bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture across the human body

Authors:

Domínguez Conde C[1,*], Gomes T[1,*], Jarvis LB[2,*], Xu C[1,*], Howlett SK[2], Rainbow DB[2], Suchanek O[3], King HW[4], Mamanova L[1], Polanski K[1], Huang N[1], Fasouli E[1], Mahbubani KT[5], Prete M[1], 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, Cambridge, 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, Cambridge, UK

*Co-first authors

#Corresponding authors

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Despite their crucial role in health and disease, our knowledge of immune cells within human tissues, in contrast to those circulating in the blood, remains limited. Here, we surveyed the immune compartment of lymphoid and non-lymphoid tissues of six adult donors by single-cell RNA sequencing, including alpha beta T-cell receptor (αβ TCR), gamma delta (γδ) TCR and B-cell receptor (BCR) variable regions. To aid systematic cell type identification we developed CellTypist, a tool for automated and accurate cell type annotation. Using this approach combined with manual curation, we determined the tissue distribution of finely phenotyped immune cell types and cell states. This revealed tissue-specific features within cell subsets, such as a subtype of activated dendritic cells in the airways (expressing *CSF2RA, GPR157, CRLF2)*, *ITGAD*-expressing γδ T cells in spleen and liver, and *ITGAX*+ splenic memory B cells. Single cell paired chain TCR analysis revealed cell type-specific biases in VDJ usage, and BCR analysis revealed characteristic patterns of somatic hypermutation and isotype usage in plasma and memory B cell subsets. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

The immune system is a dynamic and integrated network made up of many cell types that are distributed across the entire organism and 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 on how macrophages derived in embryogenesis contribute to uniquely adapted adult 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 NK cells, and non-conventional (NKT, MAIT and γδ) T cells, have circulating counterparts but are highly enriched at barrier/mucosal sites, where they maintain tissue health by sensing stress, promoting tissue repair and antimicrobial defense (*4*). In addition, long-lived tissue-resident memory T cells (TRMs) are able to migrate into peripheral tissues and take up residency after the resolution of infection, providing protection from secondary infections via their ability to rapidly regain effector functions (reviewed in (*5*, *6*)). Tissue-specific adaptations have been reported for many of these cells, such as key tissue-tropic chemokine receptor/integrin profiles and expression of markers that support tissue retention (reviewed in (*7*)). The specialisations of immune cells within tissues serve as a basis for their potentially unique roles in local environments, and in the meantime pose a challenge to our understanding of their cellular diversity throughout the human body.

Despite the important implications tissue immunity has for health and disease, much of this knowledge comes from animal studies. Historically, human immune cell analysis has focused on blood, providing a biased and incomplete view of our immune system. A number of recent studies analysing human tissues by flow cytometry (*8–11*), and several organ-focused studies utilising single-cell genomics (*12–16*) have improved upon this, but few have analysed immune cells across multiple tissues from the same individual and thus controlling for immunological experience. One such study by Szabo et al. reported an analysis of T cells in three tissues from two donors (*17*). However, murine (*18*) and human *(19, 20)* large-scale multi-tissue scRNA-seq studies investigating tissue-specific features of the immune compartment remain rare.

Here, we comprehensively profiled immune cell populations isolated from 15 donor-matched tissues from six individuals to provide novel insights into tissue-specific immunity. To capture the vast cellular diversity for annotating muti-tissue immune cells, we developed CellTypist, an immune cell resource compiled from 19 studies under a common framework followed by the use of logistic regression classifiers for the accurate prediction of cell type identities. Combining automated annotation and in-depth dissection of the cellular heterogeneity within the myeloid and lymphoid compartments, we determined the frequency and transcriptional characteristics of immune cells across human tissues. Moreover, we inferred the patterns of T and B cell migration across different tissues and their transition between cell states based on antigen-receptor sequencing, gaining insights into tissue-dependent phenotypic and clonal plasticity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Results

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

To systematically assess immune cell type heterogeneity across human tissues, we performed single-cell RNA sequencing (scRNA-seq) on 15 different tissues from six deceased organ donors (**Fig. 1A**, **Supplementary Table 1)**. Briefly, cells were isolated using the same 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 hematopoietic cells, with higher yields from the lymphoid tissues in all donors (**fig. S1A,B**). Using manual annotation, 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, including monocytes, macrophages and dendritic cells (**Fig. 1B**,**C, fig. S1C,D**). 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 bone marrow and blood; macrophage and mast cell populations were enriched in the lung; lymphocytes mainly came from the lymphoid organs (**fig. S1E**). In addition, analyses of the cell type composition within each tissue (**fig. S1F**) 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. Gut regions were also abundant in plasma cells, and the lung parenchyma immune compartment was dominated by monocytes and macrophages.

