Title:

Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture in humans

Authors:

Domínguez Conde C[1,#], Xu C[1,#], Jarvis LB[2,#], Gomes T[1,#], Howlett SK[2], Rainbow DB[2], Suchanek O[3], King HW[4], Mamanova L[1], Polanski K[1], Huang N[1], Fasouli ES[1], Mahbubani KT[5], Prete M[1], Tuck L[1], Richoz N[3], Tuong ZK[1,3], Campos L[1,6], Mousa HS[2], Needham EJ[2], Pritchard S[1,], Li T[1,], Elmentaite R[1,], Park J[1,], Menon DK[7], Bayraktar OA[1], James LK[4], Meyer KB[1,], Clatworthy MR[1,3], Saeb-Parsy K[5,*], Jones JL[2,*], Teichmann SA[1,8,*]

Affiliations:

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK
2. Department of Clinical Neurosciences, University of Cambridge
3. Molecular Immunity Unit, Department of Medicine, University of Cambridge, Cambridge, UK
4. Centre for Immunobiology, Blizard Institute, Queen Mary University of London, London, UK
5. Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, UK
6. West Suffolk Hospital NHS Trust, Bury Saint Edmunds, UK
7. Department of Anaesthesia, University of Cambridge, Cambridge, UK
8. Theory of Condensed Matter, Cavendish Laboratory, Department of Physics, University of Cambridge, JJ Thomson Ave, Cambridge CB3 0HE, UK

#Co-first authors

*Co-senior and co-corresponding authors. Email: st9@sanger.ac.uk, jls53@medschl.cam.ac.uk, ks10014@cam.ac.uk

Abstract:

Despite their crucial role in health and disease, our knowledge of immune cells within human tissues remains limited. Here, we surveyed the immune compartment of 15 tissues of six deceased adult donors by single-cell RNA sequencing and paired VDJ sequencing. To systematically resolve immune cell heterogeneity across tissues, we developed CellTypist, a machine learning tool for rapid and precise cell type annotation. Using this approach, combined with detailed curation, we determined the tissue distribution of 45 finely phenotyped immune cell types and states, revealing hitherto unappreciated tissue-specific features and clonal architecture of T and B cells. In summary, our multi-tissue approach lays the foundation for identifying highly resolved immune cell types by leveraging a common reference dataset, tissue-integrated expression analysis and antigen receptor sequencing.

One Sentence Summary:

We provide an immune cell atlas, including antigen receptor repertoire profiling, across lymphoid and non-lymphoid human tissues.

Main Text:

The immune system is a dynamic and integrated network made up of many different cell types distributed across the entire organism that act together to ensure effective host defence. In recent years, a growing appreciation of immune ontogeny and diversity across tissues has emerged. For example, we have gained new insights into how macrophages derived in embryogenesis contribute to the unique adaptation of tissue-resident myeloid cells, such as Langerhans cells in the skin, microglia in the brain and Kupffer cells in the liver (*1–3*). Other populations, such as innate lymphoid cells (ILCs), including natural killer (NK) cells and non-conventional (NKT, MAIT and γδ) T cells, have circulating counterparts but are highly enriched at barrier/mucosal sites where they mediate tissue defence and repair (*4*). In addition, following resolution of an immune response, antigen-specific, long-lived tissue-resident memory T cells (TRMs) have been shown to persist in many tissues, where they provide a first line of defence against secondary infections (reviewed in (*5–7*)).

Despite the central role of tissue immunity in homeostasis and disease, much of our current knowledge comes from animal studies. Historically, human immune cell analysis has focused on peripheral blood, providing a biased and incomplete view of human immunity. Recent work interrogating human tissue immunity using flow cytometry (*8–11*) and several organ-focused single-cell RNA sequencing (scRNA-seq) studies (*12–16*) have begun to address this knowledge gap, but few have analysed immune cells across multiple tissues from the same individual, which allows direct and controlled comparisons to be made. One such study by Szabo *et al.* has reported an analysis of T cells in three tissues from two donors (*17*). However, despite the effort to assemble murine (*18*) and human *(19, 20)* multi-tissue atlases, large-scale cross-tissue scRNA-seq studies that investigate tissue-specific features of the immune compartment remain scarce. To address this, here, we comprehensively profiled immune cell populations isolated from 15 donor-matched tissues from six deceased individuals representing in the order of 100,000 single cell transcriptomes.

Annotation of increasingly large single cell transcriptomics data sets remains a challenge, including the identification of rare cell subsets, and distinguishing previously described cell populations from novel discoveries. To capture the vast cellular diversity for annotating muti-tissue immune cells, we developed CellTypist, a machine learning framework for cell type prediction compiled from published studies across 20 human tissues. Combining this approach with the in-depth dissection of cellular heterogeneity within the myeloid and lymphoid compartments, we determined the frequency and transcriptional characteristics of immune cell types and states across human tissues. Moreover, we inferred the patterns of T and B cell migration across different tissues and their transition between cell states on the basis of antigen-receptor identity, gaining insights into tissue-dependent phenotypic and clonal plasticity.

Results

Single cell transcriptomics and antigen receptor sequence analysis of tissue mononuclear cells

To systematically assess immune cell type heterogeneity across human tissues, we performed scRNA-seq on 15 different tissues from six deceased organ donors (**Fig. 1A,B, Supplementary Table 1)**. Briefly, cells were isolated using a uniform protocol across tissues, with the exception of blood and bone marrow samples (see Methods for details). The tissues studied included primary (thymus and bone marrow) and secondary (spleen, thoracic and mesenteric lymph nodes) lymphoid organs, mucosal tissues (gut and lung), as well as liver, skeletal muscle and omentum. After stringent quality control, we obtained a total of 78,844 high quality hematopoietic cells, with higher yields from the lymphoid tissues in all donors (**Fig. S1A,B**). Using manual annotation based on marker gene expression (**Fig. S1C**), we identified 14 major cell populations represented across all donors and tissues from three major immune compartments: (i) T cells and ILCs, (ii) B cells and plasma cells, and (iii) myeloid cells (**Fig. 1C**,**D, Fig. S1D,E**). In addition, we identified four small distinct clusters consisting of mast cells, megakaryocytes/platelets (Mgk), plasmacytoid dendritic cells (pDCs) and a small population of immune progenitor cells. As expected, progenitors and megakaryocyte clusters were primarily found in the bone marrow and blood; macrophage and mast cell populations were enriched in the lung and lymphocytes mainly came from lymphoid organs (**Fig. S1E**). Analyses of the cell type composition within each tissue (**Fig. S1F**) further revealed that lymph nodes and thymus were rich in CD4+ T cells, while in the liver, spleen, bone marrow and gut, CD8+ T cells were predominant. In addition, gut regions were abundant in plasma cells, and the immune compartment of the lung parenchyma was dominated by monocytes and macrophages.

For further analysis of the identity of cell states, we developed a machine learning framework trained on published datasets from 20 tissues, as described in the next section.

CellTypist: a novel tool for annotating immune cell populations across tissues

Robust cell type annotation remains a major challenge in single-cell transcriptomics. To address the cellular heterogeneity in our cross-tissue study, we developed CellTypist (https://github.com/Teichlab/celltypist), a lightweight and interpretable cell type classification pipeline built on an expandable cell type reference atlas assembled from multiple tissues and studies (**Fig. 1E**, **Supplementary Note**). In brief, CellTypist is currently based on 20 tissues from 19 reference datasets (**Supplementary Note Figure 1A**) where cell type labels are harmonized at different hierarchically-structured levels (**Supplementary Note Figure 1B**). This hierarchy is of particular importance for determining immune cell identity at different levels of specificity (**Supplementary Note Table 1**), and is later implemented under a fine-tuned logistic regression framework using stochastic gradient descent learning. Performance of the derived models, as measured by precision, recall and global F1-score, reached ~0.9 for cell type classifications at both the high- and low-hierarchy levels (**Fig. 1F,**

Supplementary Note Figures 2 and 3A,B). Notably, representation of a given cell type in the training data was a major determinant of its prediction accuracy (**Supplementary Note Fig. 3C**), implying higher model performance can be achieved with the incorporation of more datasets.

First we applied the high-hierarchy (i.e. low-resolution, 38 cell types) classifier to our cross-tissue dataset and found high consistency when comparing the predicted cell types to our coarse-grained manual annotations (**Fig. 1G**). Among them, the monocytes and macrophages, which often form a transcriptomic continuum in scRNA-seq datasets due to their functional plasticity, were clearly resolved by CellTypist. Furthermore, as the training dataset of CellTypist contained hematopoietic tissues with definitive annotations for progenitor populations, the classifier was able to resolve the manually annotated progenitors into HSC/MPP, neutrophils and monocytes. Thus CellTypist was able to both reinforce and refine our manual annotations.

To allow for automated annotation of fine-grained immune sub-populations, we next applied the low-hierarchy (high-resolution, 93 cell types and subtypes) classifier, which is able to predict cells into more specific subtypes including subsets of T cells, B cells, ILCs and dendritic cells (**Fig. 1G**). This classification highlighted a high degree of heterogeneity within the T cell compartment, not only distinguishing between αβ and γδ T cells, but also dissecting CD4+ and CD8+ T cell subtypes and their more detailed effector and functional phenotypes. Specifically, the CD4+ T cell cluster was classified as helper, regulatory and cytotoxic subsets, and the CD8+ T cell clusters contained unconventional T cell subpopulations such as NKT. Moreover, CellTypist identified a population of germinal center B cells and also revealed three distinct subsets of dendritic cells - DC1, DC2 and migratory DCs (*21*, *22*), again highlighting the granularity CellTypist can achieve.

In summary, we have generated an in-depth map of immune cell populations across human tissues, and developed a framework for automated annotation of immune cell types and subtypes. CellTypist produced expert-grade annotations on our multi-tissue and multi-lineage dataset, and its performance as assessed on multiple metrics was comparable or better relative to other label-transfer methods with minimal computational cost (**Supplementary Note Figures 4,5**). This approach allowed us to substantially refine the description of multiple cell subtypes such as the progenitors and dendritic cell subtypes at full transcriptomic breadth, resolving 26 cell states in total across our dataset. This automated annotation forms the basis for further cross-tissue comparisons of cell compartments in the sections below. CellTypist is publicly available on github and pypi, with user-friendly documentation for application to any single cell transcriptomics data sets.

Tissue-restricted features of mononuclear phagocytes

Mononuclear phagocytes, including monocytes, macrophages and dendritic cells, are critical for immune surveillance and tissue homeostasis. Building on the CellTypist annotation, subclustering of the myeloid subsets revealed further heterogeneity, particularly within macrophages and monocytes (**Fig. 2A-C, fig. S2A**).

Among macrophages, lung-resident cells constituted the majority, and were classified into four major clusters: MAC-I and MAC-II expressing *GPNMB* and *TREM2*, which have been related to alveolar

macrophages(*23*) and disease-associated macrophages (*24*), respectively. MAC-III expressing epithelial markers (potentially due to ambient RNA, doublets or the ingestion of epithelial debris), and MAC-IV with unique expression of *TNIP3* (**Fig. 2B**). TNIP3 (TNFAIP3-interacting protein 3) binds to A20, encoded by TNFAIP3, and inhibits TNF, IL-1 and LPS induced NF-kB activation(*25*). Its expression in lung macrophages may be related to underlying pathology as it was primarily detected in donor A29, a multitrauma donor with lung contusions, and bibasal lung consolidation secondary to aspiration of gastric contents (**fig. S2B**). Consistent with this, among lung macrophages, MAC-IV also expressed the highest levels of S100A8/9, which together form calprotectin, an antimicrobial peptide due to its ability to sequester iron and zinc. Notably, MAC-IV also expressed monocyte-recruiting chemokines CCL2 and CCL3 (**Fig 2B**), providing a means of replenishing the lung macrophage pool.

Beyond these lung macrophages, other macrophage subsets in our data also showed a high degree of tissue restriction (**Fig. 2C**). Red pulp macrophages and Kupffer cells mainly populated the spleen and liver, as expected, and shared high expression of *CD5L*, *SCL40A1* and the transcription factor *SPIC(26)*, with higher expression of *LYVE1* and *ITGA9* distinguishing the Kupffer cells. Notably, a number of macrophages from other tissues, such as the bone marrow and lymph nodes, clustered together with red pulp macrophages and Kupffer cells, pointing to phenotypic similarities across subsets of iron-recycling macrophages (**Fig. 2C**).

We also found a macrophage subtype in the thoracic lymph node which expressed the chitin- and kinin-degrading enzymes *CHIT1* and *CTSK,* respectively (**Fig 2B,C**). Macrophages upregulate *CHIT1* during terminal differentiation, and its expression and activity correlates with lung disease (*27*, *28*). The *CHIT1*-expressing macrophages were restricted to the thoracic lymph node (TLN) in our data, while other publications have reported these in the lung (*29*).

Similar to macrophages, analyses of monocytes demonstrated significant heterogeneity. In addition to three subsets of classical *S100A8*/*9*/*12*+ monocytes, we identified one non-classical monocyte subtype expressing *FCGR3A* and *CX3CR1*, and an intermediate monocyte subtype mainly restricted to the bone marrow with modest *CD14* expression (**Fig. 2A,B**). Among dendritic cells, DC1 expressed *XCR1* and *CLEC9A*, consistent with their identity as conventional/classical DCs (cDC1), specialised for cross-presentation of antigen (**Fig. 2B**). Two *CD1C/IRF4*+ subsets were evident, DC2 and a *CCR7*+ subset. *CCR7* is upregulated in tissue DCs following TLR or FcγR ligation(*30*, *31*), enabling migration towards CCL19/21-expressing lymphatic endothelium and stromal cells in the T cell zone of lymph nodes(*32*, *33*). Consistent with this, we observed a marked enrichment of *CCR7*+ migratory DCs (migDCs) in thoracic and mesenteric lymph nodes, with some cells also observed in the lung (**Fig. 2C**). We validated the presence of migDCs in TLN by immunofluorescence (**fig. S2C**).

Interestingly, migDCs showed upregulation of *AIRE, PDLIM4* and *EBI3* in the TLN, and to a lesser extent in the MLN, while in the lung they upregulate *CRLF2* (encoding TLSPR), chemokines (*CCL22*, *CCL17*), *CSF2RA* and *GPR157* (**Fig. 2D, E**). TLSPR is involved in the induction of Th2 responses in asthma (*34*). Expression of these genes in lung DCs was also observed in our previously published scRNA-seq

datasets(*35*, *36*) (**fig. S2D,E**). These observations suggest that dendritic cell activation coincides with the acquisition of tissue-specific markers that differ depending on the local tissue microenvironment.

