION

The aim of this study was to investigate the homology between human and murine mononuclear phagocytes to aid in the translation of animal studies to human clinical uses. The dataset and analyses we provide are intended to improve the design of animal models mimicking human disease. Toward 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. Overall, this dataset allows 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.

As validation for our computational analysis, we analyze cell types derived from our human BL and splenic MPs, and compare the populations from several recent scRNA-seq studies profiling human BL MPs (Villani et al., 2017; Dutertre et al., 2019), splenic MPs (Brown et al., 2019), or lung MPs (Travaglini et al., 2020) (Figure S9). Examining our expression data for the top 20 marker genes defined per cluster by the original publications, human blood Clec9a[+] DCs consistently correspond to conventional DC1 populations and blood DC CD1c[−] cells correspond to plasmacytoid (pDC) populations. The monocyte subtypes express

genes identified in monocyte clusters, while AMs and IMs express the top macrophage cluster genes. In general, our bulk RNA-seq data of purified cell populations aligns favorably with the assumed cluster labels of scRNA-seq data, with the added advantage of robust measurement of thousands of genes across more MP cell types than previously published, and especially in the understudied LN MPs.

The validity and utility of our analysis for the identification of human-mouse MP counterparts is demonstrated by its unbiased prediction of known homologs. As expected, cross-species groupings include (1) human and murine AMs (group 1), (2) human BL and murine splenic monocytes (group 3), and (3) human BL Clec9a$^+$ DCs and murine splenic CD8$^+$ DCs (group 6). These pairs align best with each other when compared with all of the other MPs across species (Ingersoll et al., 2010; Caminschi et al., 2008). These three pairs provide a useful reference for assessing the confidence with which our strategy can identify less-well-defined MP subtype homologs. In addition, high degrees of overlap between human BL and lung monocytes with mouse splenic monocytes is reassuring, as it demonstrates a negligible effect on the transcriptome of cells extracted from BL versus digested tissue.

We also observe a preserved general distinction between MP lineages in both human and mouse. For example, groups 1, 2, and 3 contain exclusively macrophages (AMs and IMs) and monocytes, while groups 4, 5, and 6 contain DCs (Figures 3 and 4B), with 2 exceptions. The first exception is human LN monocytes in the IM-enriched group 2, which as previously stated represents a more mature monocyte expressing both 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; however, conversely, mouse IM3 aligns best with lung IM CD1c$^+$ (Figure 4B).

Interestingly, human BL DCs and mouse splenic DCs cluster away from human and mouse LN DCs, perhaps due to tissue imprinting. One distinction we notice is the expression of CCR7, which is a hallmark feature of DCs. Human BL DCs lack CCR7, a receptor required for lymphatic migration and localization in the T cell zone of LNs (Randolph et al., 2005). This is possibly because the cells are immature and require additional endothelial-tissue interaction to upregulate CCR7. Another observation is that human DC subtypes differentially express CD62L, a transmigration molecule used by circulating monocytes and preDCs to enter tissue and LNs (Liu et al., 2009; Jakubzick et al., 2013; León and Ardavín, 2008). CD62L is highly expressed in human BL DCs, but not in tissue DCs, suggesting that once in tissue, this gene is no longer required. Thus, one possibility could be that human BL DCs may give rise to human tissue DCs. An alternative possibility is that human BL DCs and tissue DCs share the same precursor cell, which gives rise to DCs in different compartments. There is a precedent for this in mice: murine lymphoid resident DCs do not give rise to migratory tissue DCs but share the same precursor cell, where the former DCs precursor entered the LN via the high endothelial venule (HEV) and the latter DC migrated from the tissue into its draining LN via afferent lymphatics.

One surprising finding is the relatively small effect of DC lineage on gene expression patterns. In mice, there is a strong distinction between DC1 (Irf8, Batf3 dependent) and DC2

(Irf4 dependent) lineages, as seen for CD8$^+$ (DC1) versus CD4$^+$ (DC2) splenic DCs. The existence of two DC lineages in humans is supported by the alignment of BL Clec9a$^+$ DC with murine CD8$^+$ splenic DC (group 6), compared with CD4$^+$ splenic DCs. However, the separation of human lung or LN DCs based on DC lineage appears to be masked by the stronger effects of tissue imprinting, as evidenced 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 consistently observed CD1a^{hi} and CD1a^{lo} CD1c$^+$ lung-draining LN DCs in >50 individuals, making the assumption that they may be distinct cell types. However, it is now clear that although the sorted human LN DCs are distinct at the CD1a protein level, these populations are not transcriptionally distinct. Further investigation as to how to subset DEC205$^+$CD1c$^+$ lung-draining LN DCs using different cell surface markers is needed. More important, human LN DCs are understudied. Even though we are unable to decipher more than one LN DC cluster, there is a wealth of genetic information provided in this dataset for future follow-up. For instance, CCL17 and CCL22 expression is restricted to DC populations in lung and LNs, and is not expressed in human BL DCs (Alferink et al., 2003; Stutte et al., 2010; Rapp et al., 2019; Vulcano et al., 2001). This is also true in mice, in which these chemokines distinguish DCs from other immune cell types and are important for T cell recruitment and interactions (Rapp et al., 2019).

Overall, the overlap among the top 1,000 genes for corresponding human and mouse MPs of the same type is low (13%–28%; Figure S6B). This points to the need to identify species-specific marker genes and not assume that a gene identified in one species will be an adequate marker in another species. This applies to both translational (mouse to human) and reverse translational (human to mouse) studies. However, it is important to note that this does not mean that the species have completely different transcriptional profiles overall. To qualify as a marker gene for a given MP subtype, a gene had to be expressed primarily in just that one single subtype. Genes expressed in common across multiple closely related subtypes were not included in the overlap analyses. An alternative approach would be to detect the broader differences between monocytes/macrophages versus DC lineages overall, including all genes. Broader inclusion would likely yield a much higher percentage overlap across all MP subtypes, and estimating this overlap may be useful for some applications. However, it would require a different analysis and is beyond the scope of the present study, as our goal was to focus on the subset of genes likely to be useful as markers for particular MP subtypes.

Finally, 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 (Desch et al., 2016; Nakano et al., 2015; Gautier et al., 2012; Jakubzick et al., 2013). 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.

STAR★METHODS

RESOURCE AVAILABILITY

Lead Contact—Further information and requests of resources and reagents should be directed to and will be fulfilled by Dr. Claudia Jakubzick (claudia.jakubzick@dartmouth.edu).

Materials Availability—This study did not generate new unique reagents.

Data and Code Availability—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). To protect donor anonymity human RNaseq data have not been deposited in a public repository. Requests for access to human sequencing data may be made to the lead contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects—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 PaO2/FiO2 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 were female and 4 were male. 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). Additional demographic information is shown in Table S2. 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. Note, not every donor provided all 15 MP cell types. 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.

Animals—Naïve female C57BL/6 mice were obtained from Charles River Laboratories. Mice were group housed in a specific pathogen-free environment at National Jewish Health (Denver, CO), and used at 6–10 weeks of age, in accordance with protocols approved by the Institutional Animal Care and Use Committee. Each sample represents pooled cells from 5 mice; different cohorts of mice were used to isolate lungs, LLNs and spleens.

METHOD DETAILS

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 additionally described in Desch et al., 2016 and Gibbings and Jakubzick, 2018a.

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 (Worthington Biochem) in PBS containing Ca^{2+}, Mg2+, 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, first with scissors and second with a food processor in buffer containing fetal bovine serum. The resulting tissue homogenate was filtered through a 100 mm nylon filter membrane to create a single-cell suspension. Density gradient centrifugation using 30% and 65% Percoll (Sigma Aldrich) 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 100 mm nylon filter membrane to create a single cell suspension.

Peripheral blood mononuclear cell (PBMC) isolation, blood was obtained by venipuncture from healthy adults as per IRB approved protocol. CPT tubes (BD) were used to separate PBMC from other blood components according to manufacturer's instructions.

Murine Lung, LLN, and spleen MP isolation—Mice were euthanized by CO_2 asphyxiation and lungs were perfused with cold PBS via the heart. Lungs were removed, minced finely with scissors and subjected to enzymatic digestion with 0.5mg/mL Liberase (Roche) for 30 minutes at 37°C as reported previously (Gibbings et al., 2017). Mediastinal lymph nodes or spleens were removed, minced, and digested by incubation with 2.5mg/mL Collagenase D for 30 minutes at 37°C. Single cell suspensions were prepared by manual homogenization and filtration through a 100 mm nylon filter membrane. Mouse lung macrophages were labeled with anti-mouse MerTK and enriched with anti-biotin microbeads as previously reported. Mouse lymph node and spleen DCs were enriched with anti-mouse CD11c microbeads and spleen monocytes were enriched using anti-mouse CD11b microbeads according to the manufacturer's protocol (Miltenyi Biotec).

Mononuclear phagocyte enrichment—Cells of interest were positively enriched prior to FACS sorting using Miltenyi microbeads according to manufacturer's protocols. Human 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 human lung and lymph node cells was achieved using PE conjugated anti-human CD1c and biotin-conjugated anti-human CD11c antibodies in combination with anti-PE and anti-biotin microbeads.