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://pypi.org/project/celltypist-dev/), a lightweight classification pipeline for scRNA-seq data based on an expandable cell type reference atlas assembled from multiple tissues and studies (**Fig. 1D**, Supplementary Note). In brief, CellTypist currently includes reference datasets for 20 tissues (**Supplementary Note Figure 1A**) with harmonized cell type labels at different hierarchically-structured resolutions. This hierarchy is implemented with logistic regression models, enabling the prediction of cell identity at different levels of specificity (**Supplementary Note Table 1**), a feature of particular importance for determining immune cell identity accurately. Global F1-scores for the models ranged between 0.88 and 0.94 (**Fig. 1E**). The representation of a given cell type in the training data is a major determinant in how well the model can predict it (**Supplementary Note Fig. 2B and 3C,D**); therefore this will be improved with the incorporation of more datasets.

We first applied the high-hierarchy (low-resolution) classifier to our cross-tissue dataset and found high consistency when comparing the predicted cell types to our coarse-grained manual annotations (**Fig. 1F**). Due to their functional plasticity, monocytes and macrophages often form a continuum in scRNA-seq datasets, which matches the observation of their high cross-classification frequency in our

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

study. Furthermore, as the training dataset of CellTypist contains hematopoietic tissues with resolved annotations for progenitor populations, the classifier could unravel the manually annotated progenitors into HSC/MPP, pro-myelocytes, erythrocytes and monocytes. In this way, CellTypist can inform and refine manual annotations.

To allow for automated annotation of more specific immune sub-populations, we further applied the low-hierarchy (high-resolution) classifier, which is able to predict cell subtypes including subsets of T cells, B cells, ILCs and dendritic cells (**Fig. 1F**). This classification highlighted a high degree of heterogeneity within the T cell compartment, not only distinguishing between αβ and γδ T cells, but also unravelling 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 MAIT and NKT. Moreover, CellTypist identified a population of germinal center B cells and revealed three distinct subsets of dendritic cells - DC1, DC2 and activated DCs (aDC) (*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. CellTypist produced expert-grade annotations on our multi-tissue and multi-lineage dataset, and its performance was better or comparable relative to other label-transferring methods with minimal computational cost (**Supplementary Note Figs 4, 5**). This approach allowed us to further refine the description of multiple cell subtypes such as the progenitors and dendritic cell subtypes at full transcriptomic breadth.

Tissue-restricted features of mononuclear phagocytes

Mononuclear phagocytes, including monocytes, macrophages and dendritic cells, are critical for immune surveillance and tissue homeostasis. Subclustering of the myeloid subsets unveiled further heterogeneity, particularly within the macrophage and monocyte compartments (**Fig. 2A,B, fig. S2A**). We found four major macrophage clusters in the lung: (I) and (II) expressed *GPNMB* and *TREM2*, (III) expressed epithelial markers - potentially due to ambient RNA, doublets or ingestion of epithelial debris -, and (IV) expressed *TNIP3*. TNIP3 (TNFAIP3-interacting protein 3) binds to A20 (also known as TNF alpha induced protein 3) and inhibits TNF, IL-1 and LPS induced NF-kB activation. Its expression in lung macrophages may be related to underlying pathology as it was primarily detected in a multitrauma donor (A29) with lung contusions (**fig. S2B**). Red pulp macrophages and Kupffer cells expressed *CD5L*, *SCL40A1* and the transcription factor *SPIC(23)*, with *LYVE1* and *ITGA9* distinguishing the Kupffer cells. A small population of macrophages expressed chitin and kinin degrading enzymes *CHIT1* and *CTSK,* respectively. *CHIT1* encodes chitotriosidase, an enzyme secreted by macrophages that degrades chitin, a component of the exoskeletons of mites and other arthropods (*24*). Macrophages upregulate *CHIT1* during terminal differentiation, and its expression and activity correlates with lung disease (*25*, *26*). Moreover, we identified non-classical monocytes, three

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

subsets of classical monocytes and a small cluster of intermediate monocytes based on the expression of *CD14* and *FCGR3A*. Notably, several chemokines and their receptors, as well as adhesion molecules showed specific expression patterns across the myeloid subsets.