Overall, our analysis of the myeloid compartment revealed shared and tissue-restricted features of 18 mononuclear phagocyte cell states, including rare populations of iron-recycling macrophages in mesenteric lymph nodes, chitin-degrading macrophages in thoracic lymph nodes and new subtypes of tissue-specific activated dendritic cells.

B cell subsets and immunoglobulin repertoires across tissues

B cells, comprising progenitors in the bone marrow, developmental states in lymphoid tissues and terminally differentiated memory and plasma cell states in lymphoid and non-lymphoid tissues, play a central role in humoral immunity *via* the production of antibodies that are tailored to specific body sites. Within CellTypist there are four immature and six mature B cell subsets. Using these models and manual curation, we performed an in depth analysis of the B cell compartment, revealing finer subpopulations across tissues, and describing 12 cell states in total (including two likely physiological doublet B cell states) (**Fig. 3A, fig. S3A,B**).

Within the memory B cells, which were characterized by expression of B-cell lineage markers (*MS4A1*, *CD19*) and *TNFRSF13B*, we found a transcriptomically distinct cluster (MemB_ITGAX) positive for *ITGAX*, *TBX21* and *FCRL2*, encoding CD11c, T-bet and the Fc receptor-like protein 2, respectively (**Fig. 3B**). CD11c+T-bet+ B cells, also known as age-associated B cells (ABCs), have been reported in autoimmunity and aging (*37–39*), and likely correspond to our MemB_ITGAX population. Notably, unlike conventional memory B cells, they showed low expression of *CR2* (encoding CD21) and *CD27* (**Fig. 3B**). Interestingly, this subset was mainly present in the blood, liver and bone marrow, while in secondary lymphoid organs, it was primarily found in the spleen (confirmed by flow cytometry and IF **fig. S3C,D**), with only few cells detectable in the thoracic and none in the mesenteric lymph nodes (**Fig. 3C**). This data deepens our understanding of the phenotype of this non-classical subtype of memory B cells, their adhesion properties and tissue distributions.

Within the naive B cell compartment of the blood and spleen, we observed a small population that was distinct from the bulk of naive B cells, with unique expression of *ITGA2B* and *SPARC* (encoding the matricellular protein - secreted protein acidic rich in cysteine - that plays a key role in tissue stroma remodeling) (*40*, *41*) (**Fig. 3B**). In addition, we identified two small populations of germinal center B cells expressing *AICDA* and *BCL6* differing in their proliferative states (marked by *MKI67*). Of note, we did not find any differential expression of dark zone and light zone marker genes for these two populations, probably reflecting limited germinal center activity in the adult donors. Moreover, these germinal center populations were present in lymph nodes and diverse gut regions (**Fig. 3C**), presumably representing Peyer's patches.

We also uncovered two populations of plasmablasts (Plasma_prolif) and plasma cells (Plasma_ITGA8) marked by expression of *CD38*, *XBP1* and *SDC1*. Whereas the former expressed *MKI67* and were

found in the spleen, liver, bone marrow and blood, the latter expressed the integrin alpha-8-encoding gene *ITGA8*, the adhesion molecule *CERCAM* and was detected in the large intestine (CAE and TCL) in addition to the tissues of the Plasma_prolif population (**Fig. 3B,C**). ITGA8+ plasma cells have been recently reported in the context of an analysis of bone marrow plasma cells (*42*), and are likely a long-lived plasma cell population that is resting and tissue-resident. Here we expand their tissue distribution to the liver and gut, and describe their specific clonal distribution pattern below.

B cells have an additional source of variability due to VDJ recombination, somatic hypermutation and class-switching, which can relate to cell phenotype. We performed targeted enrichment and sequencing of BCR transcripts to assess isotypes, hypermutation levels and clonal architecture of the B cell populations described above. This analysis revealed an isotype and subclass usage pattern that related to the cellular phenotypes (**Fig. S4A**). As expected, progenitors and naive B cells mainly showed a subclass preference for IgM and IgD. Interestingly, while memory B cells presented evidence of class switching to IgA1 and IgG1, plasmablasts and plasma cells also showed significant class switching to IgA2 and IgG2.

To determine to what extent this isotype subclass bias correlated with the tissue of origin, we assessed each cell state independently (requiring a minimum cell count of 19 for statistical power). Memory B cells showed a bias towards IgA1 in the omentum and mesenteric lymph node, and towards both IgA1 and IgA2 in the ileum where Peyer's patches are found (**Fig. 3D**). In the plasma cell compartment, we found an even more striking preference for IgA2 across several gut regions (caecum, duodenum and transverse colon), consistent with the known dominance of this isotype at mucosal surfaces (**Fig. 3D**). IgG-expressing plasma cells were also evident in the large intestine, but not in the duodenum, consistent with previous studies (*15*). Of note, plasma cells in the bone marrow, liver and spleen were composed of over 20% IgG2 subclass. With more limited numbers, we report isotype distributions across tissues for other B cell populations (**Fig. S4B,C**): for example, the *ITGAX*+ memory B cells are dominated by IgA subtypes in the thoracic lymph node, and by IgM in the spleen.

Somatic hypermutation (SHM) levels were, as expected, lowest in naive B cells and highest in plasma cells (**Fig. S4D**). Between isotypes and subclasses, SHM did not differ significantly. Nonetheless, there was a tendency towards higher mutation rates in the distal classes IgG2 and IgA2, which are downstream from the IgH locus and can thus accumulate more mutations during sequential switching (*43*) (**Fig. 3E**).

We also explored the occurrence of sequential class switching events in our data by assessing the isotype frequency among expanded clonotypes (>10 cells). Mixed isotype clones are rare in our data, with only a minority detected between IgM and IgD (**Fig. S4E**). We then evaluated the distribution of these expanded clones across tissues and cell types and found three major groups of clones - present in only two tissues, three to four tissues or five or more tissues, respectively (**Fig. 3F**), similar to previously reported patterns of B cell clonal tissue distribution (*44*). While the clones restricted to two tissues, typically between the spleen and the liver or bone marrow, were enriched in plasma cells, those distributed across more than five tissues, including lymph nodes, were over-represented in memory B

cells. Together, these findings suggest that tissue-restricted clones could represent a long-term immunological memory maintained by long-lived plasma cells (Plasma_ITGA8 population) resident in the bone marrow and spleen as well as liver in our data.

Overall, we describe 12 cell states in the B cell compartment, with new in depth characterisation of both naive and memory B cell as well as plasma cell subsets. We identify distinct patterns of clone distributions for the more tissue-restricted long-lived Plasma_ITGA8 cells *versus* broad tissue distribution of classical memory B cell clones.

TCR repertoire analysis reveals patterns of clonal distribution and differentiation across tissues

For annotation of the T cell/innate lymphocyte compartment, as for the B cell compartment, we combined information from CellTypist (27 cell types in this compartment) with in depth manual annotation, which further divided these populations into 15 substates by adding manual inspection to CellTypist scoring (**Fig. 4A**).

Naive/central memory CD4+ T cells were transcriptionally close to naive CD8+ T cells as defined by high expression of *CCR7* and *SELL* (**Fig. 4B**). Other CD4+ T cells identified included: follicular helper T cells (Tfh) expressing *CXCR5*, regulatory T cells (Tregs) expressing *FOXP3* and *CTLA4*, and effector memory CD4+ T cells. Within the effector memory CD8+ T compartment, we found three major subsets characterized by expression of the activation molecule *CRTAM* and the chemokine receptors *CCR9* and *CX3CR1*, respectively**.** The *CCR9*+ population, which also expressed tissue-resident markers *ITGAE* and *ITGA1* (**Fig. 4B**), was enriched in the gut regions, whereas the *CX3CR1+* population was ubiquitous (with the exception of lymph nodes) (**Fig. 4C**). A third CD8+ T cell population expressing *CRTAM* was found in both lymphoid and non-lymphoid tissues. We validated and mapped this population of *CD3D*+*CD8A*+*CRTAM*+ T cells using single-molecule FISH (smFISH) in the liver (**Fig. 4D**) and thoracic lymph nodes (**Fig. 4E**). Furthermore we validated all three effector memory CD8+ T cell populations at the protein level by flow cytometry of cells purified from human spleen (**fig. S5**). Although we could validate CRTAM at the RNA level by smFISH, the protein could not be detected without stimulation, suggesting that CRTAM is subject to post-translational regulation upon TCR activation. These three distinct cytotoxic effector memory T cell populations may represent different states of tissue adaptation and/or maturation between T effector memory and tissue-resident memory states. We also detected unconventional T cell subsets such as MAIT cells, characterised by expression of *TRAV1-2* and *SLC4A10*, and a γδ T cell population overexpressed the integrin molecule *ITGAD* (CD11d) (**Fig. 4B**). NK cells in our data were represented by two clusters with high expression of either *FCGR3A* (encoding CD16) or *NCAM1* (encoding CD56). We defined the ILC3 population, within a small cluster mixed with NK cells, based on markers including *PCDH9* (**Fig. 4A,B**).

Analyses of tissue distributions of these populations revealed that, whereas the majority of CD4+ T and ILC3 cells were located in the lymph nodes and to some extent in the spleen, cytotoxic T and NK cells were more abundant in the bone marrow, spleen and non-lymphoid tissues (**Fig. 4C**). Notably, the

CD4:CD8 ratio in our gut data was lower than previously reported (*45*), which could be related to a lower yield from lamina propria and associated lymphoid tissue due to our pan-tissue dissociation method(*46*).

The analysis of T cell clonal distribution within different tissues of a single individual, and across different individuals is key to understanding T cell-mediated protection. We identified a total of 30,842 cells with productive TCRαβ chains. Chain pairing analysis showed that cells from the T cell clusters mostly contained a single pair of chains (50-60%), with orphan (5-20%) and extra (5-10%) chains present in small fractions of cells (**Fig. S6A,B**). Notably, the frequency of extra α chains (extra VJ) was higher than that of β chains (extra VDJ), a phenomenon potentially due to more stringent and multi-layered allelic exclusion mechanisms at the TCRβ locus compared to TCRα (*47*). As expected, the NK and ILC clusters held no productive TCR chains. Within the γδ T cell cluster, only a small proportion had a productive TCR chain, which may result from cytotoxic T cells co-clustering with γδ T cells.

We next examined V(D)J gene usage in relation to T cell identity. In the MAIT population, we detected a significant enrichment of *TRAV1-2,* as expected (**Fig. 4F, Fig. S6C**), and revealed an exciting tissue-specific distribution of *TRAJ33* (spleen) and *TRAJ12* (liver) (**Fig. 4F**). Specifically, V(D)J usage bias analysis across tissues revealed a significant enrichment of *TRAJ12, TRBV6-4* and *TRBJ2-1* in liver MAIT cells *versus TRAJ33* and *TRBJ2-6* in splenic MAIT cells (**Fig. 4F**). This suggests that there may be different antigens for MAIT cells in the spleen *versus* liver corresponding to the different metabolomes in these tissues. In addition, full analysis of the TCR repertoire of the MAIT cells revealed previously unappreciated diversity of V segment usage in the beta chain (**Fig. S6D**).

We then defined clonally related cells based on identical CDR3 nucleotide sequences to investigate their TCR repertoires. Using this approach, we found that clonally expanded cells were primarily from the memory CD8+ T cell compartment, and the MAIT populations mentioned above, with fewer small CD4+ T cell clones (**fig. S6E**). These clonotypes were restricted to single individuals as expected (**fig. S6F**), and within an individual they were distributed across tissues and subsets (**fig. S6G,H**). We found a very small number of isolated CD4+ T cell clones that shared Treg and effector T cell phenotypes, illustrating the possibility but low frequency of plasticity or (trans)differentiation from the same naive precursor cell in the periphery (**fig. S6H**).

Focusing on the most expanded clonotypes (>20 cells), the majority were widespread across five or more tissues supporting the systemic nature of tissue-resident immune memory (**Fig. 4G**). Moreover, we found that several *CCR9*+ CD8+ T cell-enriched clonotypes were shared across different gut regions, and that several clonotypes present across both the liver and lung consisted of a mixture of cells of the *CRTAM*+ and *CX3CR1*+ CD8+ T cell populations. This suggests that a single naive CD8+ T cell precursor can give rise to diverse cytotoxic T cell states, which harbour immune memory across multiple non-lymphoid tissues, again emphasizing plasticity of phenotype and location within a clone.

In summary, we describe 15 T/innate cell states in our data by integrating three sources of information: CellTypist logistic regression models, V(D)J sequencing data and manual inspection. This yields new insights into the MAIT cell compartment and its antigen receptor repertoire distribution that is markedly different in spleen *versus* liver. For CD4+ T cells, we find rare sharing of Tregs and effector memory

cells within the same clone. We show that for the cytotoxic T cell memory compartment there is broad sharing of clones across gut regions for *CCR9*+ CD8+ T cell-enriched clonotypes, and mixed *CRTAM*+ and *CX3CR1*+ CD8+ T cell clonotypes with broad tissue distributions.

Two distinct subsets of γδ T cells across human tissues

γδ T cells are at the interface between adaptive and innate immune functions, and have recently attracted attention due to their potential for cell-based therapies. Accurate identification of γδ T cells in single-cell transcriptomics studies remains challenging due to their transcriptional similarity to cytotoxic αβ T cells and NK cells. Here we manually annotated a γδ T cell cluster expressing *ITGAD* which encodes CD11d, part of a CD11d/CD18 heterodimer with the potential to interact with the cell adhesion molecule VCAM1 (*48*). In the locations where γδ T cells were identified (liver, spleen and bone marrow), resident macrophages, such as red pulp macrophages, expressed high levels of *VCAM1* (**Fig. 2B, Fig. 5A**), indicating a potential interaction between these cells. Due to the lack of a robust commercially available anti-CD11d antibody, we validated this population by flow sorting and qPCR on sorted CD3+TCRγδ+ and CD3+TCRαβ+ cells from cryopreserved spleen samples from three donors (**Fig.S7**). As a small fraction of αβ T cells, marked by low expression of CD52 and CD127, were noted to express *ITGAD*, the CD3+TCRαβ population was split into CD52-CD127- and CD52+CD127+ subpopulations. In keeping with our scRNAseq data, *ITGAD* expression was high in CD3+TCRγδ and CD52-CD127- CD3+TCRαβ (**Fig. 5B**), providing additional evidence for the specific expression of this integrin alpha subunit in this subpopulation of γδ T cells. In order to further elucidate the identity of this interesting lymphocyte compartment, we carried out γδ TCR sequencing in selected spleen, bone marrow and liver samples. The γδ TCR sequencing data was subjected to a customized analysis pipeline we developed and optimised on the basis of the use of Cell Ranger followed by contig re-annotation with dandelion (see Methods), facilitating the full recovery of γδ chains in our data. This analysis confirmed that the majority of productive γδ TCR chains originated from the *ITGAD*-expressing γδ T cells (**Fig. 5C**, **fig. S8A**), further supporting the robust identification of this population. Overall, these multiple data support a γδ T cell population in the liver, spleen and bone marrow that may play a role in tissue homeostasis *via* interaction with tissue-resident macrophage populations.