FACS sorting and flow cytometry—Human 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 the key resources table. Dead cells were excluded by uptake of 4',6-diamidino-2-phenylindole (DAPI, Invitrogen/Thermo Fisher Scientific). Cells were either sorted using a FACS Aria Fusion (BD) as outlined in Figure 1 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 (Illumina San Diego, CA, USA). A Clontech SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing (Mountain View, CA, USA) and Nextera XT (Illumina 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 1×50bp 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 adaptor 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). 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 (Dobin et al., 2013). Reads mapping uniquely to each gene of the Ensembl 75 annotation 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 (Liao et al., 2019).

QUANTIFICATION AND STATISTICAL ANALYSIS

Cross-species RNA-seq data—A cross-species dataset of RNA-seq reads was obtained for all autosomal genes with a 1:1 homology between human and mouse, as determined Ensembl BioMart version 75 (*Homo sapiens* genes, Homologs subsection, Mouse Orthologs). Sex chromosome genes were omitted to avoid a strong effect of donor gender. The resulting cross-species data consisted of 11,959 genes across 93 samples total. For visualization purposes, gene expression levels that had subject effects regressed out were obtained for both the human and mouse samples, as follows. The original counts were transformed to a continuous measure using the variance stabilizing transformation (VST)

available from the DESeq2 R package version V1.20.0.(Love et al., 2014). Using these transformed values, linear mixed models were fit to each gene with the nlme R package version V3.1–137. These models included an intercept only for fixed effects along with a random intercept for each subject. The expression values with subject regressed out were obtained by adding the overall intercept with the residual error terms (thus excluding the random intercepts). For the small number of genes where the linear mixed models had a singular fit (i.e., the random intercept variance was estimated to be 0), the original expression values were retained. This was performed separately for the human and mouse samples.

To further account for differences in expression levels across species, the VST subject-regressed data were quantile-normalized using the preprocessCore package (version 1.40.0) in R version 3.4.2 (Figure S3). The quantile-normalized data were used in all subsequent analyses, except the Z-score analysis described below.

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 (Satija et al., 2015). 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 datasets. 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). (Figures 3, and S4).

Identifying cell-type specific marker genes—Marker genes were identified as those significantly overexpressed primarily in one MP subtype versus all others within a species. Overexpression 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 overexpression 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 (Kuznetsova et al., 2017). 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. 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, as well as genes that could be marker genes for multiple cell types.

Identifying human-mouse MP analogs—A popular method for determining the corresponding MP subtype between species is to minimize distances or maximize correlations between sample groups using the quantile-normalized gene expression vectors, or 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). Moreover, determining the corresponding cell types by maximizing correlation between samples is highly dependent on the set of genes over which the correlation is computed (Figure S5). Whether using all genes (Figure S5A) or only the highly variable genes (hvg) to avoid flat or non-expressed genes dominating the correlation (Figure S5B), all samples are nearly equally correlated, with little discrimination. Moreover, the pairings which maximize correlation using all genes do not always agree with the pairings using the hvg. Another alternative to determining homologous subtypes is to intersect marker gene lists and report as the best match that which has the largest marker gene overlap (Figure S6). However, for the marker gene determination used here (positive Z-score and p.adj > 0.05), ~4,000–7,000 of the initial 11,959 genes were chosen as potential marker genes per cell type. Given the large marker gene set size relative to the total set of genes, comparing any two sets would necessary overlap of nearly 50% (Figure S6A). Restricting the overlap analysis to fixed sized sets of the top ranked marker genes per candidate pair results in different 'best' matches depending on the size of the marker set (Figure S6B). To compute an unbiased correspondence between human and murine MP, the data were analyzed instead based on a summary value of average overlap percentage when considering ordered marker

set sizes from 1 to 1000 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. (Figures 4A, S7, and S8). 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, i.e., marker genes unique to the cell type. For each candidate subtype cross-species pair, the overlap percentage was averaged across all list sizes from 1 to 1000 (Figure 4B). A high mean CAT percent agreement indicates that the pair consistently agrees on strong overlap, regardless of the marker set size (from 1 to 1000). The cross-species analog was chosen as that which maximized the mean CAT percentage agreement across all list sizes from 1 to 1000 (Figure 4B).

Comparison to previous DC and monocyte signature gene sets—The signature gene sets defined by Villani et al., (Villani et al., 2017) for the six DC subsets and for four monocyte subsets. Genes not appearing in our dataset were first removed and the top 20 genes per subset were identified using the reported AUC value to rank the genes (AUC column, Tables 1, S2 and S5 in Villani et al. (2017) paper). Signature gene sets defined by Dutertre et al., (Dutertre et al., 2019) appearing in our dataset or occurring as multiples in their signatures were first removed and then the top 20 genes per subset were defined based on decreasing order of average log fold change (avg_logFC column, Table S2 in Dutertre et al. (2019) paper). The correspondence of cluster number to cluster identity was established using the signature genes highlighted for each cluster in Dutertre et al. Signature gene sets from Brown et al., (Brown et al., 2019); were created from the normalized count data available the Gene Expression Omnibus (GEO) accession GSE137710. Following the methods described in Brown et al., a score for each gene for each subset was computed as the product of the earth mover's distance (EMD) and the area under the ROC curve (AUC) using the gene to distinguish between cells within the cluster versus all cells outside the cluster. Genes not appearing in our dataset were removed and the top 20 genes per DC subset were defined based on decreasing order of the gene score. The top scoring genes were verified as the genes annotated in Brown et al. MP signature gene sets were defined by Travaglini et al. (2020). Genes not appearing in our dataset were first removed and the top 20 genes per subset were identified based on decreasing order of the average log fold change (avg_logFC column, Table S4 in Travaglini et al. (2020) paper).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

Alferink J, Lieberam I, Reindl W, Behrens A, Weiss S, Hüser N, Gerauer K, Ross R, Reske-Kunz AB, Ahmad-Nejad P, et al. (2003). Compartmentalized production of CCL17 in vivo: strong inducibility in peripheral dendritic cells contrasts selective absence from the spleen. J. Exp. Med 197, 585–599. [PubMed: 12615900]

Altmann DM (2018). A Nobel Prize-worthy pursuit: cancer immunology and harnessing immunity to tumour neoantigens. Immunology 155, 283–284. [PubMed: 30320408]

Atif SM, Nelsen MK, Gibbings SL, Desch AN, Kedl RM, Gill RG, Marrack P, Murphy KM, Grazia TJ, Henson PM, and Jakubzick CV (2015). Cutting Edge: Roles for Batf3-Dependent APCs in the Rejection of Minor Histocompatibility Antigen-Mismatched Grafts. J. Immunol 195, 46–50. [PubMed: 26034174]

Atif SM, Gibbings SL, and Jakubzick CV (2018a). Isolation and Identification of Interstitial Macrophages from the Lungs Using Different Digestion Enzymes and Staining Strategies. Methods Mol. Biol 1784, 69–76. [PubMed: 29761388]

Atif SM, Gibbings SL, Redente EF, Camp FA, Torres RM, Kedl RM, Henson PM, and Jakubzick CV (2018b). Immune Surveillance by Natural IgM Is Required for Early Neoantigen Recognition and Initiation of Adaptive Immunity. Am. J. Respir. Cell Mol. Biol 59, 580–591. [PubMed: 29953261]

Bachem A, Güttler S, Hartung E, Ebstein F, Schaefer M, Tannert A, Salama A, Movassaghi K, Opitz C, Mages HW, et al. (2010). 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. [PubMed: 20479115]

Bain CC, Hawley CA, Garner H, Scott CL, Schridde A, Steers NJ, Mack M, Joshi A, Guilliams M, Mowat AM, Geissmann F, et al. (2016). Long-lived self-renewing bone marrow-derived macrophages displace embryo-derived cells to inhabit adult serous cavities. Nat. Commun 7, ncomms11852.

Bajpai G, Schneider C, Wong N, Bredemeyer A, Hulsmans M, Nahrendorf M, Epelman S, Kreisel D, Liu Y, Itoh A, et al. (2018). The human heart contains distinct macrophage subsets with divergent origins and functions. Nat. Med 24, 1234–1245. [PubMed: 29892064]

Begley CG, and Ellis LM (2012). Drug development: raise standards for preclinical cancer research. Nature 483, 531–533. [PubMed: 22460880]

Bharat A, Bhorade SM, Morales-Nebreda L, McQuattie-Pimentel AC, Soberanes S, Ridge K, DeCamp MM, Mestan KK, Perlman H, Budinger GR, and Misharin AV (2016). Flow Cytometry Reveals Similarities Between Lung Macrophages in Humans and Mice. Am. J. Respir. Cell Mol. Biol 54, 147–149. [PubMed: 26274047]

Bolstad BM, Irizarry RA, Astrand M, and Speed TP (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19, 185–193. [PubMed: 12538238]

Bosteels C, Neyt K, Vanheerswynghels M, Van Helden MJ, Sichien D, Debeuf N, De Prijck S, Bosteels V, Vandamme N, Martens L, et al. (2020). Inflammatory Type 2 cDCs Acquire Features of cDC1s and Macrophages to Orchestrate Immunity to Respiratory Virus Infection. Immunity 52, 1039–1056.e9. [PubMed: 32392463]

Brown CC, Gudjonson H, Pritykin Y, Deep D, Lavallee VP, Mendoza A, Fromme R, Mazutis L, Ariyan C, Leslie C, Pe'Er D, et al. (2019). Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. Cell 179, 846–863.e24. [PubMed: 31668803]

Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol 36, 411–420. [PubMed: 29608179]

Caminschi I, Proietto AI, Ahmet F, Kitsoulis S, Shin Teh J, Lo JC, Rizzitelli A, Wu L, Vremec D, van Dommelen SL, et al. (2008). The dendritic cell subtype-restricted C-type lectin Clec9A is a target for vaccine enhancement. Blood 112, 3264–3273. [PubMed: 18669894]

Chakarov S, Lim HY, Tan L, Lim SY, See P, Lum J, Zhang XM, Foo S, Nakamizo S, Duan K, et al. (2019). Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches. Science 363, eaau0964.