Within the myeloid compartment, macrophage subsets showed the highest degree of tissue restriction (**Fig. 2C**). Red pulp macrophages and Kupffer cells were mainly found in spleen and liver as expected, however, the presence of these clusters in other tissues such as the bone marrow and the mesenteric lymph node points to the transcriptional similarities between iron recycling macrophages. The *CHIT1*-expressing macrophages were restricted to the thoracic lymph node (TLN), however, other datasets have reported these in the lung (*27*). For dendritic cells, which were more broadly distributed across tissues, we assessed cross-tissue differential expression. Interestingly, activated dendritic cells showed upregulation of *AIRE, PDLIM4* and *EBI3* in the TLN, and to a lesser extent MLN, while in the lung they showed *CRLF2* (encoding TLSPR), upregulated chemokines (*CCL22*, *CCL17*), *CSF2RA* and *GPR157*. TLSPR is involved in the induction of Th2 responses in asthma (*28*) (**Fig. 2D, E**). These observations suggest that dendritic cell activation coincides with the acquisition of tissue-specific markers that differ depending on the local environment.

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

B cell subsets and immunoglobulin repertoires across tissues

B cells constitute the central player in humoral immunity via the production of antibodies that are tailored to specific body sites. We performed an in-depth analysis of the B cell compartment, revealing subpopulations across tissues (**Fig. 3A-C**). Within the memory B cells, globally characterized by expression of the B-cell lineage markers *MS4A1, CD19* and *TNFRSF13B,* we found a distinct cluster (MemB_ITGAX) positive for *ITGAX, TBX21* and *FCRL2,* encoding CD11c, T-bet and the Fc receptor-like protein 2, respectively. CD11c+T-bet+ B cells, also known as “age-associated B cells” or ABC cells, have been reported in a range of human conditions (*29*). Notably, unlike conventional memory B cells, they show low expression of *CR2* (encoding CD21) and *CD27*. We primarily detect this population in the spleen, thoracic lymph nodes and bone marrow. Within the naive B cell compartment, we observed a small cluster, characterized by expression of *ITGA2B* and *SPARC,* that was only found in blood and spleen and was distinct from the bulk of naive B cells. In addition, we identified two small populations of germinal center B cells, expressing *AICDA* and *BCL6,* that differed in their proliferative state (marked by *MKI67*). We did not find any differential expression of dark zone and light zone marker genes, probably reflecting limited germinal center activity in the adult donors. These GC populations were present in lymph nodes and different gut regions, presumably

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

representing Peyer's patches. Plasmablasts (Plasma_prolife) and plasma cells (Plasma_ITGA8) were marked by expression of *CD38, XBP1* and *SDC1.* The former expressed *MKI67* and were found in spleen, liver, bone marrow and lymph nodes; the latter expressed *ITGA8*, the adhesion molecule *CERCAM* and were detected primarily in spleen, caecum, liver, bone marrow and lymph nodes. Lastly, we also detected a small progenitor population in the bone marrow expressing *IGLL1, RAG1* and *DNTT* and two doublet populations between memory B cells and T cells or macrophages that were not excluded during automated doublet exclusion.

B cells have an additional source of variability due to VDJ recombination, somatic hypermutation and class-switching, which can influence the number of cell subtypes present. We performed targeted enrichment and sequencing of BCR transcripts to assess isotypes, hypermutation levels and clonal architecture of the B cell populations described above. Isotype and subclass usage followed patterns that related to cellular phenotype (**fig. S3C**). As expected, progenitors and naive B cells mainly showed IgM and IgD. Interestingly, while memory B cells showed evidence of class switching to IgA1 and IgG1, plasmablasts and plasma cells showed, in addition, a remarkable fraction of switching to subclasses IgA2 and IgG2. To determine to what extent this isotype subclass bias is correlated with tissue of origin, we assessed each cell state independently (minimum cell count of 19). Memory B cells in omentum and mesenteric lymph nodes showed a bias towards IgA1 and terminal ileum, where Peyer's patches are found, to both IgA1 and IgA2 (**Fig. 3D**). In the plasma cell compartment, we found an even more striking preference towards IgA2 in several gut regions (caecum, duodenum and transverse colon) (**Fig. 3E**). Also of note, plasma cells in bone marrow, liver and spleen were composed of over 20% IgG2 subclass. With more limited numbers, we also report isotype distribution per tissue for plasmablasts and *ITGAX*+ memory B cells (**fig. S3D,E**).