γδ T cells in the blood have been shown to comprise two transcriptomically distinct subsets according to their T cell receptor V gene: TCRVδ1 or TCRVδ2(*49*). Our unique data allowed us to explore in detail the characteristics of these two subsets. Specifically, analyses of γδ TCR chains revealed a further subdivision of cells with productive single-pair γδ chains into the TRDV1 and TRDV2 subclusters, consistent with their separation trend in the UMAP representation (**Fig. 5D**). Comparisons of the two subpopulations with the blood TCRVδ1 and TCRVδ2 showed that they shared common markers for TCRVδ1 (*CCL3*, *FCGR3A*, *CMC1*, *CD63*) and TCRVδ2 (*IL7R*, *LTB*, *SPOCK2*, *NFKBIA*). Moreover, in comparison to blood, γδ T cells derived from the spleen, liver and bone marrow were uniquely overrepresented by specific markers, such as expression of *TOX* and *TRGV4* in the TRDV1 subcluster, as well as *CD55* and *TRGV9* in the TRDV2 subcluster (**Fig. 5D**).

In addition to *ITGAD*-expressing γδ T cells, CellTypist predicted a different group of cells (sitting within the *CCR9*+ CD8+ T cell cluster) as γδ T cells (**Fig. 5E, fig. S8B**). This CellTypist-annotated γδ T cell population was primarily derived from the gut and had high expression of the gut homing receptor CCR9, as well as several integrin-related genes including *ITGAE* and *ITGA1* (**Fig. 5F**).

Altogether, dissection of these different tissue-resident γδ T cell populations in human tissues for the first time showcases how a combination of CellTypist-based automated annotation, expert-driven cluster analysis and TCR sequencing can synergize to dissect specific and functionally relevant aspects of cell identities.

Discussion

Here, we present the first multi-donor/multi-tissue study of immune cells across the human body. By sampling multiple organs from the same individuals, which allows for robust control of age, gender, medical history, drug exposure and sampling backgrounds, we have been able to reveal novel tissue-specific expression patterns across the myeloid and lymphoid compartments.

Accurately assigning detailed cell identities in scRNA-seq experiments is a major challenge. To overcome this, we have developed CellTypist, a framework for automated immune cell type annotation. In addition to identifying major cell types, CellTypist is able to perform fine-grained cell subtype annotation that is typically very time-consuming and requires expert knowledge. This has largely been made possible due to the curation of a large collection of studies across a range of tissues with in-depth immune cell analysis comprising 93 harmonized cell type labels. Currently, manual curation still has a role to play in revealing specific cell subtypes that may be absent from the database/training set, and we showcase such an example in the γδ T cell compartment. To address this, in the longer term, the CellTypist models will be periodically updated and extended to include further immune and non-immune sub-populations, as more data become available.

Using CellTypist combined with manual annotation, we have dissected the transcriptomic features of 49 immune cell types and states across tissues. Within the myeloid compartment, macrophages showed most prominent features of tissue specificificity. Red pulp macrophages and Kupffer cells, shared features related to iron-recycling(*50*) with macrophages in other locations such as the bone marrow and mesenteric lymph nodes, suggesting that macrophages participate in iron metabolism across a wide range of tissues. Among lung macrophages, we found a subtype with high TNIP3 and calprotectin expression. Elevated serum calprotectin has been proposed as a biomarker of bacterial pneumonia(*51*), predicts pulmonary exacerbations and lung function decline in cystic fibrosis(*52*), and an S100A8/9 high macrophage subset has been identified in COVID19 bronchiolar lavage fluid(*53*). Lastly, we characterized tissue-specific features of migratory dendritic cells (CCR7+) revealing specific expression of *CRLF2, CSF2RA* and *GPR157* in the lung, and expression of *AIRE* in the thoracic lymph nodes. These novel migratory dendritic cell states are interesting targets for future in depth functional characterisation in the context of allergy, asthma and other related pathologies (*54*, *55*).

In the lymphoid compartment, we combined single-cell transcriptome and VDJ analysis, which allowed the phenotypic dissection of adaptive immune cells using complementary layers of single cell genomics data. Of note, we detected a subset of memory B cells expressing *ITGAX* (CD11c) and *TBX21* (T-bet) that resemble age-associated B cells (ABCs) previously reported to be expanded in ageing(*38*), following malaria vaccination (*56*) and in systemic lupus erythematosus (SLE) patients (*57*). In our data, these B cells did not show clonal expansion, suggesting that they may be present at low levels in healthy tissues and expand upon challenge as well as ageing. BCR analysis revealed isotype usage biased towards IgA2 in gut plasma cells, which may be related to structural differences (*58*) or higher resistance to microbial degradation as compared to IgA1 (*59*). Cross-tissue clonal distribution of B cells and plasma cells has been previously shown to follow two different patterns (*44*). In our study, we primarily detected shared clones between the spleen and lymphoid organs but were also able to detect a degree of sharing with non-lymphoid tissues including the gut. Notably, when we incorporated cell identities into the clonotype analysis, we observed that broadly distributed clones were geared towards the B cell memory phenotype while tissue-restricted clones were mainly the plasma cells. The plasma cells are therefore likely long-lived plasma cells contributing to immune memory *via* more restricted tissue localisation as compared to memory B cells.

In the T-cell compartment, cross-tissue and cross-cell type TCR sharing revealed previously unappreciated insights into the plasticity and distribution of T cell subtypes. Sharing between *CRTAM*+ and *CX3CR1*+ subtypes of CD8+ T cells supports the possibility that these transcriptomically distinct populations may represent different stages of migration or tissue adaptation, such as the differentiation of effector memory T cells (TEM) to resident memory T cells (TRM) (*60*). Lastly, using a combination of automated annotation, manual curation and γδ TCR sequencing, we found several distinct TRDV1 and TRDV2 subsets of γδ T cells showing distinct integrin gene expression and tissue distributions.

In summary, this study reveals previously unrecognized features of tissue-specific immunity in the myeloid and lymphoid compartments, and provides a comprehensive framework for future cross-tissue cell type analysis. Further investigation of human tissue-resident immunity is needed to determine the effect of important covariates such as underlying critical illness, donor age and gender as well as considering the immune cell activation status, to gain a defining picture of how human biology influences immune functions. Applications of a deep characterisation of the features of immune cells across human tissues include implications for engineering of cells for therapeutic purposes and addressing cells to the intended tissue locations, for understanding tissue-specific features of infection as well as the distinct modes of vaccine delivery to tissues.

Figures and figure legends

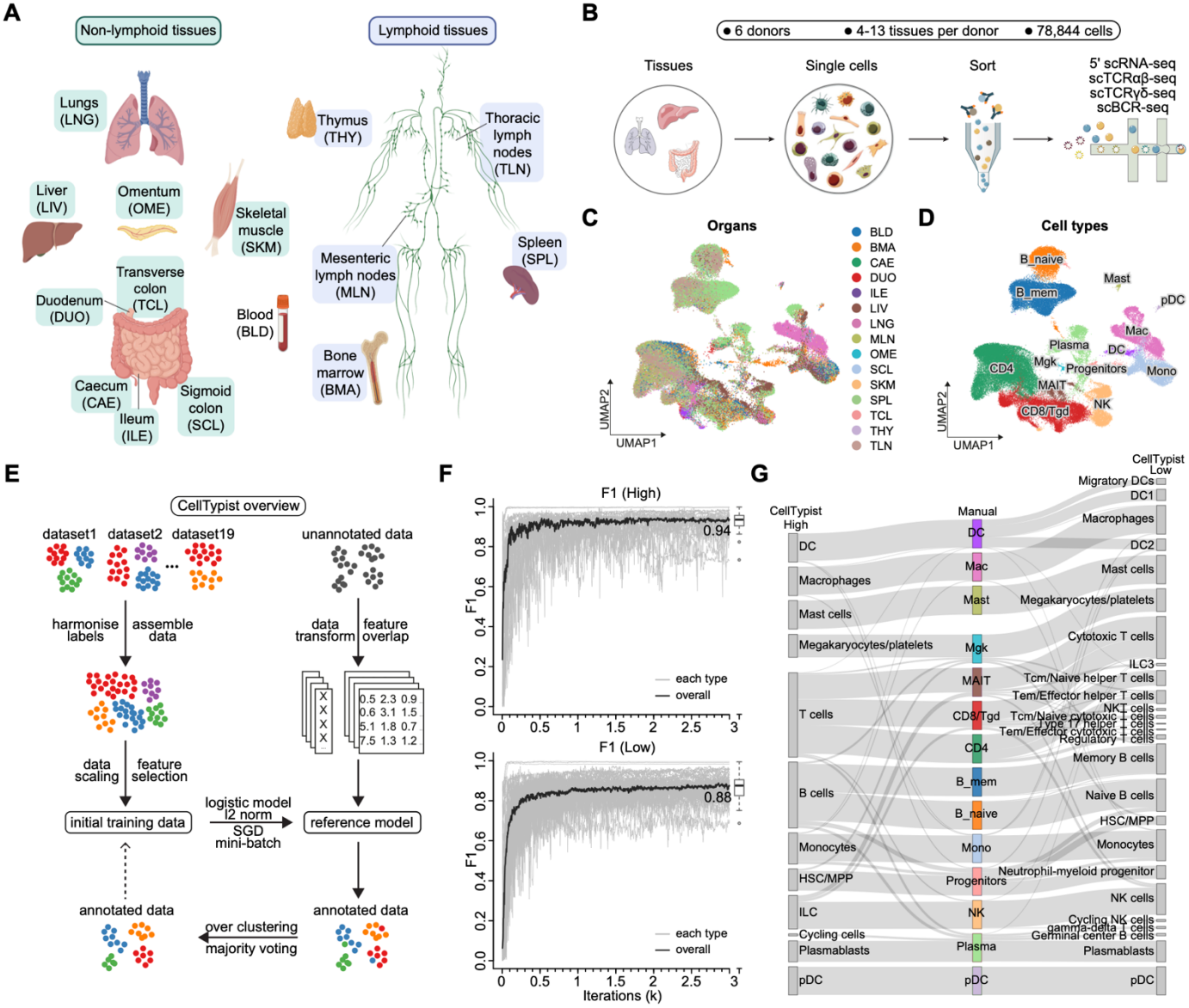

Fig. 1. Automated annotation of immune cells across human tissues using CellTypist. (**A**) Schematic showing sample collections of human lymphoid and non-lymphoid tissues and their assigned tissue name acronyms. (**B**) Schematic of single-cell transcriptome profiling and paired sequencing of αβ TCR, γδ TCR and BCR variable regions. (**C**) UMAP visualization of the immune cell compartment colored by tissues. (**D**) As in (**C**), but colored by manually annotated cell types. (**E**) Workflow of CellTypist including data collection, processing, model training and cell type prediction. (**F**) Performance curves showing the F1 score at each iteration of training using mini-batch stochastic gradient descent for high- and low-hierarchy CellTypist models, respectively. The black curve represents the median F1 score averaged across the individual F1 scores of all predicted cell types (grey curves). (**G**) Sankey plot showing the fractions of CellTypist high-hierarchy model-derived labels (left) as compared to the manually defined clusters (center), as well as the fractions of CellTypist low-hierarchy model-derived labels (right).

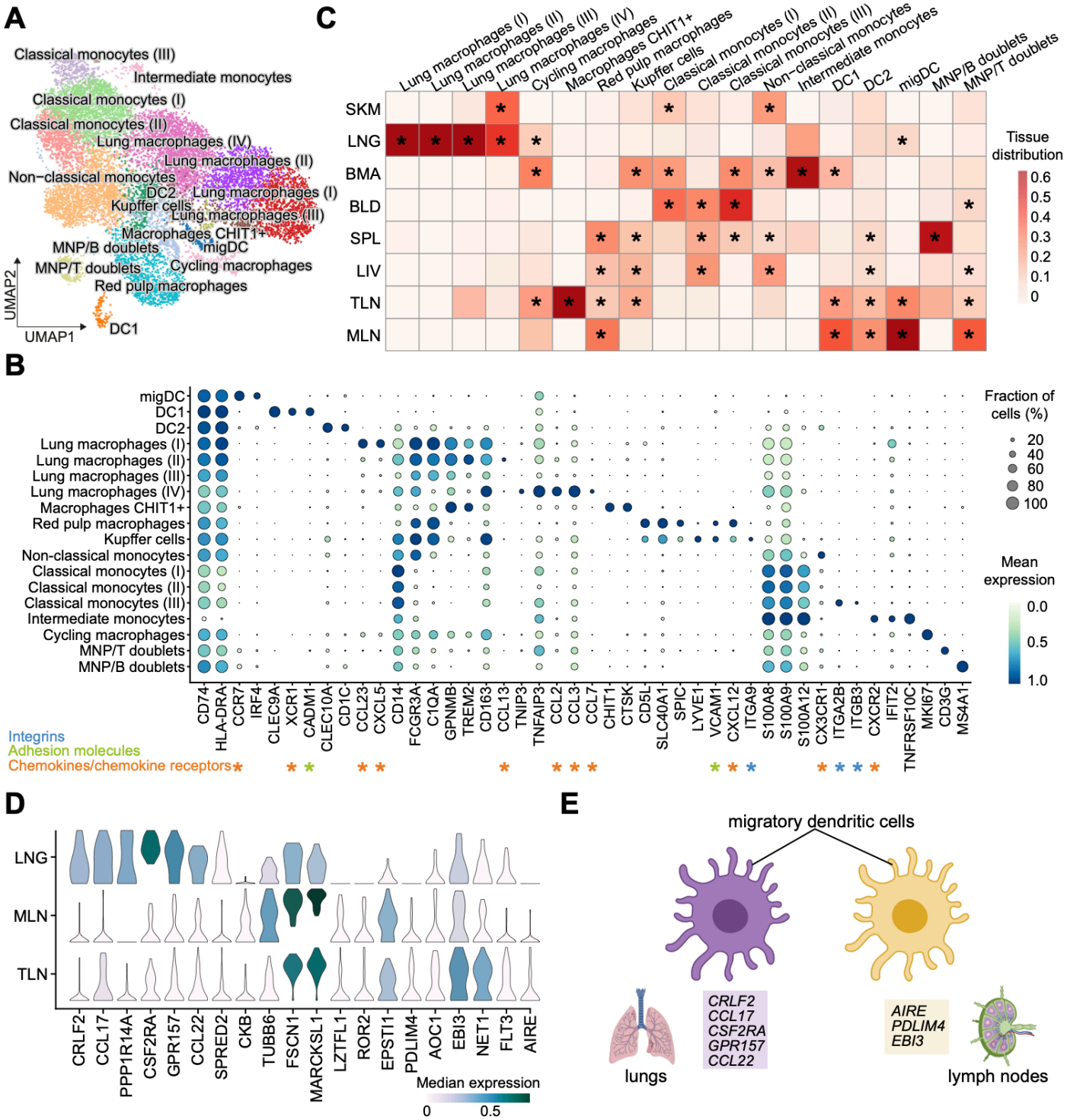
A
Classical monocytes (III)
Intermediate monocytes
Classical monocytes (I)
Classical monocytes (II)
Lung macrophages (IV)
Lung macrophages (II)
Non−classical monocytes
DC2
Lung macrophages (I)
Kupffer cells
Lung macrophages (III)
Macrophages CHIT1+
MNP/B doublets
migDC
MNP/T doublets
Cycling macrophages
Red pulp macrophages
DC1
UMAP2
UMAP1
C
Lung macrophages (I)
Lung macrophages (II)
Lung macrophages (III)
Lung macrophages (IV)
Cycling macrophages
Macrophages CHIT1+
Red pulp macrophages
Kupffer cells
Classical monocytes (I)
Classical monocytes (II)
Classical monocytes (III)
Non−classical monocytes
Intermediate monocytes
DC1
DC2
migDC
MNP/B doublets
MNP/T doublets
SKM
LNG
BMA
BLD
SPL
LIV
TLN
MLN
Tissue distribution
0.6
0.5
0.4
0.3
0.2
0.1
0
B
migDC
DC1
DC2
Lung macrophages (I)
Lung macrophages (II)
Lung macrophages (III)
Lung macrophages (IV)
Macrophages CHIT1+
Red pulp macrophages
Kupffer cells
Non-classical monocytes
Classical monocytes (I)
Classical monocytes (II)
Classical monocytes (III)
Intermediate monocytes
Cycling macrophages
MNP/T doublets
MNP/B doublets
CD74
HLA-DRA
CCR7
IRF4
CLEC9A
XCR1
CADM1
CLEC10A
CD1C
CCL23
CXCL5
CD14
FCGR3A
C1QA
GPNMB
TREM2
CD163
CCL13
TNIP3
TNFAIP3
CCL2
CCL3
CCL7
CHIT1
CTSK
CD5L
SLC40A1
SPIC
LYVE1
VCAM1
CXCL12
ITGA9
S100A8
S100A9
S100A12
CX3CR1
ITGA2B
ITGB3
CXCR2
IFIT2
TNFRSF10C
MKI67
CD3G
MS4A1
Fraction of cells (%)
20
40
60
80
100
Mean expression
0.0
0.5
1.0
Integrins
Adhesion molecules
Chemokines/chemokine receptors
D
LNG
MLN
TLN
CRLF2
CCL17
PPP1R14A
CSF2RA
GPR157
CCL22
SPRED2
CKB
TUBB6
FSCN1
MARCKSL1
LZTFL1
ROR2
EPSTI1
PDLIM4
AOC1
EBI3
NET1
FLT3
AIRE
Median expression
0
0.5
E
migratory dendritic cells
CRLF2
CCL17
CSF2RA
GPR157
CCL22
AIRE
PDLIM4
EBI3
lungs
lymph nodes