Desch AN, Randolph GJ, Murphy K, Gautier EL, Kedl RM, Lahoud MH, Caminschi I, Shortman K, Henson PM, and Jakubzick CV (2011). CD103+ pulmonary dendritic cells preferentially acquire and present apoptotic cell-associated antigen. J. Exp. Med 208, 1789–1797. [PubMed: 21859845]

Desch AN, Gibbings SL, Clambey ET, Janssen WJ, Slansky JE, Kedl RM, Henson PM, and Jakubzick C (2014). Dendritic cell subsets require cis-activation for cytotoxic CD8 T-cell induction. Nat. Commun 5, 4674. [PubMed: 25135627]

Desch AN, Gibbings SL, Goyal R, Kolde R, Bednarek J, Bruno T, Slansky JE, Jacobelli J, Mason R, Ito Y, et al. (2016). Flow Cytometric Analysis of Mononuclear Phagocytes in Nondiseased Human Lung and Lung-Draining Lymph Nodes. Am. J. Respir. Crit. Care Med 193, 614–626. [PubMed: 26551758]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Dutertre CA, Becht E, Irac SE, Khalilnezhad A, Narang V, Khalilnezhad S, Ng PY, Van Den Hoogen LL, Leong JY, Lee B, et al. (2019). Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. Immunity 51, 573–589.e8. [PubMed: 31474513]

Epelman S, Lavine KJ, Beaudin AE, Sojka DK, Carrero JA, Calderon B, Brija T, Gautier EL, Ivanov S, Satpathy AT, et al. (2014). Embryonic and adult-derived resident cardiac macrophages are maintained through distinct mechanisms at steady state and during inflammation. Immunity 40, 91–104. [PubMed: 24439267]

Förster R, Schubel A, Breitfeld D, Kremmer E, Renner-Müller I, Wolf E, and Lipp M (1999). CCR7 coordinates the primary immune response by es tablishing functional microenvironments in secondary lymphoid organs. Cell 99, 23–33. [PubMed: 10520991]

Gautier EL, Shay T, Miller J, Greter M, Jakubzick C, Ivanov S, Helft J, Chow A, Elpek KG, Gordonov S, et al.; Immunological Genome Consortium (2012). Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. Nat. Immunol 13, 1118–1128. [PubMed: 23023392]

Gibbings SL, and Jakubzick CV (2018a). A Consistent Method to Identify and Isolate Mononuclear Phagocytes from Human Lung and Lymph Nodes. Methods Mol. Biol 1799, 381–395. [PubMed: 29956166]

Gibbings SL, and Jakubzick CV (2018b). Isolation and Characterization of Mononuclear Phagocytes in the Mouse Lung and Lymph Nodes. Methods Mol. Biol 1809, 33–44. [PubMed: 29987780]

Gibbings SL, Thomas SM, Atif SM, McCubbrey AL, Desch AN, Danhorn T, Leach SM, Bratton DL, Henson PM, Janssen WJ, and Jakubzick CV (2017). Three Unique Interstitial Macrophages in the Murine Lung at Steady State. Am. J. Respir. Cell Mol. Biol 57, 66–76. [PubMed: 28257233]

Guilliams M, and Scott CL (2017). Does niche competition determine the origin of tissue-resident macrophages? Nat. Rev. Immunol 17, 451–460. [PubMed: 28461703]

Guilliams M, Lambrecht BN, and Hammad H (2013). Division of labor between lung dendritic cells and macrophages in the defense against pulmonary infections. Mucosal Immunol. 6, 464–473. [PubMed: 23549447]

Guilliams M, Dutertre CA, Scott CL, McGovern N, Sichien D, Chakarov S, Van Gassen S, Chen J, Poidinger M, De Prijck S, et al. (2016). Unsupervised High-Dimensional Analysis Aligns Dendritic Cells across Tissues and Species. Immunity 45, 669–684. [PubMed: 27637149]

Haniffa M, Shin A, Bigley V, McGovern N, Teo P, See P, Wasan PS, Wang XN, Malinarich F, Malleret B, et al. (2012). Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells. Immunity 37, 60–73. [PubMed: 22795876]

Hildner K, Edelson BT, Purtha WE, Diamond M, Matsushita H, Kohyama M, Calderon B, Schraml BU, Unanue ER, Diamond MS, et al. (2008). Batf3 deficiency reveals a critical role for CD8alpha + dendritic cells in cytotoxic T cell immunity. Science 322, 1097–1100. [PubMed: 19008445]

Huysamen C, Willment JA, Dennehy KM, and Brown GD (2008). 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. [PubMed: 18408006]

Ingersoll MA, Spanbroek R, Lottaz C, Gautier EL, Frankenberger M, Hoffmann R, Lang R, Haniffa M, Collin M, Tacke F, et al. (2010). Comparison of gene expression profiles between human and mouse monocyte subsets. Blood 115, e10–e19. [PubMed: 19965649]

Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, et al. (2005). Multiple-laboratory comparison of microarray platforms. Nat. Methods 2, 345–350. [PubMed: 15846361]

Jakubzick C, Tacke F, Llodra J, van Rooijen N, and Randolph GJ (2006). Modulation of dendritic cell trafficking to and from the airways. J. Immunol 176, 3578–3584. [PubMed: 16517726]

Jakubzick C, Helft J, Kaplan TJ, and Randolph GJ (2008). 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. [PubMed: 18662693]

Jakubzick C, Gautier EL, Gibbings SL, Sojka DK, Schlitzer A, Johnson TE, Ivanov S, Duan Q, Bala S, Condon T, et al. (2013). Minimal differentiation of classical monocytes as they survey steady-state tissues and transport antigen to lymph nodes. Immunity 39, 599–610. [PubMed: 24012416]

Jakubzick CV, Randolph GJ, and Henson PM (2017). Monocyte differentiation and antigen-presenting functions. Nat. Rev. Immunol 17, 349–362. [PubMed: 28436425]

Jiang H, Lei R, Ding SW, and Zhu S (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15, 182. [PubMed: 24925680]

Jongbloed SL, Kassianos AJ, McDonald KJ, Clark GJ, Ju X, Angel CE, Chen CJ, Dunbar PR, Wadley RB, Jeet V, et al. (2010). 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. [PubMed: 20479116]

Kim TS, Gorski SA, Hahn S, Murphy KM, and Braciale TJ (2014). 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. [PubMed: 24631155]

Kuznetsova A, Brockhoff PB, and Christensen RHB (2017). lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw 82, 25418.

Larson SR, Atif SM, Gibbings SL, Thomas SM, Prabagar MG, Danhorn T, Leach SM, Henson PM, and Jakubzick CV (2016). Ly6C(+) monocyte efferocytosis and cross-presentation of cell-associated antigens. Cell Death Differ. 23, 997–1003. [PubMed: 26990659]

León B, and Ardavín C (2008). Monocyte migration to inflamed skin and lymph nodes is differentially controlled by L-selectin and PSGL-1. Blood 111, 3126–3130. [PubMed: 18184867]

Liao Y, Smyth GK, and Shi W (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 47, e47. [PubMed: 30783653]

Lim HY, Lim SY, Tan CK, Thiam CH, Goh CC, Carbajo D, Chew SHS, See P, Chakarov S, Wang XN, et al. (2018). Hyaluronan Receptor LYVE-1-Expressing Macrophages Maintain Arterial Tone through Hyaluronan-Mediated Regulation of Smooth Muscle Cell Collagen. Immunity 49, 326–341.e7. [PubMed: 30054204]

Liu K, Victora GD, Schwickert TA, Guermonprez P, Meredith MM, Yao K, Chu FF, Randolph GJ, Rudensky AY, and Nussenzweig M (2009). In vivo analysis of dendritic cell development and homeostasis. Science 324, 392–397. [PubMed: 19286519]

Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. [PubMed: 25516281]

Masopust D, Sivula CP, and Jameson SC (2017). Of Mice, Dirty Mice, and Men: Using Mice To Understand Human Immunology. J. Immunol 199, 383–388. [PubMed: 28696328]

McCubbrey AL, Barthel L, Mohning MP, Redente EF, Mould KJ, Thomas SM, Leach SM, Danhorn T, Gibbings SL, Jakubzick CV, et al. (2018). Deletion of c-FLIP from CD11b[hi] Macrophages Prevents Development of Bleomycin-induced Lung Fibrosis. Am. J. Respir. Cell Mol. Biol 58, 66–78. [PubMed: 28850249]

Misharin AV, Morales-Nebreda L, Mutlu GM, Budinger GR, and Perlman H (2013). Flow cytometric analysis of macrophages and dendritic cell subsets in the mouse lung. Am. J. Respir. Cell Mol. Biol 49, 503–510. [PubMed: 23672262]

Misharin AV, Morales-Nebreda L, Reyfman PA, Cuda CM, Walter JM, McQuattie-Pimentel AC, Chen CI, Anekalla KR, Joshi N, Williams KJN, et al. (2017). Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. J. Exp. Med 214, 2387–2404. [PubMed: 28694385]

Mould KJ, Barthel L, Mohning MP, Thomas SM, McCubbrey AL, Danhorn T, Leach SM, Fingerlin TE, O'Connor BP, Reisz JA, et al. (2017). Cell Origin Dictates Programming of Resident versus Recruited Macrophages during Acute Lung Injury. Am. J. Respir. Cell Mol. Biol 57, 294–306. [PubMed: 28421818]

Mould KJ, Jackson ND, Henson PM, Seibold M, and Janssen WJ (2019). Single cell RNA sequencing identifies unique inflammatory airspace macrophage subsets. JCI Insight 4, e126556.