Somatic hypermutation (SHM) levels were, as expected, lowest in naive B cells and highest in plasma cells (**fig. S3F**). SHM did not differ significantly between isotypes or subclasses. Nonetheless, there was a tendency towards higher mutation rates in distal classes IgG2 and IgA2, which are downstream in the IgH locus and could accumulate more mutations during sequential switching (*30*) (**Fig. 3F**). We explored the occurrence of sequential switching events in our data by assessing isotype frequency among expanded clonotypes (>10 cells), however, the detection of mixed isotype clones was very infrequent in our data and mostly between IgM and IgD (**fig. S3G**). We then evaluated the distribution of these expanded clones across tissues and cell types. We found three major groups of clones. Those present in only 2 tissues, 3-4 tissues or in more than 5 tissues (**Fig. 3G**), similar to previously reported patterns of B cell clone tissue distribution (*31*). While those restricted to 2 tissues, typically spleen, liver and bone marrow, were enriched in plasma cells, those widespread across more than 5 tissues, including lymph nodes, were enriched in memory B cells. Together, these findings suggest that, as seen in the bone marrow of rhesus macaques (*32*), tissue-restricted clones could represent a long-term immunological memory maintained by long-lived plasma cells resident in the spleen, liver and bone marrow.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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

Most of the cells we captured are T cells or ILCs, including NK cells, further divided into 16 sub-clusters (**Fig. 4A and 4B**). Naive/central memory CD8+ and CD4+ T cells were closely located and defined by high *CCR7* and *SELL* expression. Follicular helper T cells (Tfh) expressing *CXCR5*, regulatory T cells (Tregs) expressing *FOXP3* and *CTLA4*, and other effector CD4+ cells were also identified. MAIT cells were characterised by expression of *TRAV1-2* and *SLC4A10*. In the CD8+ memory compartment, we found three major subsets characterized by expression of the chemokine receptors *CCR9* and *CX3CR1* as well as the activation molecule *CRTAM*. Furthermore, within the γδ T cell cluster we found expression of the integrin molecule *ITGAD* (CD11d). Notably, the NK cells were represented by two clusters, expressing high levels of either *FCGR3A* or *NCAM1*. The ILC3 population, marked by *PCDH9* expression, was mixed in a cluster with NK cells. Tissue distribution of these populations reveals that whereas the majority of CD4+ and ILC3 cells are located in the lymph nodes and to some extent in the spleen, cytotoxic T and NK cells are more abundant in the bone marrow, spleen and non-lymphoid tissues (**Fig. 4C**). Specifically within the CD8+ T cells, *CCR9+* were enriched in the gut regions whereas *CX3CR1+* were absent from mesenteric and thoracic lymph nodes and *CRTAM+* were found in both lymphoid and non-lymphoid tissues. We validated and mapped this last population of *CD3D+CD8A+CRTAM+* T cells using single-molecule FISH (smFISH) in the liver and thoracic lymph nodes (**Fig. 4D and 4E**).

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 T cell clusters mostly contained a single pair of chains, with orphan and extra chains being present on a relatively small fraction of cells (**fig. S4A,B**). Notably, the frequency of extra α chains (extra VJ) is more common than that of β chains (extra VDJ), thought to be due to more stringent and multi-layered allelic exclusion mechanisms at the TCRβ locus compared to TCRα (*33*). As expected, the NK and ILC clusters had no productive TCR and only a small proportion of the γδ T cell cluster had a productive TCR, which may be due to cytotoxic T cell co-clustering. We next examined V(D)J gene usage in relation to cell identity. We detected a significant enrichment of *TRAV1-2, TRAJ33* and *TRAJ12* in the MAIT population, as expected (**fig.S4C**). 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. Furthermore, we also found a significant enrichment of specific gene segments in the *CD8+CX3CR1+* population in the lung, bone marrow, liver, spleen and blood (**Fig. 4F**).

We then defined clonally related cells based on identical CDR3 nucleotide sequences to investigate TCR repertoires. Using this approach, we found that clonally expanded cells were primarily within the

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

memory CD8+ compartment and the MAIT population (**fig. S4D**). This data reveals that detected clonotypes were restricted to single individuals, however, within donors, they were commonly found across tissues and subsets (**fig. S4E-G**). Focusing on expanded clonotypes (>20 cells), the majority of them were widespread across five or more tissues (**Fig. 4G**). We also found that several CD8+CCR9+ T cell-enriched clonotypes were shared across different gut regions, and that several clonotypes present in the liver and lung consisted of a mixture of cells within both the CD8+CRTAM+ and CD8+CX3CR1+ populations.