Fig. 2. Myeloid compartment across tissues. (**A**) UMAP visualization of the cell populations in the myeloid compartment. (**B**) Dot plot for expression of marker genes of the identified myeloid populations. Color represents maximum-normalized mean expression of cells expressing marker genes, and size represents the percentage of cells expressing these genes. (**C**) Heatmap showing the distribution of each myeloid cell population across different tissues. Cell numbers are normalised with each tissue and later calculated as proportions across tissues. Asterisks mark significant enrichment in a given tissue relative to the remaining tissues (poisson regression stratified by donors, $p < 0.05$). (**D**) Violin plot for genes differentially expressed in migratory dendritic cells across tissues. Color represents maximum-normalized mean expression of cells expressing marker genes. (**E**) Schematic illustration of tissue-specific features of migratory dendritic cells.

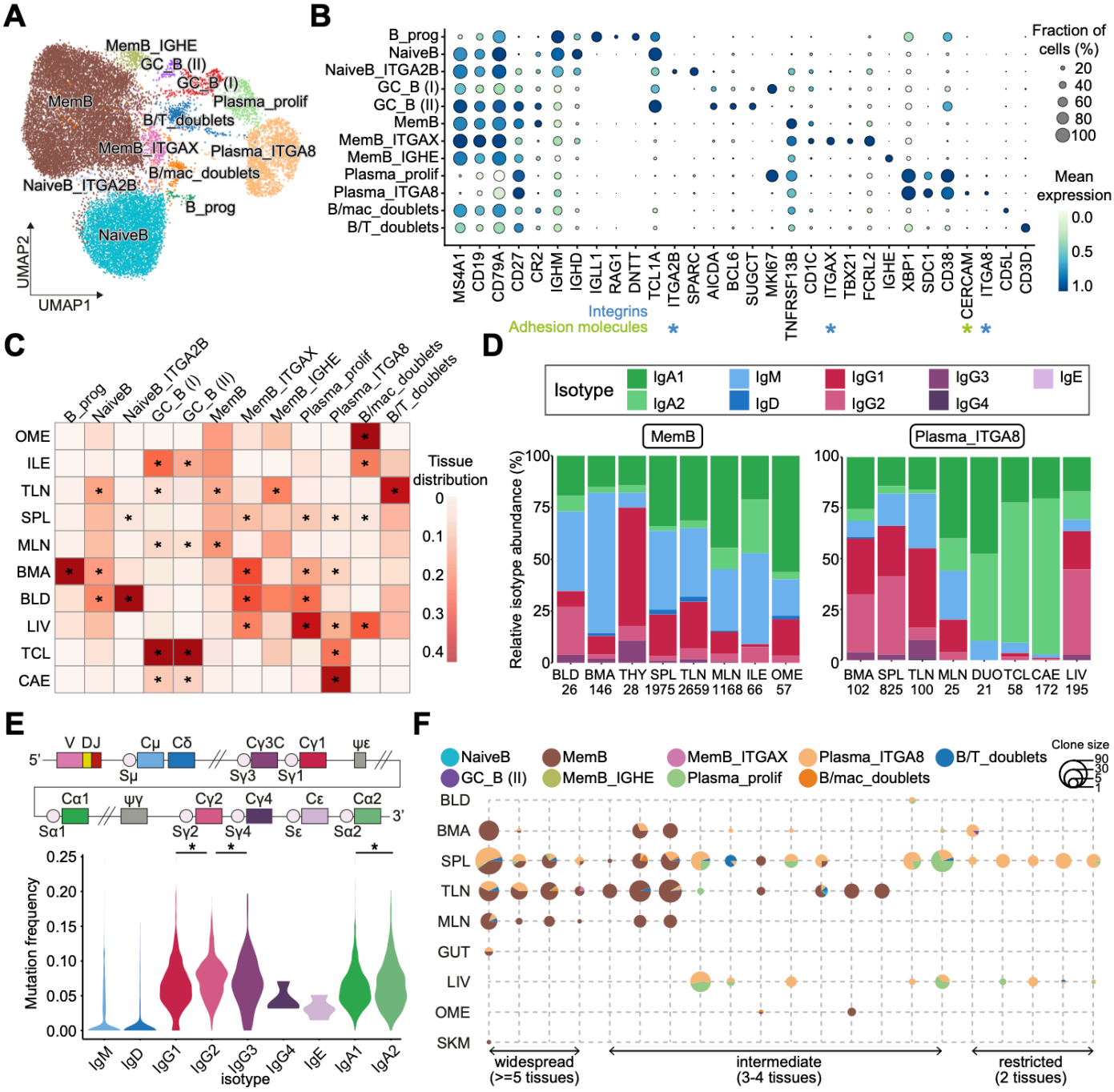
A
MemB_IGHE
GC_B (II)
GC_B (I)
MemB
Plasma_prolif
B/T_doublets
MemB_ITGAX
Plasma_ITGA8
B/mac_doublets
NaiveB_ITGA2B
B_prog
NaiveB
UMAP2
UMAP1
B
B_prog
NaiveB
NaiveB_ITGA2B
GC_B (I)
GC_B (II)
MemB
MemB_ITGAX
MemB_IGHE
Plasma_prolif
Plasma_ITGA8
B/mac_doublets
B/T_doublets
MS4A1 CD19 CD79A CD27 CR2 IGHM IGHD IGLL1 RAG1 DNTT TCL1A ITGA2B SPARC AICDA BCL6 SUGCT MKI67 TNFRSF13B CD1C ITGAX TBX21 FCRL2 IGHE XBP1 SDC1 CD38 CERCAM ITGA8 CD5L CD3D
Integrins
Adhesion molecules
Fraction of cells (%)
20
40
60
80
100
Mean expression
0.0
0.5
1.0
C
B_prog NaiveB NaiveB_ITGA2B GC_B (I) GC_B (II) MemB MemB_ITGAX MemB_IGHE Plasma_prolif Plasma_ITGA8 B/mac_doublets B/T_doublets
OME
ILE
TLN
SPL
MLN
BMA
BLD
LIV
TCL
CAE
Tissue distribution
0
0.1
0.2
0.3
0.4
D
Isotype
IgA1
IgA2
IgM
IgD
IgG1
IgG2
IgG3
IgG4
IgE
MemB
Plasma_ITGA8
Relative isotype abundance (%)
100
75
50
25
0
BLD BMA THY SPL TLN MLN ILE OME
26 146 28 1975 2659 1168 66 57
BMA SPL TLN MLN DUO TCL CAE LIV
102 825 100 25 21 58 172 195
E
V DJ Cμ Cδ Cγ3C Cγ1 ψε
5'
Sμ Sγ3 Sγ1
Cα1 ψγ Cγ2 Cγ4 Cε Cα2
3'
Sα1 Sγ2 Sγ4 Sε Sα2
Mutation frequency
0.25
0.20
0.15
0.10
0.05
0.00
IgM IgD IgG1 IgG2 IgG3 IgG4 IgE IgA1 IgA2
isotype
F
NaiveB
GC_B (II)
MemB
MemB_IGHE
MemB_ITGAX
Plasma_prolif
Plasma_ITGA8
B/mac_doublets
B/T_doublets
Clone size
90
30
5
1
BLD
BMA
SPL
TLN
MLN
GUT
LIV
OME
SKM
widespread (>=5 tissues)
intermediate (3-4 tissues)
restricted (2 tissues)

Fig. 3. B cell compartment across tissues. (**A**) UMAP visualization of the cell populations in the B cell compartment. (**B**) Dot plot for expression of marker genes of the identified B cell populations. Color represents maximum-normalized mean expression of cells expressing marker genes, and size represents the percentage of cells expressing these genes. (**C**) Heatmap showing the distribution of each B cell population across different tissues. Cell numbers are normalised with each tissue and later calculated as proportions across tissues. Asterisks mark significant enrichment in a given tissue relative to the remaining tissues (poisson regression stratified by donors, $p < 0.05$). (**D**) Stacked bar plots showing the isotype distribution per tissue within the memory B cell cluster and the plasma cell cluster. (**E**) Violin plot of hypermutation frequency on the IgH chain across isotypes. Significant difference between IgG2 and IgG1/3, as well as between IgA2 and IgA1 is marked by asterisks (wilcoxon rank sum test, $p < 0.05$). (**F**) Scatterpie plot showing the tissue distribution and B cell subsets of expanded clonotypes (>10 cells). Each vertical line represents one clonotype. Clonotypes are grouped based on their tissue distributions.