Murphy KM (2013). Transcriptional control of dendritic cell development. Adv. Immunol 120, 239–267. [PubMed: 24070387]

Nakano H, Moran TP, Nakano K, Gerrish KE, Bortner CD, and Cook DN (2015). Complement receptor C5aR1/CD88 and dipeptidyl peptidase-4/CD26 define distinct hematopoietic lineages of dendritic cells. J. Immunol 194, 3808–3819. [PubMed: 25769922]

Plantinga M, Guilliams M, Vanheerswynghels M, Deswarte K, Branco-Madeira F, Toussaint W, Vanhoutte L, Neyt K, Killeen N, Malissen B, et al. (2013). 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. [PubMed: 23352232]

Randolph GJ, Angeli V, and Swartz MA (2005). Dendritic-cell trafficking to lymph nodes through lymphatic vessels. Nat. Rev. Immunol 5, 617–628. [PubMed: 16056255]

Rapp M, Wintergerst MWM, Kunz WG, Vetter VK, Knott MML, Lisowski D, Haubner S, Moder S, Thaler R, Eiber S, et al. (2019). CCL22 controls immunity by promoting regulatory T cell communication with dendritic cells in lymph nodes. J. Exp. Med 216, 1170–1181. [PubMed: 30910796]

Reyfman PA, Walter JM, Joshi N, Anekalla KR, Mcquattie-Pimentel AC, Chiu S, Fernandez R, Akbarpour M, Chen CI, Ren Z, et al. (2019). Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. Am. J. Respir. Crit. Care Med 199, 1517–1536. [PubMed: 30554520]

Reynolds G, and Haniffa M (2015). Human and Mouse Mononuclear Phagocyte Networks: A Tale of Two Species? Front. Immunol 6, 330. [PubMed: 26124761]

Rosenthal R (1978). Combining Results of Independent Studies. Psychol. Bull 85, 185–193.

Satija R, Farrell JA, Gennert D, Schier AF, and Regev A (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol 33, 495–502. [PubMed: 25867923]

Satpathy AT, Kc W, Albring JC, Edelson BT, Kretzer NM, Bhattacharya D, Murphy TL, and Murphy KM (2012). Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages. J. Exp. Med 209, 1135–1152. [PubMed: 22615127]

Schlitzer A, McGovern N, Teo P, Zelante T, Atarashi K, Low D, Ho AW, See P, Shin A, Wasan PS, et al. (2013). IRF4 transcription factor-dependent CD11b+ dendritic cells in human and mouse control mucosal IL-17 cytokine responses. Immunity 38, 970–983. [PubMed: 23706669]

Schyns J, Bai Q, Ruscitti C, Radermecker C, De Schepper S, Chakarov S, Farnir F, Pirottin D, Ginhoux F, Boeckxstaens G, et al. (2019). Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung. Nat. Commun 10, 3964. [PubMed: 31481690]

Scott CL, Henri S, and Guilliams M (2014). Mononuclear phagocytes of the intestine, the skin, and the lung. Immunol. Rev 262, 9–24. [PubMed: 25319324]

Stutte S, Quast T, Gerbitzki N, Savinko T, Novak N, Reifenberger J, Homey B, Kolanus W, Alenius H, and Förster I (2010). Requirement of CCL17 for CCR7- and CXCR4-dependent migration of cutaneous dendritic cells. Proc. Natl. Acad. Sci. USA 107, 8736–8741. [PubMed: 20421491]

Tamoutounour S, Henri S, Lelouard H, de Bovis B, de Haar C, van der Woude CJ, Woltman AM, Reyal Y, Bonnet D, Sichien D, et al. (2012). 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. [PubMed: 22936024]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Tamoutounour S, Guilliams M, Montanana Sanchis F, Liu H, Terhorst D, Malosse C, Pollet E, Ardouin L, Luche H, Sanchez C, et al. (2013). Origins and functional specialization of macrophages and of conventional and monocyte-derived dendritic cells in mouse skin. Immunity 39, 925–938. [PubMed: 24184057]

Theisen DJ, Davidson JTT 4th, Briseño CG, Gargaro M, Lauron EJ, Wang Q, Desai P, Durai V, Bagadia P, Brickner JR, et al. (2018). WDFY4 is required for cross-presentation in response to viral and tumor antigens. Science 362, 694–699. [PubMed: 30409884]

Trapnell BC, Carey BC, Uchida K, and Suzuki T (2009). Pulmonary alveolar proteinosis, a primary immunodeficiency of impaired GM-CSF stimulation of macrophages. Curr. Opin. Immunol 21, 514–521. [PubMed: 19796925]

Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Conley SD, Mori Y, Seita J, et al. (2020). A molecular cell atlas of the human lung from single cell RNA sequencing. BioRxiv. 10.1101/742320.

Tussiwand R, Lee WL, Murphy TL, Mashayekhi M, Kc W, Albring JC, Satpathy AT, Rotondo JA, Edelson BT, Kretzer NM, et al. (2012). Compensatory dendritic cell development mediated by BATF-IRF interactions. Nature 490, 502–507. [PubMed: 22992524]

Tussiwand R, Everts B, Grajales-Reyes GE, Kretzer NM, Iwata A, Bagaitkar J, Wu X, Wong R, Anderson DA, Murphy TL, et al. (2015). Klf4 expression in conventional dendritic cells is required for T helper 2 cell responses. Immunity 42, 916–928. [PubMed: 25992862]

Ural BB, Yeung ST, Damani-Yokota P, Devlin JC, de Vries M, Vera-Licona P, Samji T, Sawai CM, Jang G, Perez OA, et al. (2020). Identification of a nerve-associated, lung-resident interstitial macrophage subset with distinct localization and immunoregulatory properties. Sci. Immunol 5, eaax8756.

Vermaelen KY, Carro-Muino I, Lambrecht BN, and Pauwels RA (2001). Specific migratory dendritic cells rapidly transport antigen from the airways to the thoracic lymph nodes. J. Exp. Med 193, 51–60. [PubMed: 11136820]

Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S, et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356, eaah4573.

Vulcano M, Albanesi C, Stoppacciaro A, Bagnati R, D'Amico G, Struyf S, Transidico P, Bonecchi R, Del Prete A, Allavena P, et al. (2001). Dendritic cells as a major source of macrophage-derived chemokine/CCL22 in vitro and in vivo. Eur. J. Immunol 31, 812–822. [PubMed: 11241286]

Wu X, Briseño CG, Durai V, Albring JC, Haldar M, Bagadia P, Kim KW, Randolph GJ, Murphy TL, and Murphy KM (2016). Mafb lineage tracing to distinguish macrophages from other immune lineages reveals dual identity of Langerhans cells. J. Exp. Med 213, 2553–2565. [PubMed: 27810926]

Yu YA, and Tighe RM (2018). Isolation and Characterization of Human Lung Myeloid Cells. Methods Mol. Biol 1809, 111–119. [PubMed: 29987786]

Yu YA, Hotten DF, Malakhau Y, Volker E, Ghio AJ, Noble PW, Kraft M, Hollingsworth JW, Gunn MD, and Tighe RM (2016). Flow Cytometric Analysis of Myeloid Cells in Human Blood, Bronchoalveolar Lavage, and Lung Tissues. Am. J. Respir. Cell Mol. Biol 54, 13–24. [PubMed: 26267148]

Highlights

- Transcriptional profiling defines cross-species MPs from lung, LN, blood, and spleen

- 13%–23% of the top 1,000 marker genes overlap in corresponding H-M MP counterparts