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 much attention due to their potential in cell-based therapies. Their accurate identification in single-cell studies remains a challenge due to their transcriptional similarity to cytotoxic αβ T cells and NK cells. We manually annotated a γδ T cell cluster expressing *ITGAD* encoding CD11d, which as part of a heterodimer with CD18 has the potential to interact with VCAM1(*34*). In the liver and the spleen, where we identified these γδ T cells, resident macrophages, such as red pulp macrophages and Kupffer cells, express high levels of *VCAM1* (**Fig. 2B, Fig. 5B**). We hypothesized that these γδ T cells may be interacting with resident macrophages and used smFISH to explore this possibility. Imaging of liver sections incubated with *TRDC* and *ITGAD* probes to identify *ITGAD*+ γδ T cells and *CD5L* probes to mark Kupffer cells validated the presence of these populations and showed a number of *ITGAD*+ γδ T cells in the vicinity of Kupffer cells (**Fig. 5B**). We also performed γδ TCR sequencing in selected spleen samples and found that productive γδ TCR were primarily part of the *ITGAD*-expressing γδ T cells (**fig. S5A**), further supporting the robust identification of these cells. Interestingly, CellTypist misclassified this population as cytotoxic-T cells and NK cells, but predicted a different group of cells within the CD8+CCR9+ cluster as γδ T cells (**Fig. 5C, fig. S5B**), primarily derived from the gut. While both γδ T cell populations were positive for *TRDV1*, they presented differences in the expression of specific genes such as *ITGAD, CCR9, ITGAE* and *ITGA1* (**Fig. 5D**). Altogether, this 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 identity.

Discussion

Here, we present the first multi-donor study of immune cells across the human body. We sampled multiple organs from individuals, which allowed for a more robust control of age, gender, medical history and sampling backgrounds, and consequently revealed tissue-specific expression patterns.

Achieving an accurate assignment of detailed cell identities in scRNA-seq experiments is a major challenge. We have developed CellTypist, a framework for automated cell type annotation of immune

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

populations. We obtain accurate assignments both for major cell types, but also for fine-grained annotation that is typically time-consuming and requires expert knowledge. This has been largely possible due to the curation of a collection of 22 studies across a range of tissues with an in-depth immune cell analysis. Still, manual curation of cell clusters can reveal specific subtypes of cells that may be absent from the database/training set and we showcase an example on the γδ T cell compartment. To address this in the longer run, the CellTypist models will be periodically updated and extended to non-immune and further immune sub-populations as more data becomes available in the future.

Using CellTypist combined with manual annotation, we dissected the transcriptomic features of immune cells across tissues. Within the myeloid compartment, macrophages showed the most prominent features of tissue-specificificity. It is well established that macrophages seed tissues during development in mice (*35*) and recently direct evidence has also been reported in humans (*36*). We here detect several of the resident macrophage subtypes such as red pulp macrophages, Kupffer cells and related iron-recycling macrophages in bone marrow and lymph nodes. Cross-tissue characterization of activated dendritic cells revealed specific expression of *CRLF2, CSF2RA* and *GPR157* in the lung and *AIRE* in the thoracic lymph nodes. Transient expression of *AIRE* by aDCs outside of the thymus has been reported (*37*), however, its functional implications remain unclear. Moreover, signaling via TSLPR, encoded by *CRLF2,* has been shown to play a role in allergy and asthma (*38*, *39*).

In the lymphoid compartment, we combined transcriptome and VDJ analysis, which allowed the phenotypic dissection of adaptive immune cells using complementary layers of cell identity. Of note, we detected a subset of memory B cells expressing *ITGAX* and *TBX21*, that resembles populations reported to expand upon malaria vaccination (*40*) and in systemic lupus erythematosus (SLE) patients (*41*). In our data, these B cells do not show clonal expansion, suggesting that they may be present at low levels in healthy individuals. BCR analysis revealed isotype usage bias towards IgA2 in gut plasma cells, which may be related to structural differences (*42*) or higher resistance to microbial degradation as compared to IgA1 (*43*). Cross-tissue clonal distribution of B cells and plasma cells has been previously shown to follow two different patterns (*31*). In our study, due to limited sampling depth in the B cell compartment, we primarily detect shared clones between the spleen and lymphoid organs but can also detect a degree of sharing with non-lymphoid tissues including the gut. Notably, when we incorporate cell identity and clonotype analysis we can distinguish that broadly distributed clones are geared towards the B cell memory phenotype while tissue-restricted clones are more frequently plasma cells. In the T-cell compartment, cross-tissue and cross-cell type TCR sharing reveals until now 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 TEM to TRMs. Lastly, using a combination of automated annotation, manual curation

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and γδ TCR sequencing, we found two distinct subsets of γδ T cells showing distinct marker genes and tissue distributions.