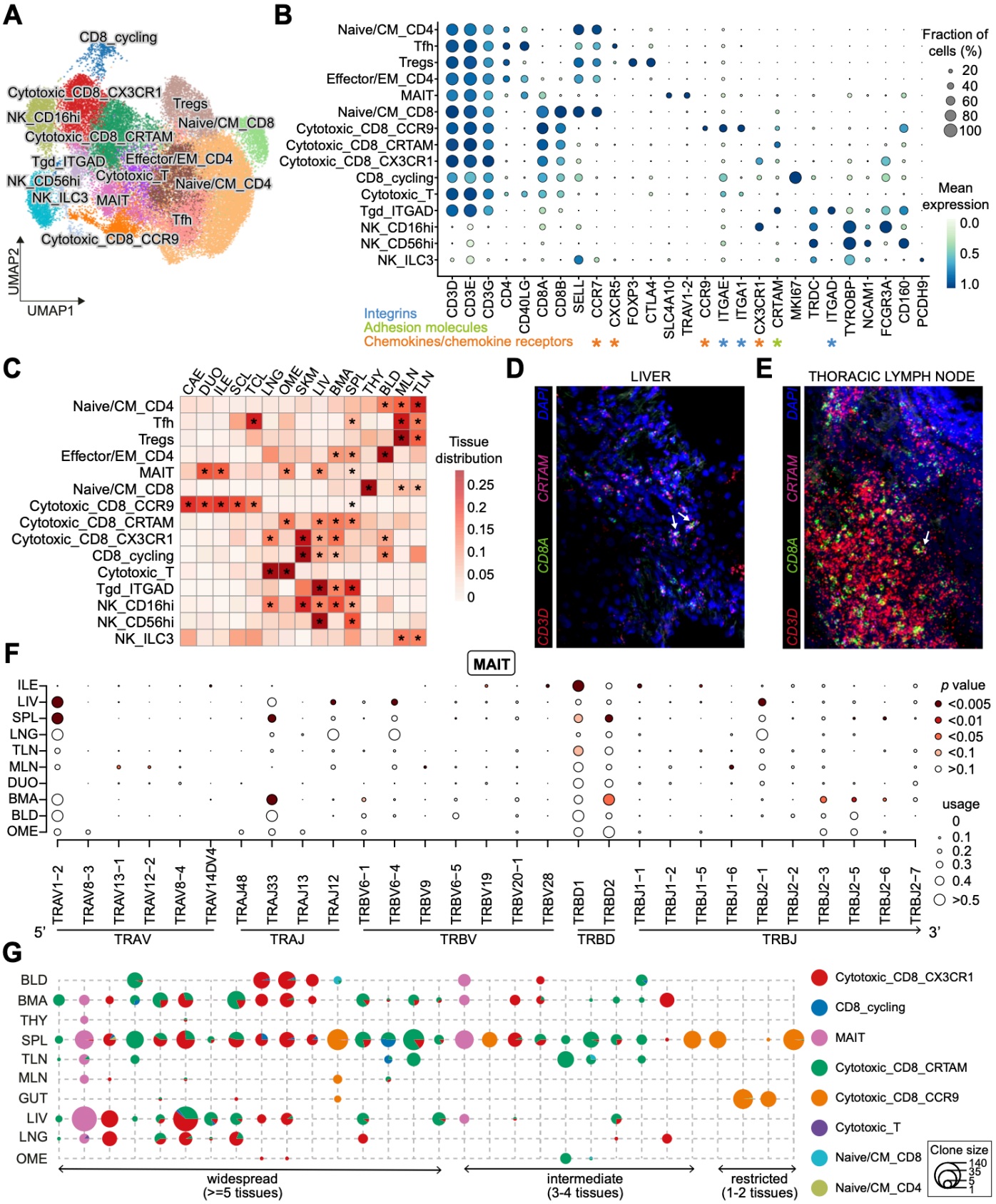
A
CD8_cycling
Cytotoxic_CD8_CX3CR1
Tregs
NK_CD16hi
Naive/CM_CD8
Cytotoxic_CD8_CRTAM
Tgd_ITGAD
Effector/EM_CD4
NK_CD56hi
Cytotoxic_T
Naive/CM_CD4
NK_ILC3
MAIT
Tfh
Cytotoxic_CD8_CCR9
UMAP2
UMAP1
B
Naive/CM_CD4
Tfh
Tregs
Effector/EM_CD4
MAIT
Naive/CM_CD8
Cytotoxic_CD8_CCR9
Cytotoxic_CD8_CRTAM
Cytotoxic_CD8_CX3CR1
CD8_cycling
Cytotoxic_T
Tgd_ITGAD
NK_CD16hi
NK_CD56hi
NK_ILC3
CD3D CD3E CD3G CD4 CD40LG CD8A CD8B SELL CCR7 CXCR5 FOXP3 CTLA4 SLC4A10 TRAV1-2 CCR9 ITGAE ITGA1 CX3CR1 CRTAM MKI67 TRDC ITGAD TYROBP NCAM1 FCGR3A CD160 PCDH9
Integrins
Adhesion molecules
Chemokines/chemokine receptors
Fraction of cells (%)
20
40
60
80
100
Mean expression
0.0
0.5
1.0
C
CAE DUO ILE SCL TCL LNG OME SKM LIV BMA SPL THY BLD MLN TLN
Naive/CM_CD4
Tfh
Tregs
Effector/EM_CD4
MAIT
Naive/CM_CD8
Cytotoxic_CD8_CCR9
Cytotoxic_CD8_CRTAM
Cytotoxic_CD8_CX3CR1
CD8_cycling
Cytotoxic_T
Tgd_ITGAD
NK_CD16hi
NK_CD56hi
NK_ILC3
Tissue distribution
0.25
0.2
0.15
0.1
0.05
0
D
LIVER
DAPI
CRTAM
CD8A
CD3D
E
THORACIC LYMPH NODE
DAPI
CRTAM
CD8A
CD3D
F
MAIT
ILE
LIV
SPL
LNG
TLN
MLN
DUO
BMA
BLD
OME
TRAV1-2 TRAV8-3 TRAV13-1 TRAV12-2 TRAV8-4 TRAV14DV4 TRAJ48 TRAJ33 TRAJ13 TRAJ12 TRBV6-1 TRBV6-4 TRBV9 TRBV6-5 TRBV19 TRBV20-1 TRBV28 TRBD1 TRBD2 TRBJ1-1 TRBJ1-2 TRBJ1-5 TRBJ1-6 TRBJ2-1 TRBJ2-2 TRBJ2-3 TRBJ2-5 TRBJ2-6 TRBJ2-7
5'
TRAV
TRAJ
TRBV
TRBD
TRBJ
3'
p value
<0.005
<0.01
<0.05
<0.1
>0.1
usage
0
0.1
0.2
0.3
0.4
>0.5
G
BLD
BMA
THY
SPL
TLN
MLN
GUT
LIV
LNG
OME
widespread (>=5 tissues)
intermediate (3-4 tissues)
restricted (1-2 tissues)
Cytotoxic_CD8_CX3CR1
CD8_cycling
MAIT
Cytotoxic_CD8_CRTAM
Cytotoxic_CD8_CCR9
Cytotoxic_T
Naive/CM_CD8
Naive/CM_CD4
Clone size
140
35
5
1

Fig. 4. TCR repertoire analysis reveals patterns of clonal distribution and differentiation across tissues. (**A**) UMAP visualization of T cells and innate lymphoid cells (ILCs) across human tissues colored by cell types. (**B**) Dot plot for expression of marker genes of the identified immune populations. Color represents maximum-normalized mean expression of cells expressing marker genes, and size represents the percentage of cells expressing these genes. (**C**) Heatmap showing the distribution of each T cell or ILC population across the different tissues. Cell numbers are normalised with each tissue and later calculated as proportions across tissues. Asterisks mark significant enrichment in a given tissue relative to the remaining tissues (poisson regression stratified by donors, $p < 0.05$). (**D**,**E**) smFISH visualisation of *CD3D, CD8A* and *CRTAM* transcripts, validating this tissue-resident memory CD8+ T cell population in the liver and thoracic lymph node. (**F**) Dot plot denoting the TRA and TRB V(D)J usage across tissues for MAIT cells. Only gene segments with a usage of greater than 10% in at least one tissue are included in the plots, with the sizes of dots indicating the gene segment usage and the colors indicating the significance of the difference between a given tissue and the remaining tissues. Significance is assessed by the generalized linear model stratified by donors with a Poisson structure. (**G**) Scatterpie plot showing the tissue distribution and T cell subsets of expanded clonotypes (>20 cells). Each vertical line represents one clonotype. Clonotypes are grouped based on their tissue distributions.

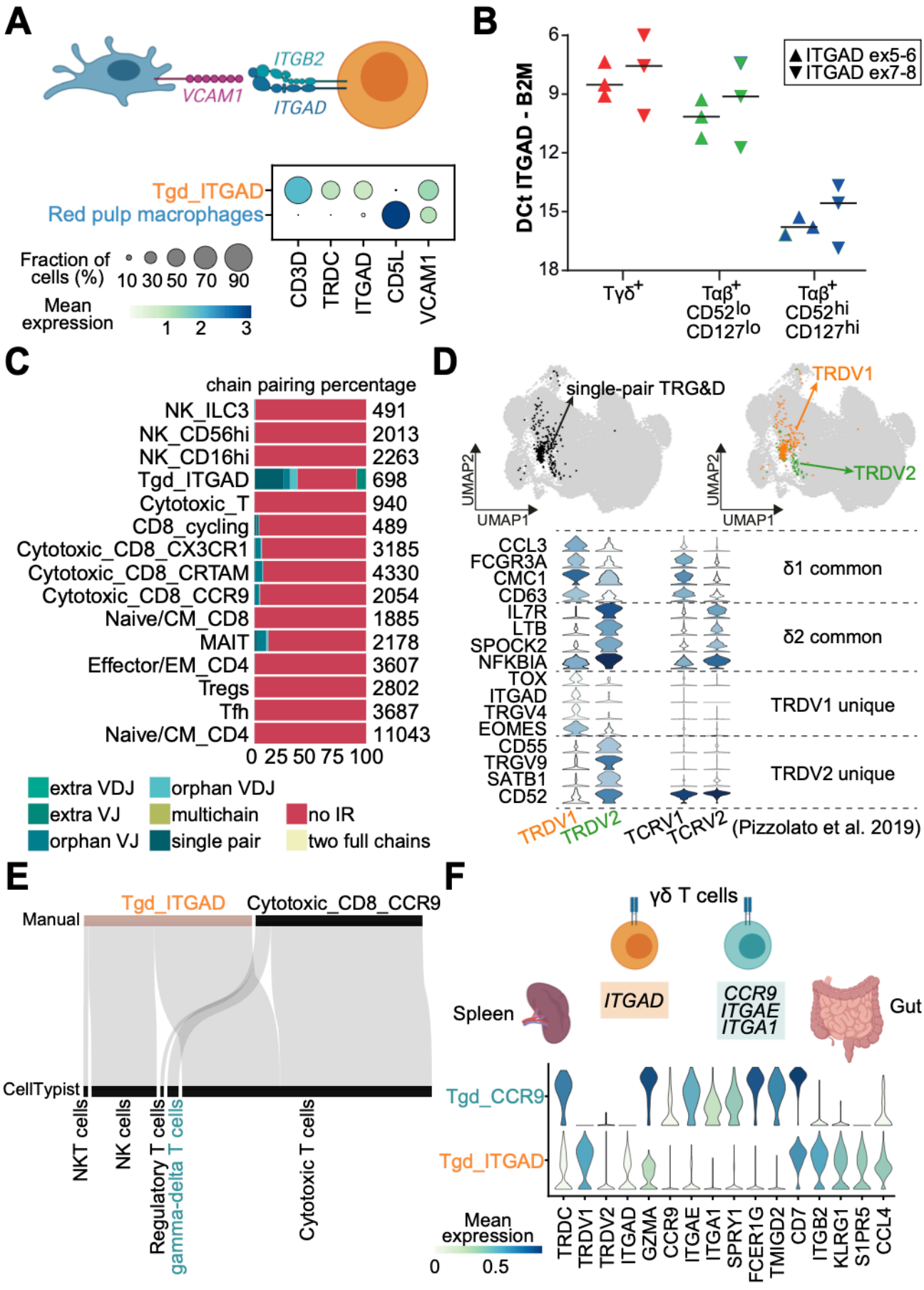
A
ITGB2
VCAM1
ITGAD
Tgd_ITGAD
Red pulp macrophages
Fraction of cells (%)
10 30 50 70 90
Mean expression
1 2 3
CD3D
TRDC
ITGAD
CD5L
VCAM1
B
DCt ITGAD - B2M
6
9
12
15
18
▲ITGAD ex5-6
▼ITGAD ex7-8
Tγδ+
Tαβ+ CD52lo CD127lo
Tαβ+ CD52hi CD127hi
C
chain pairing percentage
NK_ILC3 491
NK_CD56hi 2013
NK_CD16hi 2263
Tgd_ITGAD 698
Cytotoxic_T 940
CD8_cycling 489
Cytotoxic_CD8_CX3CR1 3185
Cytotoxic_CD8_CRTAM 4330
Cytotoxic_CD8_CCR9 2054
Naive/CM_CD8 1885
MAIT 2178
Effector/EM_CD4 3607
Tregs 2802
Tfh 3687
Naive/CM_CD4 11043
0 25 50 75 100
extra VDJ
orphan VDJ
extra VJ
multichain
no IR
orphan VJ
single pair
two full chains
D
single-pair TRG&D
TRDV1
TRDV2
UMAP2
UMAP1
CCL3
FCGR3A
CMC1
CD63
δ1 common
IL7R
LTB
SPOCK2
NFKBIA
δ2 common
TOX
ITGAD
TRGV4
EOMES
TRDV1 unique
CD55
TRGV9
SATB1
CD52
TRDV2 unique
TRDV1
TRDV2
TCRV1
TCRV2
(Pizzolato et al. 2019)
E
Tgd_ITGAD
Cytotoxic_CD8_CCR9
Manual
CellTypist
NKT cells
NK cells
Regulatory T cells
gamma-delta T cells
Cytotoxic T cells
F
γδ T cells
Spleen
ITGAD
CCR9
ITGAE
ITGA1
Gut
Tgd_CCR9
Tgd_ITGAD
Mean expression
0 0.5
TRDC
TRDV1
TRDV2
ITGAD
GZMA
CCR9
ITGAE
ITGA1
SPRY1
FCER1G
TMIGD2
CD7
ITGB2
KLRG1
S1PR5
CCL4

Fig. 5. Two distinct subsets of γδ T cells across human tissues. (**A**) Dot plot showing the expression of *VCAM1* by *CD5L+* macrophages (marking red pulp macrophages) and *TRDC* and *ITGAD* transcripts by γδ T cells in the spleen. (**B**) qPCR validation of *ITGAD* in T cell populations as measured by two assays to *ITGAD* and normalised to the housekeeping gene *B2M*. (**C**) Stacked bar plot depicting the distributions of γδ TCR chain pairing types across cell populations. No IR = No immune receptor. (**D**) UMAP plots showing the cells with single-pair γδ TCR chains (upper left) and the subdivision of these cells into the TRDV1 and TRDV2 clusters (upper right). Bottom violin plots demonstrate the expression of common and unique marker genes of the two clusters as compared to the γδ T subsets from Pizzolato et al. (**E**) Sankey plot of the *CCR9*-expressing γδ T cells identified using CellTypist. (**F**) Stacked violin plot for expression of marker genes of the two identified γδ T cell subsets. Color represents maximum-normalized mean expression of cells expressing marker genes.

References:

1. D.-M. Popescu, R. A. Botting, E. Stephenson, K. Green, S. Webb, L. Jardine, E. F. Calderbank, K. Polanski, I. Goh, M. Efremova, M. Acres, D. Maunder, P. Vegh, Y. Gitton, J.-E. Park, R. Vento-Tormo, Z. Miao, D. Dixon, R. Rowell, D. McDonald, J. Fletcher, E. Poyner, G. Reynolds, M. Mather, C. Moldovan, L. Mamanova, F. Greig, M. D. Young, K. B. Meyer, S. Lisgo, J. Bacardit, A. Fuller, B. Millar, B. Innes, S. Lindsay, M. J. T. Stubbington, M. S. Kowalczyk, B. Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L. Tickle, M. Slyper, O. Rozenblatt-Rosen, A. Filby, P. Carey, A.-C. Villani, A. Roy, A. Regev, A. Chédotal, I. Roberts, B. Göttgens, S. Behjati, E. Laurenti, S. A. Teichmann, M. Haniffa, Decoding human fetal liver haematopoiesis. *Nature*. **574**, 365–371 (2019).

2. L. C. Davies, S. J. Jenkins, J. E. Allen, P. R. Taylor, Tissue-resident macrophages. *Nat. Immunol.* **14**, 986–995 (2013).

3. E. Gomez Perdiguero, K. Klapproth, C. Schulz, K. Busch, E. Azzoni, L. Crozet, H. Garner, C. Trouillet, M. F. de Bruijn, F. Geissmann, H.-R. Rodewald, Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature*. **518**, 547–551 (2015).

4. X. Fan, A. Y. Rudensky, Hallmarks of Tissue-Resident Lymphocytes. *Cell*. **164**, 1198–1211 (2016).

5. J. M. Schenkel, D. Masopust, Identification of a resident T-cell memory core transcriptional signature. *Immunology & Cell Biology*. **92** (2014), pp. 8–9.

6. D. L. Woodland, J. E. Kohlmeier, Migration, maintenance and recall of memory T cells in peripheral tissues. *Nat. Rev. Immunol.* **9**, 153–161 (2009).

7. D. J. Topham, E. C. Reilly, Tissue-Resident Memory CD8+ T Cells: From Phenotype to Function. *Front. Immunol.* **9**, 515 (2018).