- CD88 expression distinguishes monocytes/macrophages from DCs in both species

- This study creates a resource to inform future translational studies targeting MPs
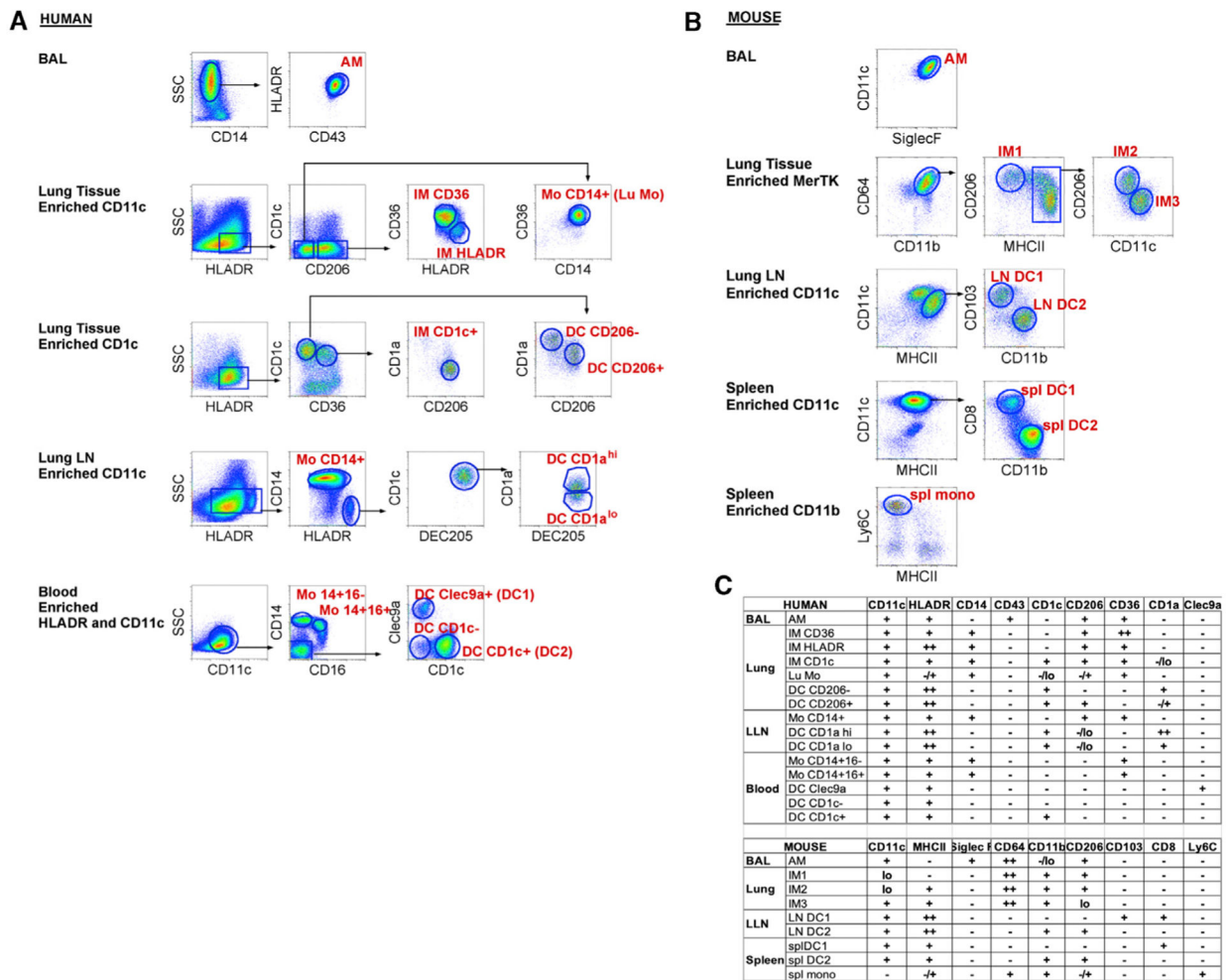
Figure 1. Gating Strategy to Sort Human and Mouse Mononuclear Phagocytes for RNA Isolation

(A and B) Fluorescence-activated cell sorting (FACS) gating strategy used to isolate human (A) or mouse (B) mononuclear phagocytes (MPs, labeled in red) from the indicated tissues. Before sorting, cell suspensions were magnetically enriched as indicated and gated to select Live; Single; CD45$^+$; Lineage$^-$ cells as shown in Figures S1 and S2. Arrows indicate where further analysis is performed on the specified subpopulation. The data presented are representative of 3–6 replicates per sort.

(C) Table illustrating cell surface marker expression by MP subsets in human (top) and mouse (bottom). – or lo signifies no or little expression, + or ++ indicates expression or strong expression, –/+ indicates heterogeneous expression of marker proteins.
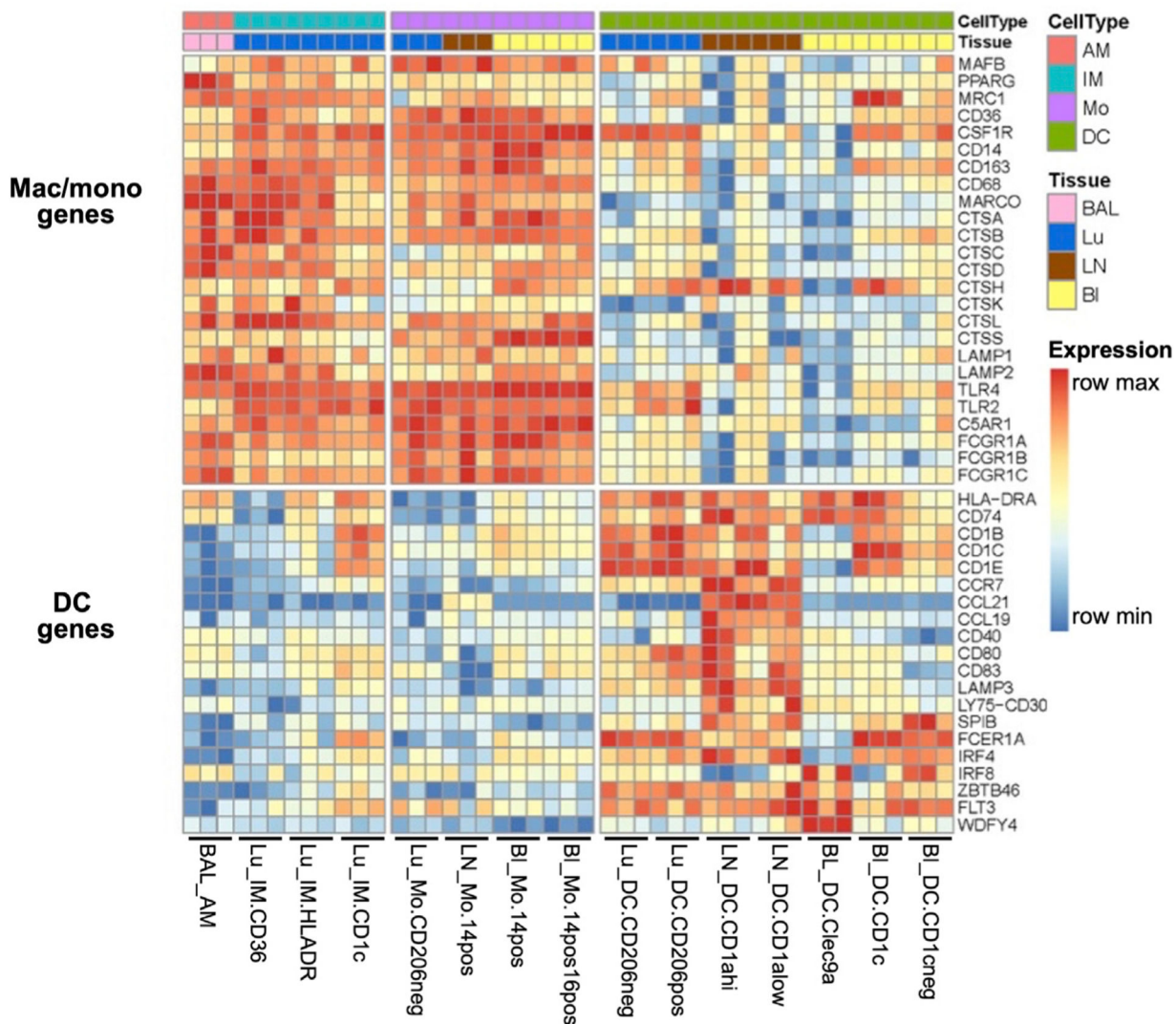
Figure 2. Confirmation of Human MP Subtypes Based on Classical Gene Signatures
Expression of signature DC and macrophage transcripts from RNA-seq are shown as
normalized values. Data first undergo a variance stabilizing transformation on read counts,
in which the donor identity has been regressed out 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. Although the number of
replicates ranges from 3–6, only the first 3 replicates are shown for each subtype.
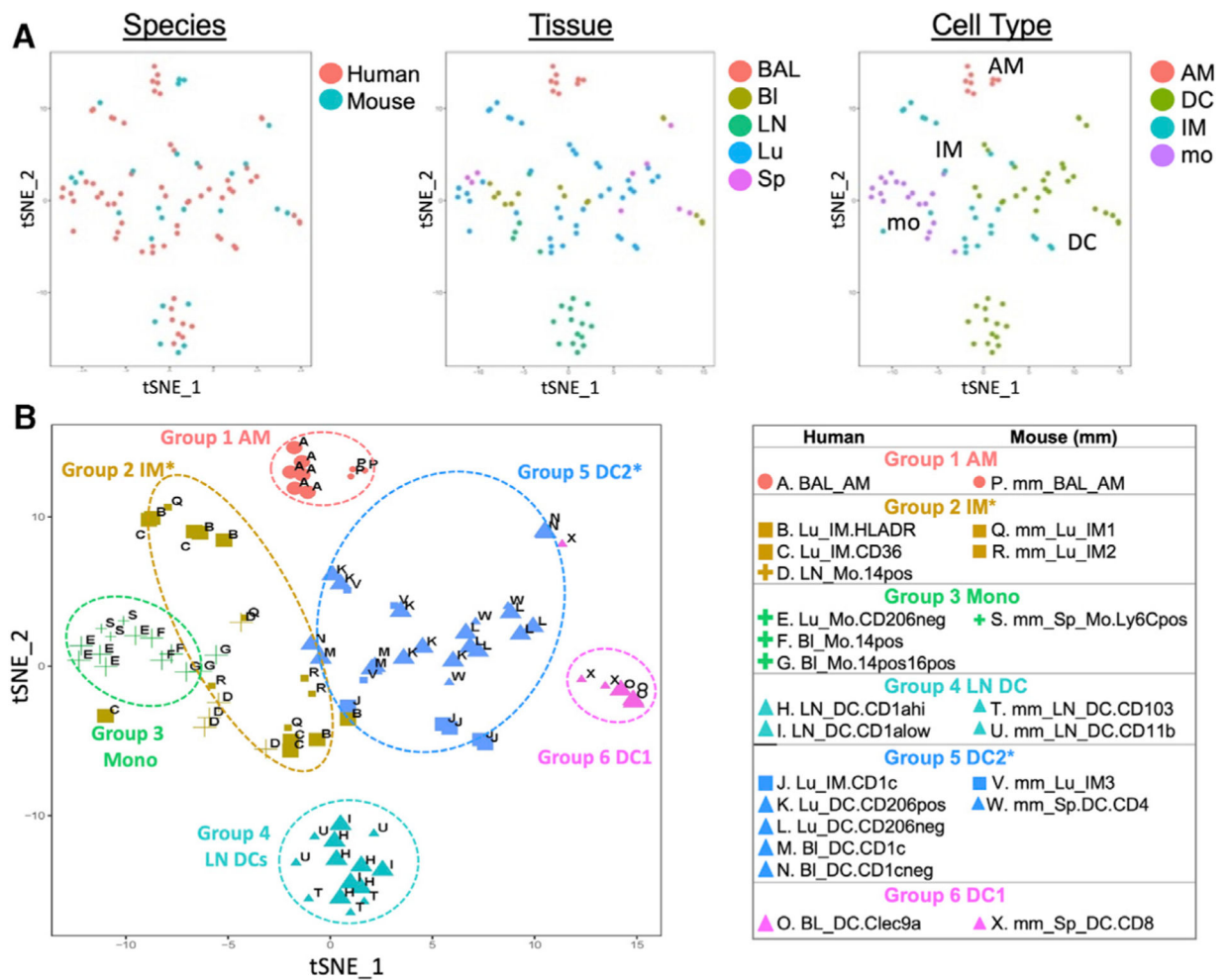
Figure 3. Visualization of MP Subtypes across Species