Overall, this study unraveled previously unrecognized features of tissue-specific immunity in the myeloid and lymphoid compartment and provides a comprehensive framework for future cross-tissue cell type analysis. Further investigation is needed to determine the effect of important covariates such as donor age and gender as well as considering the immune cell activation status, to gain a defining picture of how human biology influences immune function.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

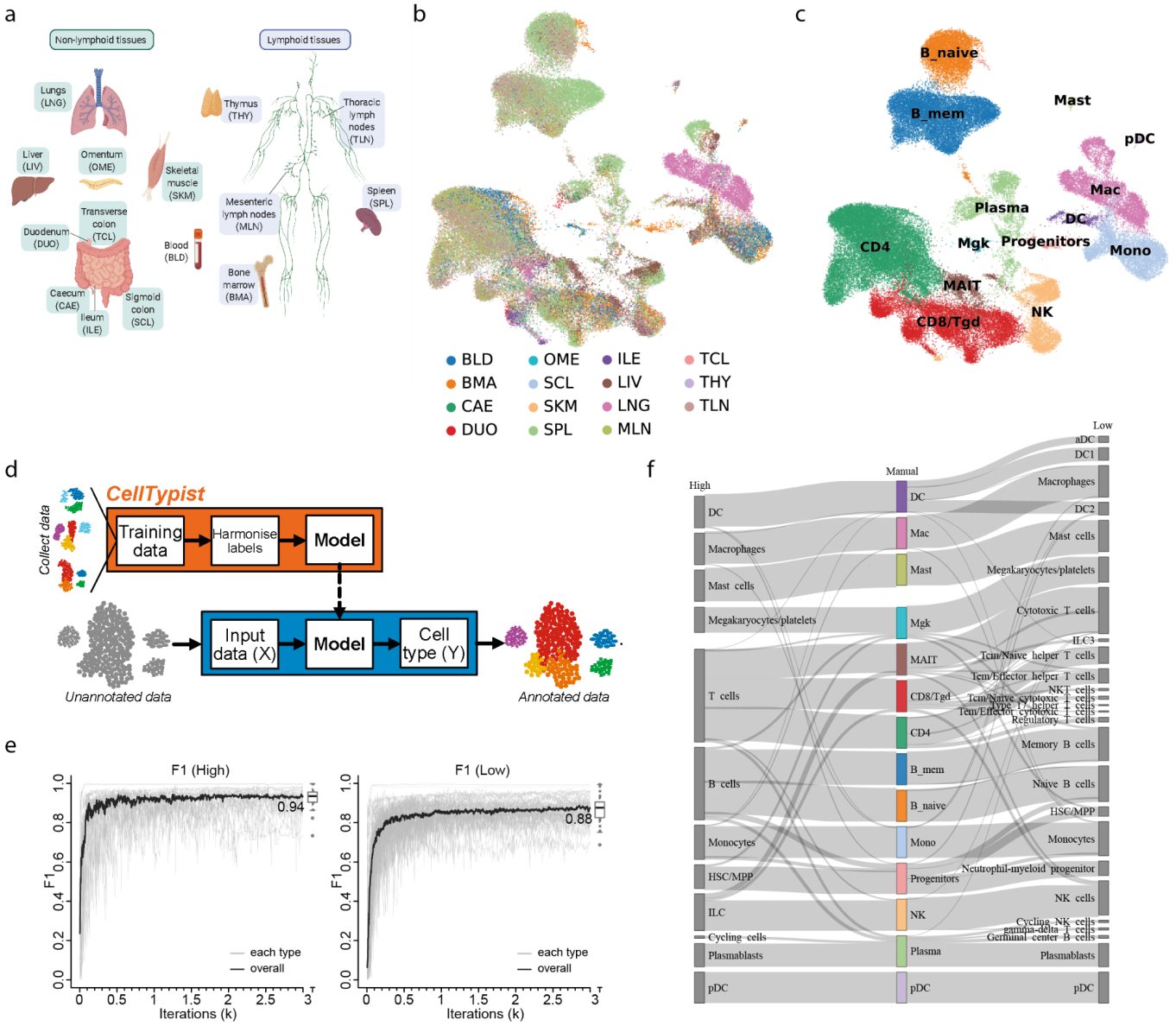
Acknowledgements

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

This research was supported by funding from the European Research Council (grant no. 646794 ThDEFINE to S.A.T.) and the Wellcome Trust (grant no. RG79413 to J.J.). T.G. was supported by the European Union's H2020 research and innovation program "ENLIGHT-TEN" under the Marie Sklodowska-Curie grant agreement 675395. 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 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. LBJ and JLJ were funded directly by the Wellcome Trust (RG79413).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