8. M. T. Wong, D. E. H. Ong, F. S. H. Lim, K. W. W. Teng, N. McGovern, S. Narayanan, W. Q. Ho, D. Cerny, H. K. K. Tan, R. Anicete, B. K. Tan, T. K. H. Lim, C. Y. Chan, P. C. Cheow, S. Y. Lee, A. Takano, E.-H. Tan, J. K. C. Tam, E. Y. Tan, J. K. Y. Chan, K. Fink, A. Bertoletti, F. Ginhoux, M. A. Curotto de Lafaille, E. W. Newell, A High-Dimensional Atlas of Human T Cell Diversity Reveals Tissue-Specific Trafficking and Cytokine Signatures. *Immunity*. **45**, 442–456 (2016).

9. P. Dogra, C. Rancan, W. Ma, M. Toth, T. Senda, D. J. Carpenter, M. Kubota, R. Matsumoto, P. Thapa, P. A. Szabo, M. M. Li Poon, J. Li, J. Arakawa-Hoyt, Y. Shen, L. Fong, L. L. Lanier, D. L. Farber, Tissue Determinants of Human NK Cell Development, Function, and Residence. *Cell*. **180**, 749–763.e13 (2020).

10. N. A. Yudanin, F. Schmitz, A.-L. Flamar, J. J. C. Thome, E. Tait Wojno, J. B. Moeller, M. Schirmer, I. J. Latorre, R. J. Xavier, D. L. Farber, L. A. Monticelli, D. Artis, Spatial and Temporal Mapping of Human Innate Lymphoid Cells Reveals Elements of Tissue Specificity. *Immunity*. **50**, 505–519.e4 (2019).

11. B. V. Kumar, T. J. Connors, D. L. Farber, Human T Cell Development, Localization, and Function throughout Life. *Immunity*. **48**, 202–213 (2018).

12. N. Aizarani, A. Saviano, Sagar, L. Mailly, S. Durand, J. S. Herman, P. Pessaux, T. F. Baumert, D. Grün, A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature*. **572**, 199–204 (2019).

13. J. Zhao, S. Zhang, Y. Liu, X. He, M. Qu, G. Xu, H. Wang, M. Huang, J. Pan, Z. Liu, Z. Li, L. Liu, Z. Zhang, Single-cell RNA sequencing reveals the heterogeneity of liver-resident immune cells in human. *Cell Discov*. **6**, 22 (2020).

14. J.-E. Park, R. A. Botting, C. Domínguez Conde, D.-M. Popescu, M. Lavaert, D. J. Kunz, I. Goh, E. Stephenson, R. Ragazzini, E. Tuck, A. Wilbrey-Clark, K. Roberts, V. R. Kedlian, J. R. Ferdinand, X. He, S. Webb, D. Maunder, N. Vandamme, K. T. Mahbubani, K. Polanski, L. Mamanova, L. Bolt, D. Crossland, F. de Rita, A. Fuller, A. Filby, G. Reynolds, D. Dixon, K. Saeb-Parsy, S. Lisgo, D. Henderson, R. Vento-Tormo, O. A. Bayraktar, R. A. Barker, K. B. Meyer, Y. Saeys, P. Bonfanti, S.

Behjati, M. R. Clatworthy, T. Taghon, M. Haniffa, S. A. Teichmann, A cell atlas of human thymic development defines T cell repertoire formation. *Science*. **367** (2020), doi:10.1126/science.aay3224.

15. K. R. James, T. Gomes, R. Elmentaite, N. Kumar, E. L. Gulliver, H. W. King, M. D. Stares, B. R. Bareham, J. R. Ferdinand, V. N. Petrova, K. Polański, S. C. Forster, L. B. Jarvis, O. Suchanek, S. Howlett, L. K. James, J. L. Jones, K. B. Meyer, M. R. Clatworthy, K. Saeb-Parsy, T. D. Lawley, S. A. Teichmann, Distinct microbial and immune niches of the human colon. *Nat. Immunol.* **21**, 343–353 (2020).

16. B. J. Stewart, J. R. Ferdinand, M. D. Young, T. J. Mitchell, K. W. Loudon, A. M. Riding, N. Richoz, G. L. Frazer, J. U. L. Staniforth, F. A. Vieira Braga, R. A. Botting, D.-M. Popescu, R. Vento-Tormo, E. Stephenson, A. Cagan, S. J. Farndon, K. Polanski, M. Efremova, K. Green, M. Del Castillo Velasco-Herrera, C. Guzzo, G. Collord, L. Mamanova, T. Aho, J. N. Armitage, A. C. P. Riddick, I. Mushtaq, S. Farrell, D. Rampling, J. Nicholson, A. Filby, J. Burge, S. Lisgo, S. Lindsay, M. Bajenoff, A. Y. Warren, G. D. Stewart, N. Sebire, N. Coleman, M. Haniffa, S. A. Teichmann, S. Behjati, M. R. Clatworthy, Spatiotemporal immune zonation of the human kidney. *Science*. **365**, 1461–1466 (2019).

17. P. A. Szabo, H. M. Levitin, M. Miron, M. E. Snyder, T. Senda, J. Yuan, Y. L. Cheng, E. C. Bush, P. Dogra, P. Thapa, D. L. Farber, P. A. Sims, Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019).

18. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Principal investigators, Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. **562**, 367–372 (2018).

19. S. He, L.-H. Wang, Y. Liu, Y.-Q. Li, H.-T. Chen, J.-H. Xu, W. Peng, G.-W. Lin, P.-P. Wei, B. Li, X. Xia, D. Wang, J.-X. Bei, X. He, Z. Guo, Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs, , doi:10.1101/2020.03.18.996975.

20. X. Han, Z. Zhou, L. Fei, H. Sun, R. Wang, Y. Chen, H. Chen, J. Wang, H. Tang, W. Ge, Y. Zhou, F. Ye, M. Jiang, J. Wu, Y. Xiao, X. Jia, T. Zhang, X. Ma, Q. Zhang, X. Bai, S. Lai, C. Yu, L. Zhu, R. Lin, Y. Gao, M. Wang, Y. Wu, J. Zhang, R. Zhan, S. Zhu, H. Hu, C. Wang, M. Chen, H. Huang, T. Liang, J. Chen, W. Wang, D. Zhang, G. Guo, Construction of a human cell landscape at single-cell level. *Nature*. **581**, 303–309 (2020).

21. C. C. Brown, H. Gudjonson, Y. Pritykin, D. Deep, V.-P. Lavallée, A. Mendoza, R. Fromme, L. Mazutis, C. Ariyan, C. Leslie, D. Pe'er, A. Y. Rudensky, Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell*. **179**, 846–863.e24 (2019).

22. A.-C. Villani, R. Satija, G. Reynolds, S. Sarkizova, K. Shekhar, J. Fletcher, M. Griesbeck, A. Butler, S. Zheng, S. Lazo, L. Jardine, D. Dixon, E. Stephenson, E. Nilsson, I. Grundberg, D. McDonald, A. Filby, W. Li, P. L. De Jager, O. Rozenblatt-Rosen, A. A. Lane, M. Haniffa, A. Regev, N. Hacohen, Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. **356** (2017), doi:10.1126/science.aah4573.

23. E. Evren, E. Ringqvist, K. P. Tripathi, N. Sleiers, I. C. Rives, A. Alisjahbana, Y. Gao, D. Sarhan, T. Halle, C. Sorini, R. Lepzien, N. Marquardt, J. Michaëlsson, A. Smed-Sörensen, J. Botling, M. C. I. Karlsson, E. J. Villablanca, T. Willinger, Distinct developmental pathways from blood monocytes generate human lung macrophage diversity. *Immunity*. **54**, 259–275.e7 (2021).

24. A. Deczkowska, A. Weiner, I. Amit, The Physiology, Pathology, and Potential Therapeutic Applications of the TREM2 Signaling Pathway. *Cell*. **181**, 1207–1217 (2020).

25. A. Martens, G. van Loo, A20 at the Crossroads of Cell Death, Inflammation, and Autoimmunity. *Cold Spring Harb. Perspect. Biol.* **12** (2020), doi:10.1101/cshperspect.a036418.

26. M. Haldar, M. Kohyama, A. Y.-L. So, W. Kc, X. Wu, C. G. Briseño, A. T. Satpathy, N. M. Kretzer, H. Arase, N. S. Rajasekaran, L. Wang, T. Egawa, K. Igarashi, D. Baltimore, T. L. Murphy, K. M. Murphy, Heme-mediated SPI-C induction promotes monocyte differentiation into iron-recycling

macrophages. *Cell*. **156**, 1223–1234 (2014).

27. M. A. Seibold, S. Donnelly, M. Solon, A. Innes, P. G. Woodruff, R. G. Boot, E. G. Burchard, J. V. Fahy, Chitotriosidase is the primary active chitinase in the human lung and is modulated by genotype and smoking habit. *J. Allergy Clin. Immunol.* **122**, 944–950.e3 (2008).

28. S. J. Van Dyken, H.-E. Liang, R. P. Naikawadi, P. G. Woodruff, P. J. Wolters, D. J. Erle, R. M. Locksley, Spontaneous Chitin Accumulation in Airways and Age-Related Fibrotic Lung Disease. *Cell*. **169**, 497–509.e13 (2017).

29. K. J. Travaglini, A. N. Nabhan, L. Penland, R. Sinha, A. Gillich, R. V. Sit, S. Chang, S. D. Conley, Y. Mori, J. Seita, G. J. Berry, J. B. Shrager, R. J. Metzger, C. S. Kuo, N. Neff, I. L. Weissman, S. R. Quake, M. A. Krasnow, A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature*. **587**, 619–625 (2020).

30. F. Sallusto, P. Schaerli, P. Loetscher, C. Schaniel, D. Lenig, C. R. Mackay, S. Qin, A. Lanzavecchia, Rapid and coordinated switch in chemokine receptor expression during dendritic cell maturation. *Eur. J. Immunol.* **28**, 2760–2769 (1998).

31. M. R. Clatworthy, C. E. P. Aronin, R. J. Mathews, N. Y. Morgan, K. G. C. Smith, R. N. Germain, Immune complexes stimulate CCR7-dependent dendritic cell migration to lymph nodes. *Nat. Med.* **20**, 1458–1463 (2014).

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

33. J. G. Cyster, Chemokines and the homing of dendritic cells to the T cell areas of lymphoid organs. *J. Exp. Med.* **189** (1999), pp. 447–450.

34. B. N. Lambrecht, H. Hammad, Biology of lung dendritic cells at the origin of asthma. *Immunity*. **31**, 412–424 (2009).

35. F. A. Vieira Braga, G. Kar, M. Berg, O. A. Carpaij, K. Polanski, L. M. Simon, S. Brouwer, T. Gomes, L. Hesse, J. Jiang, E. S. Fasouli, M. Efremova, R. Vento-Tormo, C. Talavera-López, M. R. Jonker, K. Affleck, S. Palit, P. M. Strzelecka, H. V. Firth, K. T. Mahbubani, A. Cvejic, K. B. Meyer, K. Saeb-Parsy, M. Luinge, C.-A. Brandsma, W. Timens, I. Angelidis, M. Strunz, G. H. Koppelman, A. J. van Oosterhout, H. B. Schiller, F. J. Theis, M. van den Berge, M. C. Nawijn, S. A. Teichmann, A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* **25**, 1153–1163 (2019).

36. W. Sungnak, N. Huang, C. Bécavin, M. Berg, R. Queen, M. Litvinukova, C. Talavera-López, H. Maatz, D. Reichart, F. Sampaziotis, K. B. Worlock, M. Yoshida, J. L. Barnes, HCA Lung Biological Network, SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).

37. M.-L. Golinski, M. Demeules, C. Derambure, G. Riou, M. Maho-Vaillant, O. Boyer, P. Joly, S. Calbo, CD11c+ B Cells Are Mainly Memory Cells, Precursors of Antibody Secreting Cells in Healthy Donors. *Front. Immunol.* **11**, 32 (2020).

38. E. Blanco, M. Pérez-Andrés, S. Arriba-Méndez, T. Contreras-Sanfeliciano, I. Criado, O. Pelak, A. Serra-Caetano, A. Romero, N. Puig, A. Remesal, J. T. Canizales, E. López-Granados, T. Kalina, A. E. Sousa, M. van Zelm, M. van der Burg, J. J. M. van Dongen, A. Orfao, Age-associated distribution of normal B-cell and plasma cell subsets in peripheral blood. *Journal of Allergy and Clinical Immunology*. **141** (2018), pp. 2208–2219.e16.

39. C. Wei, J. Anolik, A. Cappione, B. Zheng, A. Pugh-Bernard, J. Brooks, E.-H. Lee, E. C. B. Milner, I. Sanz, A new population of cells lacking expression of CD27 represents a notable component of the B cell memory compartment in systemic lupus erythematosus. *J. Immunol.* **178**, 6624–6633 (2007).

40. Z. Luo, Y. Zhou, P. Luo, Q. Zhao, N. Xiao, Y. Yu, Q. Yan, G. Lu, L. Cheng, SPARC deficiency affects bone marrow stromal function, resulting in impaired B lymphopoiesis. *J. Leukoc. Biol.* **96**, 73–82 (2014).

41. S. Sangaletti, C. Tripodo, C. Vitali, P. Portararo, C. Guarnotta, P. Casalini, B. Cappetti, S. Miotti, P. Pinciroli, F. Fuligni, F. Fais, P. P. Piccaluga, M. P. Colombo, Defective stromal remodeling and neutrophil extracellular traps in lymphoid tissues favor the transition from autoimmunity to lymphoma. *Cancer Discov.* **4**, 110–129 (2014).

42. A. Kassambara, T. Rème, M. Jourdan, T. Fest, D. Hose, K. Tarte, B. Klein, GenomicScape: An Easy-to-Use Web Tool for Gene Expression Data Analysis. Application to Investigate the Molecular Events in the Differentiation of B Cells into Plasma Cells. *PLOS Computational Biology*. **11** (2015), p. e1004077.

43. K. Kitaura, H. Yamashita, H. Ayabe, T. Shini, T. Matsutani, R. Suzuki, Different Somatic Hypermutation Levels among Antibody Subclasses Disclosed by a New Next-Generation Sequencing-Based Antibody Repertoire Analysis. *Front. Immunol.* **8**, 389 (2017).

44. W. Meng, B. Zhang, G. W. Schwartz, A. M. Rosenfeld, D. Ren, J. J. C. Thome, D. J. Carpenter, N. Matsuoka, H. Lerner, A. L. Friedman, T. Granot, D. L. Farber, M. J. Shlomchik, U. Hershberg, E. T. Luning Prak, An atlas of B-cell clonal distribution in the human body. *Nat. Biotechnol.* **35**, 879–884 (2017).