Quantile-normalized Variance Stabilizing Transformed (VST) subject-regressed data from 3–6 donors per subtype for 15 human and 9 murine MP subtypes were subjected to the Seurat version 2.0 pipeline to visualize human and mouse data within a common 2-dimensional (2D) space.

(A) Samples in the resulting 2D projection are colored by species, tissue, or cell type.

(B) Samples are grouped into 6 clusters based on later shared marker gene analysis (see Figure 4). Shape indicates cell type, size indicates species, and color indicates group assignment. MP subtypes are denoted by the letters A through X. Note that each group contains related subtypes across species.
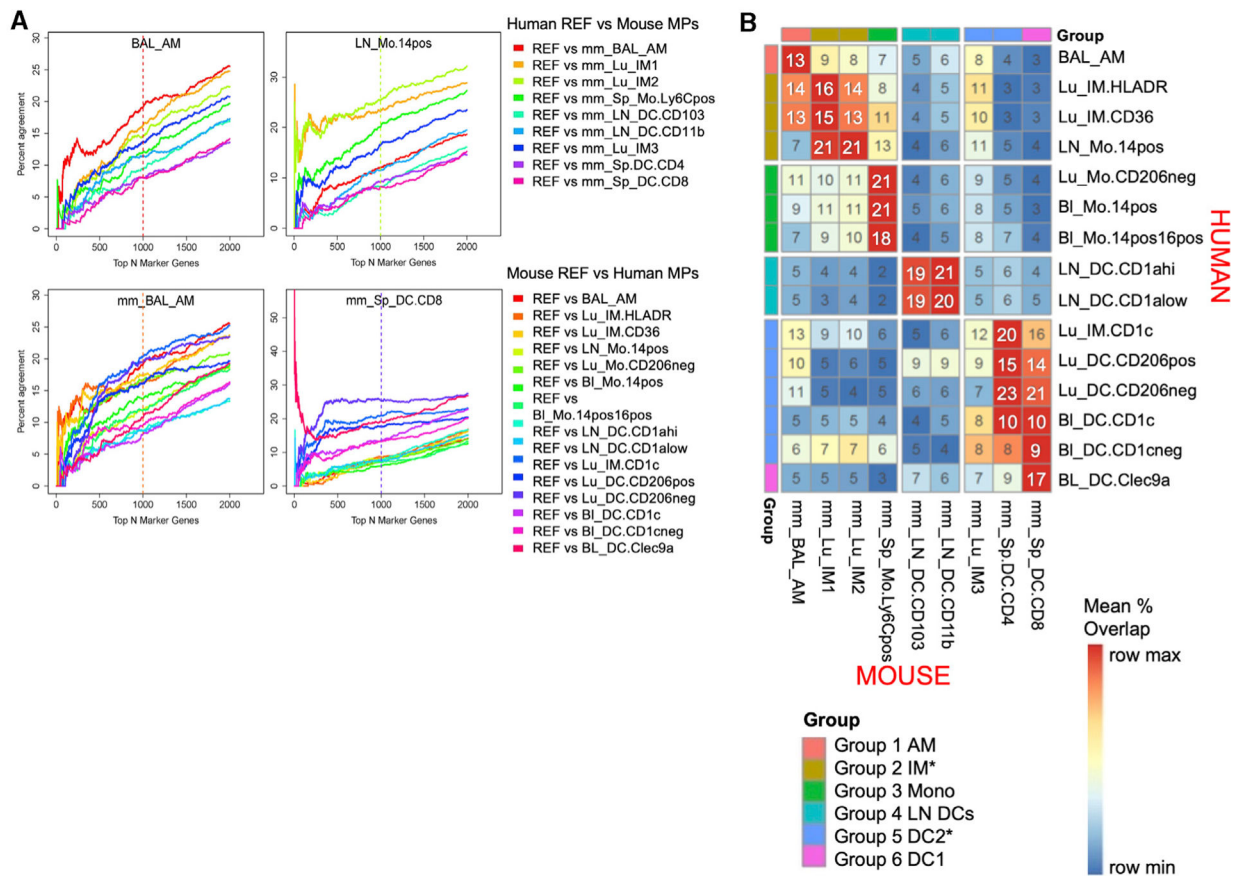
Figure 4. Correspondence of MPs across Species

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 Method Details).

(A) For each reference subtype (graph title) from either human (top) or mouse (bottom, signified by mm), 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. Percentage agreement is the percentage of genes identified as critical markers in the reference species that are also identified as critical markers for the comparison species up to that rank. For example, at rank 100 on the x axis, a value of 10% would indicate that of the marker genes in the top 100 in the reference species, 10% are in the top 100 of the comparison species. Each line shows a different cell population in the comparator species. The color of the vertical line indicates which subtype had the highest mean CAT overlap across all list sizes from 1 to 1,000 with respect to the reference.

(B) The mean CAT overlap over list sizes 1–1,000 is shown for each subtype pair. Rows are scaled by the maximum and the minimum to emphasize the highest subtype correspondence for the human subtypes. Rows and columns are ordered to cluster similar subtypes. Clusters of MP subtypes are then annotated by group number, and these groupings are used for the coloring in Figure 3B.
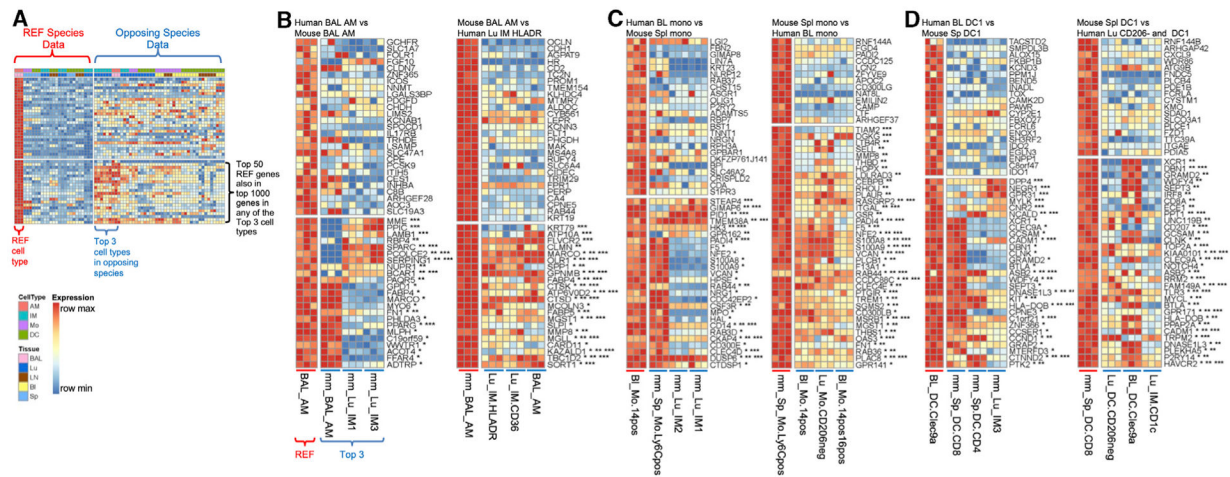
Figure 5. Well-Defined Human-Mouse Homologs