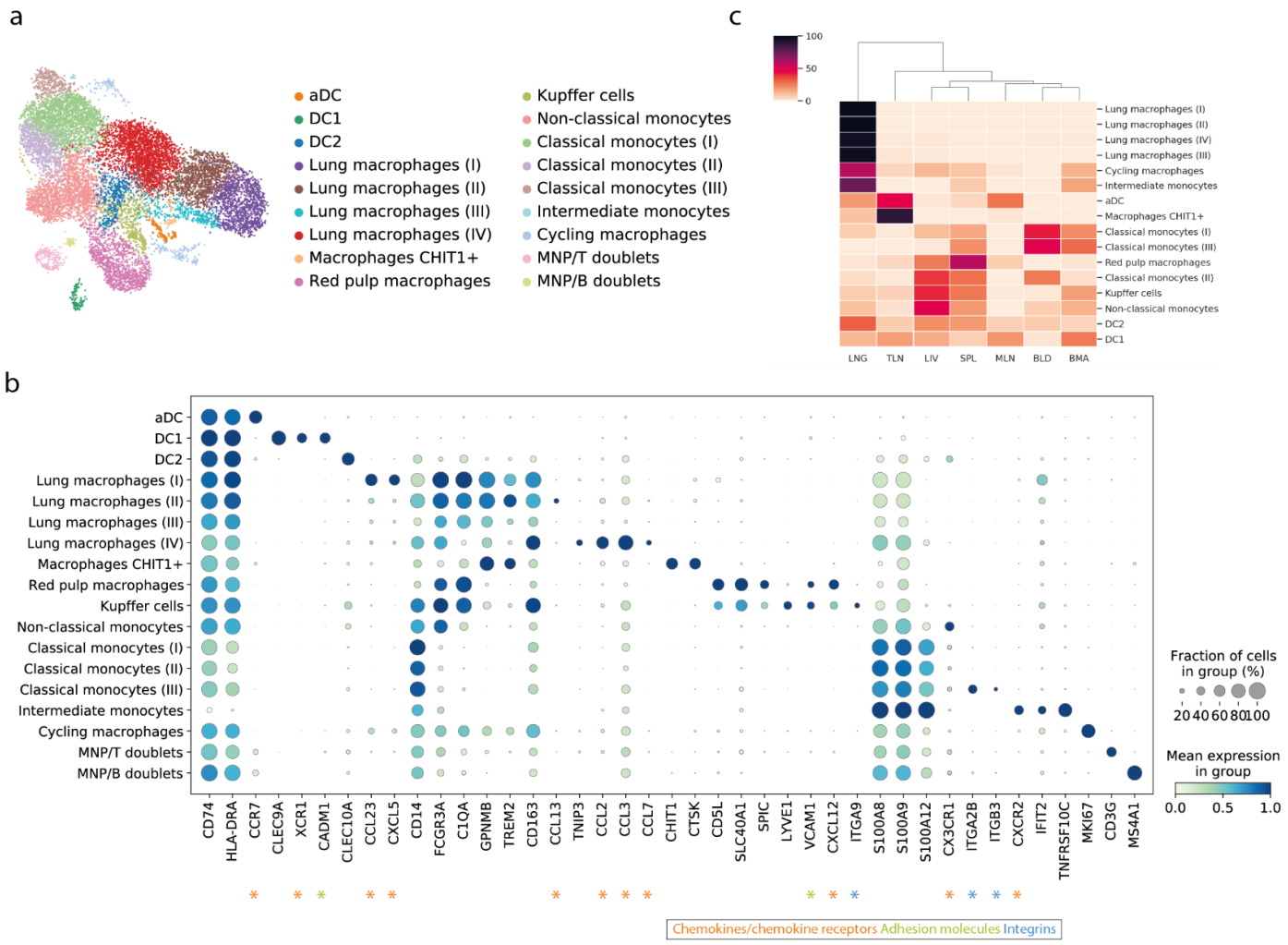
Figures and figure legends



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 1. Automated annotation of immune cells across human tissues using CellTypist. (**A**) Schematic of single-cell transcriptome profiling of human lymphoid and non-lymphoid tissues and their assigned tissue name acronyms. (**B**) UMAP visualization of the immune cell compartment colored by tissues. (**C**) UMAP visualization of the immune cell compartment colored by cell types. (**D**) Workflow of CellTypist including data collection, model training and cell type prediction. (**E**) 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).