45. D. J. Carpenter, T. Granot, N. Matsuoka, T. Senda, B. V. Kumar, J. J. C. Thome, C. L. Gordon, M. Miron, J. Weiner, T. Connors, H. Lerner, A. Friedman, T. Kato, A. D. Griesemer, D. L. Farber, Human immunology studies using organ donors: Impact of clinical variations on immune parameters in tissues and circulation. *Am. J. Transplant*. **18**, 74–88 (2018).

46. T. M. Fenton, P. B. Jørgensen, K. Niss, S. J. S. Rubin, U. M. Mörbe, L. B. Riis, C. Da Silva, A. Plumb, J. Vandamme, H. L. Jakobsen, S. Brunak, A. Habtezion, O. H. Nielsen, B. Johansson-Lindbom, W. W. Agace, Immune Profiling of Human Gut-Associated Lymphoid Tissue Identifies a Role for Isolated Lymphoid Follicles in Priming of Region-Specific Immunity. *Immunity*. **52**, 557–570.e6 (2020).

47. N. J. Schuldt, B. A. Binstadt, Dual TCR T Cells: Identity Crisis or Multitaskers? *J. Immunol.* **202**, 637–644 (2019).

48. M. Van der Vieren, D. T. Crowe, D. Hoekstra, R. Vazeux, P. A. Hoffman, M. H. Grayson, B. S. Bochner, W. Michael Gallatin, D. E. Staunton, The Leukocyte Integrin αDβ2 Binds VCAM-1: Evidence for a Binding Interface Between I Domain and VCAM-1. *The Journal of Immunology*. **163**, 1984–1990 (1999).

49. G. Pizzolato, H. Kaminski, M. Tosolini, D.-M. Franchini, F. Pont, F. Martins, C. Valle, D. Labourdette, S. Cadot, A. Quillet-Mary, M. Poupot, C. Laurent, L. Ysebaert, S. Meraviglia, F. Dieli, P. Merville, P. Milpied, J. Déchanet-Merville, J.-J. Fournié, Single-cell RNA sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRVδ1 and TCRVδ2 γδ T lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11906–11915 (2019).

50. M. P. Soares, I. Hamza, Macrophages and Iron Metabolism. *Immunity*. **44**, 492–504 (2016).

51. A. Havelka, K. Sejersen, P. Venge, K. Pauksens, A. Larsson, Calprotectin, a new biomarker for diagnosis of acute respiratory infections. *Sci. Rep.* **10**, 4208 (2020).

52. P. A. Reid, D. A. McAllister, A. C. Boyd, J. A. Innes, D. Porteous, A. P. Greening, R. D. Gray, Measurement of serum calprotectin in stable patients predicts exacerbation and lung function decline in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **191**, 233–236 (2015).

53. P. A. Szabo, P. Dogra, J. I. Gray, S. B. Wells, T. J. Connors, S. P. Weisberg, I. Krupska, R. Matsumoto, M. M. L. Poon, E. Idzikowski, S. E. Morris, C. Pasin, A. J. Yates, A. Ku, M. Chait, J. Davis-Porada, X. V. Guo, J. Zhou, M. Steinle, S. Mackay, A. Saqi, M. R. Baldwin, P. A. Sims, D. L. Farber, Longitudinal profiling of respiratory and systemic immune responses reveals myeloid cell-driven lung inflammation in severe COVID-19. *Immunity*. **54**, 797–814.e6 (2021).

54. T. Ito, Y.-H. Wang, O. Duramad, T. Hori, G. J. Delespesse, N. Watanabe, F. X.-F. Qin, Z. Yao, W. Cao, Y.-J. Liu, TSLP-activated dendritic cells induce an inflammatory T helper type 2 cell response through OX40 ligand. *J. Exp. Med.* **202**, 1213–1223 (2005).

55. V. Soumelis, P. A. Reche, H. Kanzler, W. Yuan, G. Edward, B. Homey, M. Gilliet, S. Ho, S. Antonenko, A. Lauerma, K. Smith, D. Gorman, S. Zurawski, J. Abrams, S. Menon, T. McClanahan, R. de Waal-Malefyt Rd, F. Bazan, R. A. Kastelein, Y.-J. Liu, Human epithelial cells trigger dendritic cell mediated allergic inflammation by producing TSLP. *Nat. Immunol.* **3**, 673–680 (2002).

56. H. J. Sutton, R. Aye, A. H. Idris, R. Vistein, E. Nduati, O. Kai, J. Mwacharo, X. Li, X. Gao, T. D. Andrews, M. Koutsakos, T. H. O. Nguyen, M. Nekrasov, P. Milburn, A. Eltahla, A. A. Berry, N. Kc, S. Chakravarty, B. K. L. Sim, A. K. Wheatley, S. J. Kent, S. L. Hoffman, K. E. Lyke, P. Bejon, F. Luciani, K. Kedzierska, R. A. Seder, F. M. Ndungu, I. A. Cockburn, Atypical B cells are part of an alternative lineage of B cells that participates in responses to vaccination and infection in humans. *Cell Rep.* **34**, 108684 (2021).

57. D. Nehar-Belaid, S. Hong, R. Marches, G. Chen, M. Bolisetty, J. Baisch, L. Walters, M. Punaro, R. J. Rossi, C.-H. Chung, R. P. Huynh, P. Singh, W. F. Flynn, J.-A. Tabanor-Gayle, N. Kuchipudi, A. Mejias, M. A. Collet, A. L. Lucido, K. Palucka, P. Robson, S. Lakshminarayanan, O. Ramilo, T. Wright, V. Pascual, J. F. Banchereau, Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat. Immunol.* **21**, 1094–1106 (2020).

58. A. Bonner, A. Almogren, P. B. Furtado, M. A. Kerr, S. J. Perkins, The nonplanar secretory IgA2 and near planar secretory IgA1 solution structures rationalize their different mucosal immune responses. *J. Biol. Chem.* **284**, 5077–5087 (2009).

59. M. Kilian, J. Reinholdt, H. Lomholt, K. Poulsen, E. V. Frandsen, Biological significance of IgA1 proteases in bacterial colonization and pathogenesis: critical evaluation of experimental evidence. *APMIS*. **104**, 321–338 (1996).

60. B. V. Kumar, W. Ma, M. Miron, T. Granot, R. S. Guyer, D. J. Carpenter, T. Senda, X. Sun, S.-H. Ho, H. Lerner, A. L. Friedman, Y. Shen, D. L. Farber, Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep.* **20**, 2921–2934 (2017).

61. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst*. **8**, 281–291.e9 (2019).

62. B. Pijuan-Sala, J. A. Griffiths, C. Guibentif, T. W. Hiscock, W. Jawaid, F. J. Calero-Nieto, C. Mulas, X. Ibarra-Soria, R. C. V. Tyser, D. L. L. Ho, W. Reik, S. Srinivas, B. D. Simons, J. Nichols, J. C. Marioni, B. Göttgens, A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*. **566**, 490–495 (2019).

63. G. Sturm, T. Szabo, G. Fotakis, M. Haider, D. Rieder, Z. Trajanoski, F. Finotello, Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics*. **36**, 4817–4818 (2020).

64. H. W. King, N. Orban, J. C. Riches, A. J. Clear, G. Warnes, S. A. Teichmann, L. K. James, Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci Immunol*. **6** (2021), doi:10.1126/sciimmunol.abe6291.

65. J. A. V. Heiden, J. A. Vander Heiden, G. Yaari, M. Uduman, J. N. H. Stern, K. C. O'Connor, D. A. Hafler, F. Vigneault, S. H. Kleinstein, pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics*. **30** (2014), pp. 1930–1932.

66. J. Ye, N. Ma, T. L. Madden, J. M. Ostell, IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–40 (2013).

67. D. Gadala-Maria, G. Yaari, M. Uduman, S. H. Kleinstein, Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E862–70 (2015).

68. N. T. Gupta, J. A. Vander Heiden, M. Uduman, D. Gadala-Maria, G. Yaari, S. H. Kleinstein, Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics*. **31**, 3356–3358 (2015).

69. V. Svensson, E. da Veiga Beltrame, L. Pachter, A curated database reveals trends in single-cell transcriptomics. *Database* . **2020** (2020), doi:10.1093/database/baaa073.

70. A. Haque, J. Engel, S. A. Teichmann, T. Lönnberg, A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 1–12 (2017).

71. T. Abdelaal, L. Michielsen, D. Cats, D. Hoogduin, H. Mei, M. J. T. Reinders, A. Mahfouz, A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol.* **20**, 1–19 (2019).

72. A. Alavi, M. Ruffalo, A. Parvangada, Z. Huang, Z. Bar-Joseph, A web server for comparative analysis of single-cell RNA-seq data. *Nat. Commun.* **9**, 1–11 (2018).

73. M. Lotfollahi, M. Naghipourfar, M. D. Luecken, M. Khajavi, M. Büttner, Z. Avsec, A. V. Misharin, F. J. Theis, Query to reference single-cell integration with transfer learning. *bioRxiv* (2020), p. 2020.07.16.205997.

74. N. D. Köhler, M. Büttner, F. J. Theis, Deep learning does not outperform classical machine learning for cell-type annotation. *bioRxiv* (2019), p. 653907.

75. G. La Manno, D. Gyllborg, S. Codeluppi, K. Nishimura, C. Salto, A. Zeisel, L. E. Borm, S. R. W. Stott, E. M. Toledo, J. C. Villaescusa, P. Lönnerberg, J. Ryge, R. A. Barker, E. Arenas, S. Linnarsson, Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell*. **167**, 566–580.e19 (2016).

76. R. Elmentaite, A. D. B. Ross, K. Roberts, K. R. James, D. Ortmann, T. Gomes, K. Nayak, L. Tuck, S. Pritchard, O. A. Bayraktar, R. Heuschkel, L. Vallier, S. A. Teichmann, M. Zilbauer, Single-Cell Sequencing of Developing Human Gut Reveals Transcriptional Links to Childhood Crohn's Disease. *Dev. Cell*. **55**, 771–783.e5 (2020).

77. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

78. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zagar, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. B. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data, , doi:10.1101/2020.10.12.335331.

79. J. C. Kimmel, D. R. Kelley, Semi-supervised adversarial neural networks for single-cell classification. *Genome Res.* (2021), doi:10.1101/gr.268581.120.

80. Tabula Muris Consortium, A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*. **583**, 590–595 (2020).

81. T. Gomes, S. A. Teichmann, C. Talavera-López, Immunology Driven by Large-Scale Single-Cell Sequencing. *Trends Immunol.* **40**, 1011–1021 (2019).

Acknowledgements:

We thank the deceased organ donors, donor families and the extended Cambridge Biorepository for Translational Medicine team for access to the tissue samples.

We thank the CellTypist Annotation Team (composed by Simone Webb, Laura Jardine (Muzlifah Haniffa laboratory), Benjamin Stewart, Kelvin Tuong (Menna Clatworthy laboratory), Regina Hoo (Roser Vento-Tormo laboratory), Kylie James, Jongeun Park, Elo Madissoon, Waradon Sungnak, Rasa Elmentaite, Peng He (Sarah Teichmann laboratory) for contributions to the harmonization of CellTypist labels. We acknowledge J. Eliasova and BioRender for graphical images. We thank Carlos Talavera-López and Natsuhiko Kumasaka for discussions on data analysis. We also acknowledge the support received from the Cellular Generation and Phenotyping (CGaP) core facility, the Cellular Genetics Informatics team and Core DNA Pipelines at the Wellcome Sanger Institute and the Cambridge NIHR BRC Cell Phenotyping Hub (Dept. Medicine, University of Cambridge). We are grateful to the CZI Immune Ageing Seed Network consortium for helpful discussions.

Funding:

This research was funded in part, by the Wellcome Trust (Grant number RG79413 to JLJ). For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The work was also supported by funding from the European Research Council (grant no. 646794 ThDEFINE to S.A.T.) and by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed here are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.T.G. was supported by the European Union's H2020 research and innovation program "ENLIGHT-TEN" under the Marie Sklodowska-Curie grant agreement 675395, LBJ and JLJ were funded directly by the Wellcome Trust (RG79413).

Supplementary Materials:

Materials and Methods

Supplementary Figures

Supplemental Note on CellTypist

Materials and methods

Tissue acquisition

All work was completed under ethically approved studies. Tissue was obtained from deceased organ donors following circulatory death (DCD) via the Cambridge Biorepository for Translational Medicine (CBTM, https://www.cbtm.group.cam.ac.uk/), REC 15/EE/0152. Briefly, donors proceeded to organ donation after cessation of circulation. Organs were then perfused *in situ* with cold organ preservation solution and cooled with topical application of ice. Samples for the study were obtained within 60 minutes of cessation of circulation and placed in University of Wisconsin (UW) organ preservation solution for transport at 4°C to the laboratory. Gut samples were taken from the locations indicated in Figure 1. Additional samples were obtained from the left lower lobe of the lung and the right lobe of the liver. Skeletal muscle was taken from the third intercostal space and bone marrow was obtained from the vertebral bodies. In addition, two donor-matched blood samples were taken just prior to treatment withdrawal, under REC approval 97/290.

Donor metadata table

Donor ID	Sex	Age	Primary cause of death	Multitrauma	Days in hospital	BMI	CMV/EBV/TOXO	Smoking	Alcohol (u/day)	Vasoactive agents	Antibiotics within 2 weeks of death	Steroids
A29	F	65-70	ICH	Y	2	30-35	CMV+/EBV+/TOXO-	N/K	<1	Desmopressin Metaraminol	Nil	N
A31	M	50-55	ICH	Y	8	30-35	CMV+/EBV+/TOXO-	Y	<1	Noradrenaline	Co-amoxiclav, Tazocin	N
A35	M	60-65	ICH	N	2	20-25	CMV-/EBV+/TOXO-	Y	>9	Noradrenaline Desmopressin	Gentamicin, Flucloxacillin	Y Dexamethasone
A37	F	55-60	ICH	N	3	20-25	CMV-/EBV+/TOXO-	Y	>9	Noradrenaline Metaraminol	Co-amoxiclav	N
A36	M	70-75	ICH	N	5	25-30	CMV-/EBV+/TOXO+	Y	<2	Noradrenaline Metaraminol	Amoxicillin (pre-admission) Flucloxacillin, Gentamicin, Clarithromycin, Co-amoxiclav	Y Prednisolone pre-treatment
A52	M	60-65	ICH	N	2	35-40	CMV-/EBV+/TOXO-	N/K	N/K	Noradrenaline Metaraminol Desmopressin	Co-amoxiclav	N

F = Female; M = Male; ICH = intracranial haemorrhage; CMV = Cytomegalovirus; EBV = Epstein-Barr virus; TOXO = Toxoplasmosis; N/K=not known; Y = Yes; N = No;

Tissue processing

All tissues were processed using a uniform protocol. Briefly, solid tissues were transferred to a 100mm tissue culture dish, cut into small pieces and transferred to C-tubes (Miltenyi Biotec) at a maximum of 5g/tube in 5mL of X-vivo15 media containing 0.13U/m Liberase TL (Roche), 10U/mL DNase (benzonase nuclease, Millipore/Merck) supplemented with 2% (v/v) heat-inactivated fetal bovine serum (FBS; Gibco), penicillin (100 U/ml, Sigma-Aldrich), streptomycin (0.1 mg/ml, Sigma-Aldrich), and 10mM HEPES (Sigma Aldrich). The samples were then dissociated using a GentleMACS Octo dissociator (Miltenyi Biotec) using a protocol that provided gradual ramping up of homogenisation speed along with 2x 15 minute heating/mixing steps at 37°C. Digested tissue was filtered through a 70-µm MACS Smartstrainer (miltenyi biotec) and washed with media containing 2mM EDTA, prior to washing with PBS. A ficoll density centrifugation step (400g for 30min at RT) was performed to isolate mononuclear cells (MNCs). After gradient centrifugation, cells were washed once with PBS prior to counting and resuspending in 10X buffer (PBS containing 0.04% (v/v) BSA).