The expression values are shown for the top 50 marker genes in the reference cell type. The cell types are ordered by degree of overlap with the reference cell type, if within the same species, or by the mean CAT overlap shown in Figure 4B if in the other species. The data are scaled across the row by the gene minimum and maximum within each species to emphasize in which cell types the gene is most expressed. Genes among the top 50 in the reference that also appear in the top 1,000 ranked genes in the 1st-, 2nd-, or 3rd-best match marker list appear toward the bottom and are annotated with *, **, or ***, respectively. Note that annotation by an asterisk does not mean that the gene must also appear in the top 50 marker genes in the other species, only that it appears in the top 1,000 genes for the top 3 matches. Also, visualizing the top 3 matches rather than just the best match can help reveal cases when the best match, as determined using 1,000 reference marker genes, may differ from the best match suggested by comparing expression of just the top 50 reference genes, in which case the CAT plots can inform the reason for the perceived discrepancy (Figures S7 and S8).

(A) Example depiction of all of the data in all of the samples for a given reference and the top 3 closest matches according to Figure 4B.

(B) Highest correspondence for human and mouse AMs. Three replicates for the reference and 3 closest subtypes of the opposing species are shown.

(C) Reciprocal highest correspondence for human BL and mouse Spl monocytes. Three replicates for the reference and the 3 closest subtypes are shown.

(D) Highest correspondence for human BL DC Clec9a+ and mouse Spl CD8+ DC1. Three replicates for the reference and the 3 closest subtypes are shown.

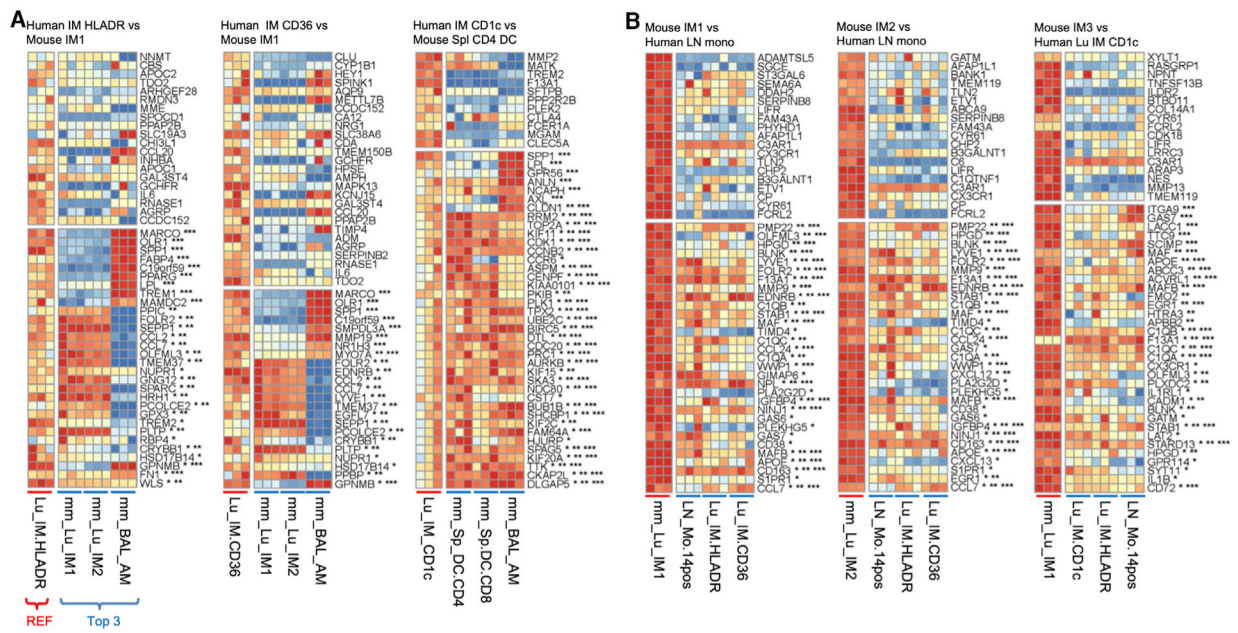**Figure 6. Corresponding Human and Mouse IM Subtypes**

Expression is shown for the first 50 ranked genes using the scaled data, as in Figure 5. Three replicates for the reference and the 3 closest subtypes of the opposing species are shown. Genes are annotated by whether they are also in the top 1,000 marker genes in the 3 highest corresponding subtypes with *, **, or ***, respectively.
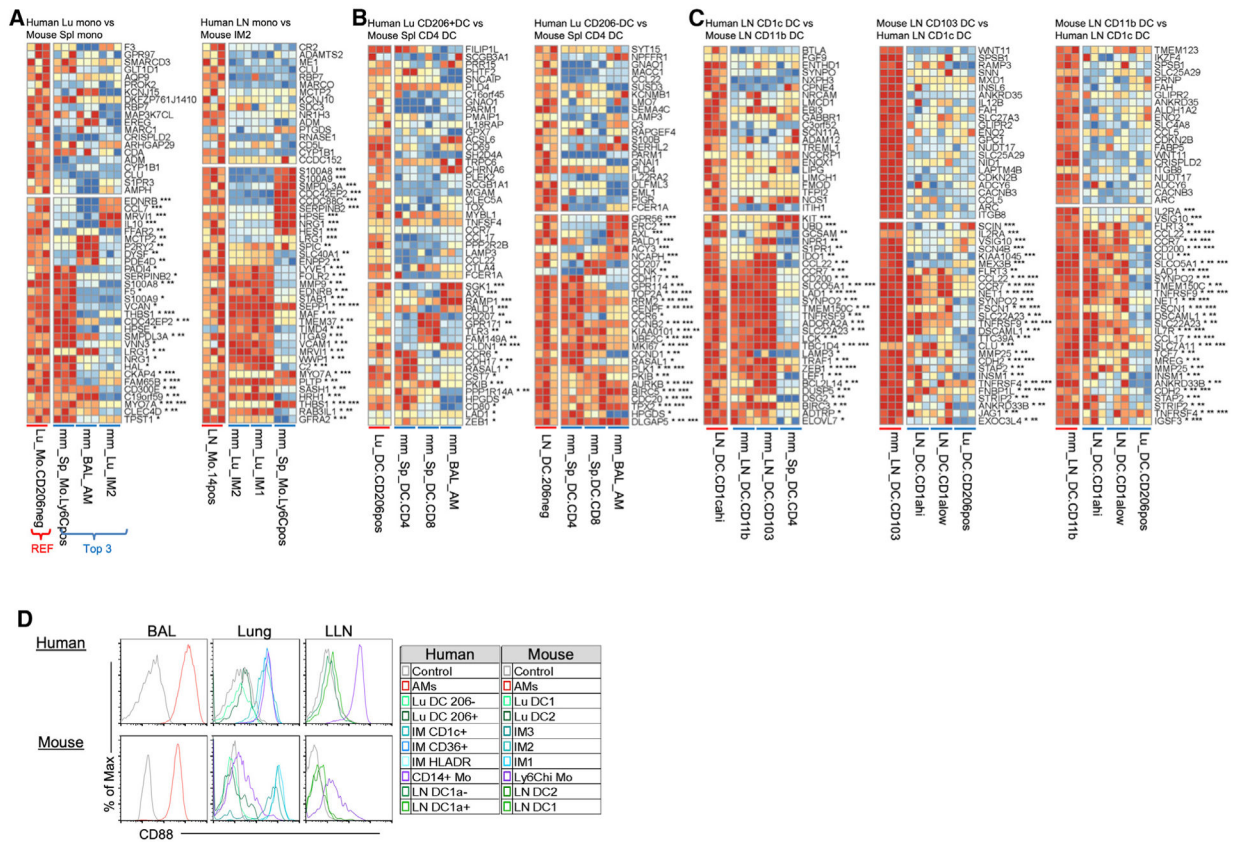
(A) Human IMs.

(B) Mouse IMs.

Figure 7. Corresponding Human and Mouse Monocytes and DC Subtypes; CD88 Expression Distinguishes Monocytes/Macrophages from DCs

Expression is shown for the first 50 ranked genes using the scaled data as in Figure 5. Three replicates for the reference and the 3 closest subtypes of the opposing species are shown. Genes are annotated by whether they are also in the top 1,000 marker genes in the 3 highest corresponding subtypes with *, **, or ***, respectively.

(A) Human monocytes.

(B) Human lung DCs.

(C) Human and mouse LN DCs.

(D) Histogram shows the protein expression of CD88 in human and mouse MPs in Figure 1.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti human CD45 PerCP clone Hl30	BioLegend	RRID:AB_893337; BioLegend Cat# 304026
Mouse anti human CD3 Pacific Blue, clone UCHT1	BioLegend	RRID:AB_493094; BioLegend Cat# 300417
Mouse anti human CD15 Pacific Blue, clone W6D3	BioLegend	RRID:AB_2105359; BioLegend Cat# 323024
Mouse anti human CD20 Pacific Blue, clone 2H7	BioLegend	RRID:AB_493651; BioLegend Cat# 302320
Mouse anti human CD56 eFluor450, clone TULY56	Invitrogen/ThermoFisher Scientific	RRID:AB_2574021; Invitrogen/ThermoFisher Cat# 48-0566-42
Mouse anti human CD14 V500, clone Mf P9	BD Biosciences	RRID:AB_2737727; BD Biosciences Cat# 562693
Mouse anti human HLA-DR APC-Cy7, clone L243	BD Biosciences	RRID:AB_399974; BD Biosciences Cat# 335796
Mouse anti human CD43 FITC, clone eBio84–3C1	Invitrogen/ThermoFisher Scientific	RRID:AB_10717075; Invitrogen/ThermoFisher Cat# 11-0439-42
Mouse anti human CD1c PE, clone AD5–8E7	Miltenyi Biotec	RRID:AB_244184; Miltenyi Biotec Cat# 130-090-507
Mouse anti human CD206 PerCp-Cy5.5, clone 15–2	BioLegend	RRID:AB_10899411; BioLegend Cat# 321122
Mouse anti human CD36 APC, clone AC106	Miltenyi Biotec	RRID:AB_10829944; Miltenyi Biotec Cat# 130-095-475
Mouse anti human CD1a FITC, clone CB-T6	Ancell	Ancell Cat# 145–040
Mouse anti human Dec205 APC, clone HD30	BioLegend	RRID:AB_2281392; BioLegend Cat# 342208
Mouse anti human CD16 PerCP-Cy5.5, clone CB16	Invitrogen/ThermoFisher Scientific	RRID:AB_1834389; Invitrogen/ThermoFisher Cat# 46-0168-42
Mouse anti human Clec9a	BioLegend	RRID:AB_10965546; BioLegend Cat# 353804
Mouse anti human CD11c PE-Cy7, clone 3.9	Invitrogen/ThermoFisher Scientific	RRID:AB_1582274; Invitrogen/ThermoFisher Cat# 25-0116-42
Mouse anti human CD88 PerCP-Cy5.5, clone S5/1	BioLegend	RRID:AB_11203904; BioLegend Cat# 344312
Mouse anti human CD141 APC, clone AD5–14H12	Miltenyi Biotec	RRID:AB_244170; Miltenyi Biotec Cat# 130-090-907
Mouse anti human CD64 PE, clone 10.1	BioLegend	RRID:AB_314492; BioLegend Cat# 305008
Rat anti mouse CD45 V500, clone 30-F11	BD Biosciences	RRID:AB_10697046; BD Biosciences Cat# 561487
Rat anti mouse Ly6G BV421, clone 1A8	BD Biosciences	RRID:AB_2737756; BD Biosciences Cat# 562737
Rat anti mouse CD3 eF450, clone 17A2	Invitrogen/ThermoFisher Scientific	RRID:AB_1272193; Invitrogen/ThermoFisher Cat# 48-0032-82
Rat anti mouse NK1.1 eF450 clone PK136	Invitrogen/ThermoFisher Scientific	RRID:AB_2043877; Invitrogen/ThermoFisher Cat# 48-5941-82
Rat anti mouse B220 eF450, clone RA3–6B2	Invitrogen/ThermoFisher Scientific	RRID:AB_1548761; Invitrogen/ThermoFisher Cat# 48-0452-82
Armenian Hamster anti mouse CD11c PE-Cy7, clone N418	Invitrogen/ThermoFisher Scientific	RRID:AB_469590; Invitrogen/ThermoFisher Cat# 25-0114-82
Rat anti mouse Siglec F PE, clone E50–2440	BD Biosciences	RRID:AB_394341; BD Biosciences Cat# 552126
Mouse anti mouse CD64 Alexa Fluor 647, clone X54–5/7.1	BD Biosciences	RRID:AB_647120; BD Biosciences Cat# 558539
Goat anti mouse MerTK biotinylated, polyclonal	R&D Systems	RRID:AB_2098563; R&D Systems Cat# BAF591
Rat anti mouse CD11b PerCp-Cy5.5, clone M1/70	BioLegend	RRID:AB_893232; BioLegend Cat# 101228
Rat anti mouse CD206 PE, clone C068C2	BioLegend	RRID:AB_10895754; BioLegend Cat# 141706

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rat anti mouse IA-IE V500, clone M5/114.15.2	BD Biosciences	RRID:AB_11153488; BD Biosciences Cat# 562366
Armenian Hamster anti mouse CD103 PE, clone 2E7	Invitrogen/ThermoFisher Scientific	RRID:AB_465799; Invitrogen/ThermoFisher Cat# 12-1031-82
Rat anti mouse CD8 APC, clone 53–6.7	BioLegend	RRID:AB_312751; BioLegend Cat# 100712
Rat anti mouse Ly6C PerCP-Cy5.5, clone HK1.4	BioLegend	RRID:AB_1659241; BioLegend Cat# 128012
Rat anti mouse CD88 PE, clone 20/70	BioLegend	RRID:AB_2243735; BioLegend Cat# 135806
Rat anti mouse CD43 BV421, clone BV421	BD Biosciences	RRID:AB_2665409; BD Biosciences Cat# 562958
Rat anti mouse Gr1 eF450, clone RB6–8C5	Invitrogen/ThermoFisher Scientific	RRID:AB_1548788; Invitrogen/ThermoFisher Cat# 48-5931-82
Mouse anti human CD141 biotin, clone AD5–14H12	Miltenyi Biotec	RRID:AB_2726093; Miltenyi Biotec Cat# 130-113-316
Mouse anti human CD1c biotin, clone AD5–8E7	Miltenyi Biotec	RRID:AB_244312; Miltenyi Biotec Cat# 130-090-692
Mouse anti human CD11c biotin, clone 3.9	Invitrogen/ThermoFisher Scientific	RRID:AB_466366; Invitrogen/ThermoFisher Cat# 13-0116-82
Chemicals, Peptides, and Recombinant Proteins		
Elastase	Worthington Biochemical Corporation	Worthington Cat# LS002294
Collagenase D	Roche, via Sigma Aldrich	Roche Cat# 11088882001
Liberase TM	Roche, via Sigma Aldrich	Roche Cat# 5401127001
4',6-diamidino-2-phenylindole	Invitrogen/ThermoFisher Scientific	Invitrogen/ThermoFisher Scientific Cat# D1306
Critical Commercial Assays		
Anti-biotin microbeads	Miltenyi Biotec	Miltenyi Cat# 130-090-485
Anti-PE microbeads	Miltenyi Biotec	Miltenyi Cat# 130-048-801
Anti-mouse CD11c microbeads	Miltenyi Biotec	Miltenyi Cat# 130-125-835
Anti-mouse CD11b microbeads	Miltenyi Biotec	Miltenyi Cat# 130-049-601
RNeasy Micro Kit	QIAGEN	QIAGEN Cat# 74004
QIAshredder	QIAGEN	QIAGEN Cat# 79654
SMART-Seq® v4 Ultra® Low Input RNA Kit	Clontech	Takara Bio Cat# 634891
Nextera XT DNA Library Preparation Kit	Illumina	Illumina Cat# FC-131–1096
Deposited Data		
Murine monocyte and DC RNaseq	This paper	NCBI Gene Expression Omnibus; Accession number GSE132911
Murine AM and IM RNaseq data	(Gibbings et al., 2017)	NCBI Gene Expression Omnibus; Accession number GSE94135
Ensembl 75 mouse	Ensembl	https://feb2014.archive.ensembl.org/
Ensembl 75 human	Ensembl	https://feb2014.archive.ensembl.org/
Ensembl BioMart version 75	Ensembl	https://feb2014.archive.ensembl.org/biomart/
Mouse and Human DC heterogeneity	(Brown et al., 2019)	NCBI Gene Expression Omnibus; Accession number GSE137710
Experimental Models: Organisms/Strains		
C57BL/6 mouse	Charles River Laboratories	CRL:027
Software and Algorithms		
bcl2fastq converter (version 2.17)	Illumina	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html

REAGENT or RESOURCE	SOURCE	IDENTIFIER
skewer (version 0.2.2)	(Jiang et al., 2014)	https://github.com/relipmoc/skewer
FastQC (version 0.11.5)	Babraham Institute	http://www.bioinformatics.babraham.ac.uk/projects/fastqc
STAR	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
subread software package (version 1.5.2)	(Liao et al., 2019)	http://bioconductor.org/packages/release/bioc/html/Rsubread.html
DESeq2 R package version V1.20.0	(Love et al., 2014)	https://github.com/mikelove/DESeq2
nlme R package version V3.1–137	R package version 3.1–148	https://cran.r-project.org/web/packages/nlme/index.html
preprocessCore package (version 1.40.0) in R version 3.4.2	Ben Bolstad	https://github.com/bmbolstad/preprocessCore
Seurat 2.3.2 package for R version 3.4.2	(Satija et al., 2015)	https://www.satijalab.org/seurat, https://github.com/satijalab/seurat
lmerTest R package version 3.1–0	(Kuznetsova et al., 2017)	https://github.com/runehaubo/lmerTestR