Bone marrow aspirates and blood samples were diluted 1:1 with PBS and layered directly onto ficoll for mononuclear cell isolation as described above. Cells taken from the interphase layer were washed with PBS and exposed to the same enzymatic conditions as solid tissues by resuspending cell pellets in tissue dissociation media containing liberase TL for 30 minutes at 37°C prior to counting and resuspending in 10X buffer.

Single-cell RNA sequencing

For scRNA-seq experiments, single cells were loaded onto the channels of a Chromium chip (10x Genomics) for a target recovery of 5,000 cells. Single-cell cDNA synthesis, amplification, and sequencing libraries were generated using the Single Cell 5′ Reagent Kit following the manufacturer's instructions. The libraries from up to eight loaded channels were multiplexed together and sequenced on an Illumina NovaSeq. The libraries were distributed over eight lanes per flow cell and sequenced at a target depth of 50,000 reads per cell using the following parameters: Read1: 26 cycles, i7: 8 cycles, i5: 0 cycles; Read2: 98 cycles to generate 75-bp paired-end reads. VDJ libraries for B and T cells were sequenced on HiSeq 4000.

Single-cell RNA sequencing data pre-processing

scRNA-seq data was aligned and quantified using the cellranger software (version 3.0.2, 10x Genomics Inc.) using the GRCh38 human reference genome (official Cell Ranger reference, version 1.2.0). Cells with fewer than 1,000 UMI counts and 600 detected genes were excluded from downstream analysis. scTCR-seq data was aligned and quantified using the cellranger-vdj software (version 2.1.1, 10x Genomics Inc). scBCR-seq data was aligned and quantified using the cellranger-vdj software (version

4.0, 10x Genomics Inc). For TCRγδ we implemented a customized pipeline due to the cell ranger vdj annotation algorithm being tuned towards alpha/beta TCR chains. Briefly, TCR gamma/delta libraries were mapped with cell ranger 4.0.0, using the 10x VDJ 4.0.0 reference. All contigs deemed high quality were selected, and reannotated with IgBlast via the workflow provided in dandelion 0.1.3 (https://github.com/zktuong/dandelion). We provide an example notebook showcasing this workflow (https://sc-dandelion.readthedocs.io/en/latest/notebooks/gamma_delta.html).

Doublet detection

Doublet detection was performed on a per sample basis using the scrublet algorithm (https://github.com/AllonKleinLab/scrublet,(*61*)) with percolation as previously described (*62*). Briefly, scrublet scores were obtained per cell and the percolation step was performed on over-clustered data using the *scanpy.tl.louvain* function from the scanpy package. Each cluster was subsequently separately clustered again, yielding an over-clustered manifold, and each of the resulting clusters had its Scrublet scores replaced by the median of the observed values. The resulting scores were assessed for statistical significance, with P values computed using a right-tailed test from a normal distribution centred on the score median and a median absolute deviation (MAD)-derived standard deviation estimate. The P-values were corrected for false discovery rate with the Bonferroni procedure, and a significance threshold of 0.01 was imposed. Cells with a Bonferroni-corrected P-value under 0.01 were excluded from downstream analysis.

Clustering, batch alignment and annotation

Downstream analysis included data normalisation (*scanpy.pp.normalize_per_cell* method, scaling factor 10000), log-transformation (*scanpy.pp.log1p*), variable gene detection (*sc.pp.highly_variable_genes*), data feature scaling (*scanpy.pp.scale*), PCA analysis (*scanpy.pp.pca*, from variable genes), and Leiden graph-based clustering (*scanpy.tl.leiden*, clustering resolution manually adjusted) performed using the python package scanpy (version 1.6.0). Data integration across donors was done using batch-balanced KNN (https://github.com/Teichlab/bbknn). Cluster cell identity was assigned by manual annotation using known marker genes as well as computed differentially expressed genes (DEGs). Differential expression across clusters was assessed using rank biserial correlation (https://github.com/Teichlab/rbcde) and *markers.py* functions from the thymus atlas (https://github.com/Teichlab/thymusatlas) (*14*). Cross-tissue differential expression was assessed using a ridge regression model that included an interaction term for tissue and cell type. For each tissue, genes were ranked according to their coefficients specific to the given interaction term associated with the tissue of interest. To achieve a high-resolution annotation, we sub-clustered T, B and myeloid cells and repeated the procedure of variable gene selection, which allowed for fine-grained cell type annotation. Donor-dependent batch effects were aligned using the *scanpy.pp.bbknn* function and we used the batch-aligned manifold to annotate cell types.

CellTypist

Full details on the data collection, processing, curation, model training, and testing of the CellTypist pipeline can be found in the **Supplementary Note**. Briefly, immune cells from 20 tissues of 19 studies were collected and harmonised into consistent labels with the inputs from experts followed by semi-supervised label propagation to non-annotated cells. These cells were randomly split into equal-sized mini-batches, with each batch containing 1,000 cells. In each epoch, 100 such batches were sequentially trained by the l2 regularized logistic regression using stochastic gradient descent learning. This step was repeated 30 epochs, enabling the CellTypist models to see cell numbers with six orders of magnitude. Lastly, feature selection was performed to choose the top 500 genes from each cell type by ranking the genes according to their absolute weights associated with the given cell type, and the union of these genes across cell types were supplied as the input for a second round of training, with the detailed training procedure being constant as the initial round of training.

scTCR-seq downstream analysis

VDJ sequence information was extracted from the output file “filtered_contig_annotations.csv” using the scirpy package(*63*). We determined productive TCR chain pairing features using the *scirpy.tl.chain_pairing()* function and selected cells with a single pair of productive αβ TCR chains for downstream analysis. Bias in VDJ usage across tissues was estimated using the generalized linear model (glm) on the basis of the Poisson family. Specifically, for each gene segment, we calculated the number of cells with this segment in each donor of a given tissue, and then compared their distributions with the remaining tissues which were stratified by donors as well. During the glm fitting, the total number of cells in each donor of a given tissue was logarithmized and used as the offset, accounting for the variance imposed by rate estimation (glm(counts~tissues + offset(log(total_counts)), family = 'poisson') in R). Clonotypes were determined using the *scirpy.pp.tcr_neighbors()* function using the CDR3 nucleotide sequence identity from both TCR chains as a metric.

scBCR-seq downstream analysis

VDJ sequence information was extracted from the output file “filtered_contig_annotations.csv”. Further single-cell VDJ analysis for B cells was performed broadly as described previously (*15*, *64*), with all sequences from a given patient grouped together for analysis. AssignGenes.py(*65*) and IgBLAST(*66*) were used to reannotate IgH sequences prior to correction of ambiguous V gene assignments using TIgGER (v1.0.0)(*67*). Clonally-related IgH sequences were identified using DefineClones.py with a nearest neighbour distance threshold of 0.15 before running CreateGermlines.py (ChangeO)(*68*) to infer germline sequences for each clonal family and calculate somatic hypermutation frequencies with

observedMutations (Shazam)(*68*). IgH diversity analyses were performed using the rarefyDiversity and testDiversity of Alakazam (v1.0.2; (*68*)). scVDJ sequences were then integrated with single-cell gene expression objects by determining the number of high quality annotated IgH, IgK or IgL per unique cell barcode. If more than one contig per chain was identified, metadata for that cell was ascribed as "Multi". To assess clonal relationships between scRNA-seq clusters, co-occurrence of expanded clone members between cell types and tissues was reported as a binary event for each clone that contained a member within two different cell types or tissues in single-cell repertoires.

Single molecule FISH

Samples were either snap frozen in chilled isopentane (-40ºC for striated muscle, -70ºC for other tissues) or fixed in 10% NBF, dehydrated through an ethanol series, and embedded in paraffin wax. Samples were run using the RNAscope 2.5 LS fluorescent multiplex assay (automated). Briefly, FFPE tissue sections (5 µm) and fresh frozen tissue sections (10um) were cut. Fresh frozen tissues were pre-treated offline (4% PFA fixation 4°C 15 mins followed by 90mins at room temperature, sequential dehydration steps (50%, 70%, 100%, 100% ethanol, air dry)) and protease III was used. FFPE tissues required no pretreatment offline, but a Heat Induced Epitope Retrieval (HIER) step was performed by the instrument for 15mins using Epitope Retrieval 2 (ER2) at 95°C. These tissues also had protease III treatment. RNAscope probes used included Hs-CD3D-C2 (599398-C2), Hs-CD8A-C3 (560398-C3), Hs-CRTAM (430248). Opal fluorophores (Opal 520, Opal 570 and Opal 650) were used at 1:300 dilution. Slides were imaged on the Perkin Elmer Opera Phenix High-Content Screening System, in confocal mode with 1 µm z-step size, using 20X (NA 0.16, 0.299 µm/pixel) and 40X (NA 1.1, 0.149 µm/pixel) water-immersion objectives.

Flow cytometry

Spleen, bone marrow and thoracic lymph nodes from additional donors different to the scRNA-seq study were used to validate the discovered cell populations. The mononuclear cells (MNCs) were either stained ex vivo or post activation with PMA+I (eBioscience, Cell Stimulation Cocktail) for two hours. Cells were washed with PBS and then stained with the live/dead marker Zombie Aqua for 10 minutes at room temperature, and then washed with PBS+0.5%FCS. The MNCs were stained in PBS+0.5% FCS at 4°C for 45 minutes with the following panels of antibodies:

CD8 panel: CD3-BUV395 (SK7, BioLegend), CD56-BUV737 (NCAM16.2, BD Horizon), CCR9-BV421 (L053E8, BioLegend), CD4-BV605 (OKT4, BioLegend), TCRgd-Fitc (B1.1, Invitrogen), CX3CR1-PE (2A9-1, BioLegend), CRTAM-PECy7 (CR24.1, Invitrogen), CD16-APC (3G8, BioLegend), CD8-APCCy7 (RPA-T8, BioLegend).

B cell panel: IgD-BUV395 (IA6-2, BD Horizon), CCR7-BV421 (G043H7, BioLegend), CD3-BV605 (SK7, BioLegend), CD11c-BV785 (3.9, BioLegend), CD27-PE (0323, eBioscience), CD19-APC (HIB19, BioLegend), Tbet-PECy7 (eBio4810, eBioscience,Tbet staining was done after the surface staining using the eBioscience Foxp3 transcription factor staining buffer kit).

Cells were fixed with PBS+0.25%PFA and stored at 4°C until they were run on the Fortessa flow cytometry instrument, located in the Cambridge NIHR BRC Cell Phenotyping Hub. Spleen MNCs were used for single stain controls to calculate compensation and FMOs were used to calculate background fluorescence. FlowJo was used to analyse the flow cytometry data.

qPCR

qPCR was performed to validate the existence ITGAD-expressing γδ T cells in the spleen using three additional samples that were different to those used in the scRNA-seq study. Spleen MNCs were stained with the live/dead marker Zombie Aqua for 10 minutes at room temperature, and then washed with PBS+0.5%FCS. Cells were then stained with the following antibodies at 4°C for 45 minutes: CD56-BV421 (HCD56, BioLegend), CD4-BV605 (OKT4, BioLegend), TCRgd-Fitc (B1.1, Invitrogen), TCRab-PerCPCy5.5 (IP26, BioLegend), CD52-PE (MHCD5204, Life Technologies), CD127-PECy7 (eBioRDR5, eBioscience), CD8-APC (RPA-T8, BioLegend), CD3-APCfire (UCHT1, BioLegend). Cells were washed with PBS+0.5% FCS and passed through a Celltrics (Partec) 30 µm filter prior to cell sorting. Cell sorting was performed on a BD Fusion 4 laser sorter and an example of the gating strategy used is shown in (**fig. S7**).

Sorted cells were pelleted at 600g for 5 minutes and lysed in RNA lysis buffer and frozen until RNA could be extracted. RNA was extracted using a Zymo Research RNA micro kit with on column DNAse digestion. The standard protocol was followed and RNA eluted in 11 µl of water. This RNA was then used to make cDNA using both oligo dT and random hexamer primers with the reverse transcriptase SuperscriptIII. Probes to B2M (housekeeping gene) and two assays to ITGAD were purchased from ThermoFisher, and qPCR reactions were performed in duplicate with the following recipe: 8 µl master mix, 0.7 µl probe, 4.3 µl water and 3 µl cDNA. A ThermoFIsher QuantStudio7 instrument was used for the qPCR, and the Ct values were determined with the DCt being calculated as the Ct of ITGAD - Ct of B2M.

Immunofluorescence

Spleen and thoracic lymph node samples from unrelated donors were fixed in 1% paraformaldehyde (Electron Microscopy Services) for 24 hours followed by 8 hours in 30% sucrose in PBS. 30µm sections were permeabilized and blocked in 0.1M TRIS, containing 0.1% Triton (Sigma), 1% normal mouse serum, 1% normal rat serum and 1% BSA (R and D). Samples were stained for 2h at RT in a wet

chamber with the appropriate antibodies, washed 3 times in PBS and mounted in Fluoromount-G® (Southern Biotech). Images were acquired using a TCS SP8 (Leica, Milton Keynes, UK) confocal microscope. Raw imaging data were processed using Imaris (Bitplane).

Antibodies used: CD3-AF488, clone UCHT1, 1/100 dilution, BioLegend; CD1c-PE, clone L161, 1/50 dilution, BioLegend; CCR7-PE, clone 3D12, 1/50 dilution, eBioscience; CD19-AF594, clone HIB19, 1/100 dilution, BioLegend; CD11c-APC, clone MJ4-27G12, 1/100 dilution, Miltenyi; HLA-DR-AF647, clone TAL 1B5, 1/100 dilution, abcam. Nuclei were stained with Hoechst 33258, 1/10,000 dilution (Biotum).