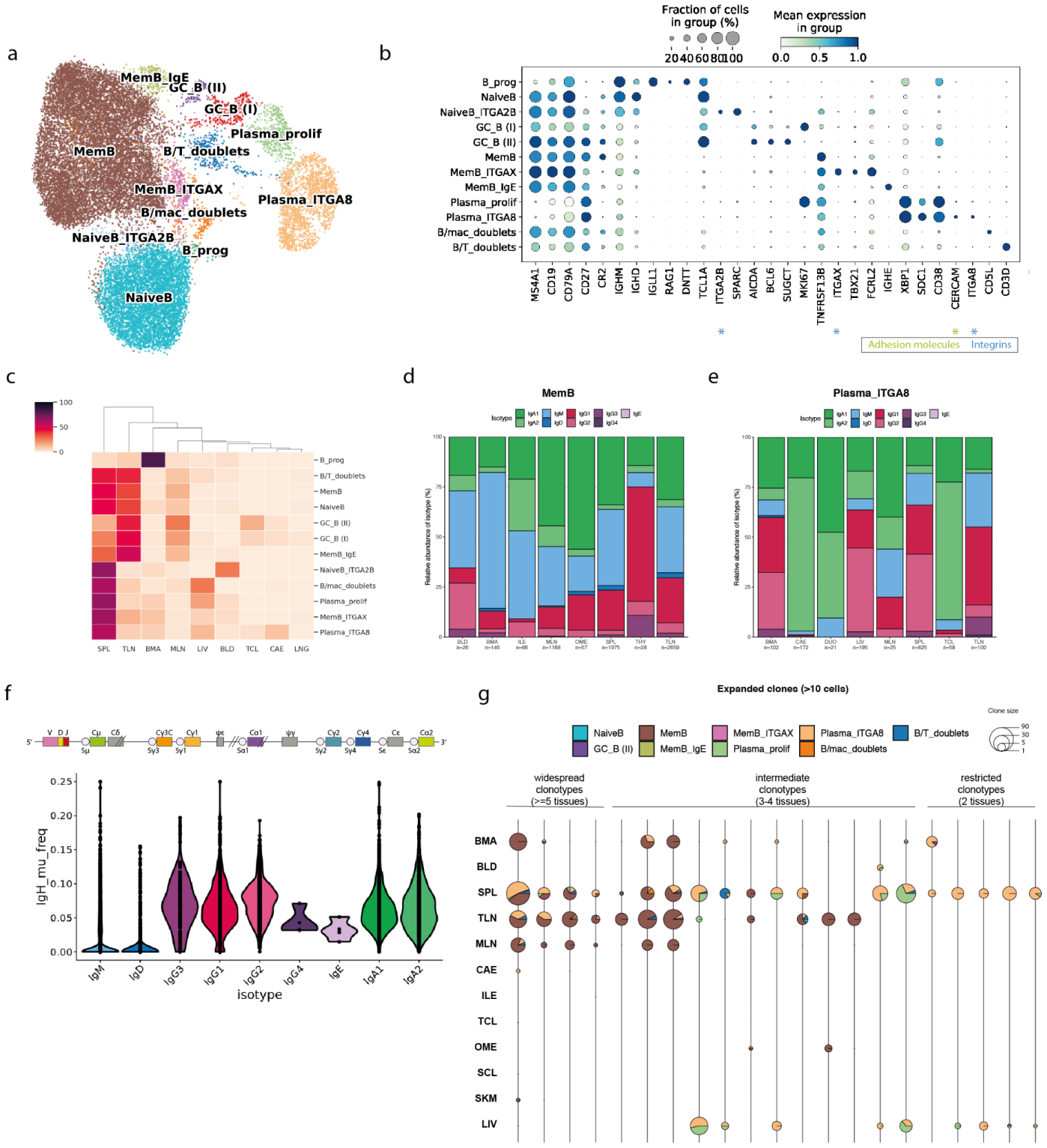
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 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. (**C**) Clustermap showing the distribution of each myeloid cell population across the different tissues. (**D**) Violin plot for genes differentially expressed in activated dendritic cells across tissues. Color represents maximum-normalized mean expression of cells expressing marker genes. (**E**) Schematic illustration of tissue-specific features of activated dendritic cells.

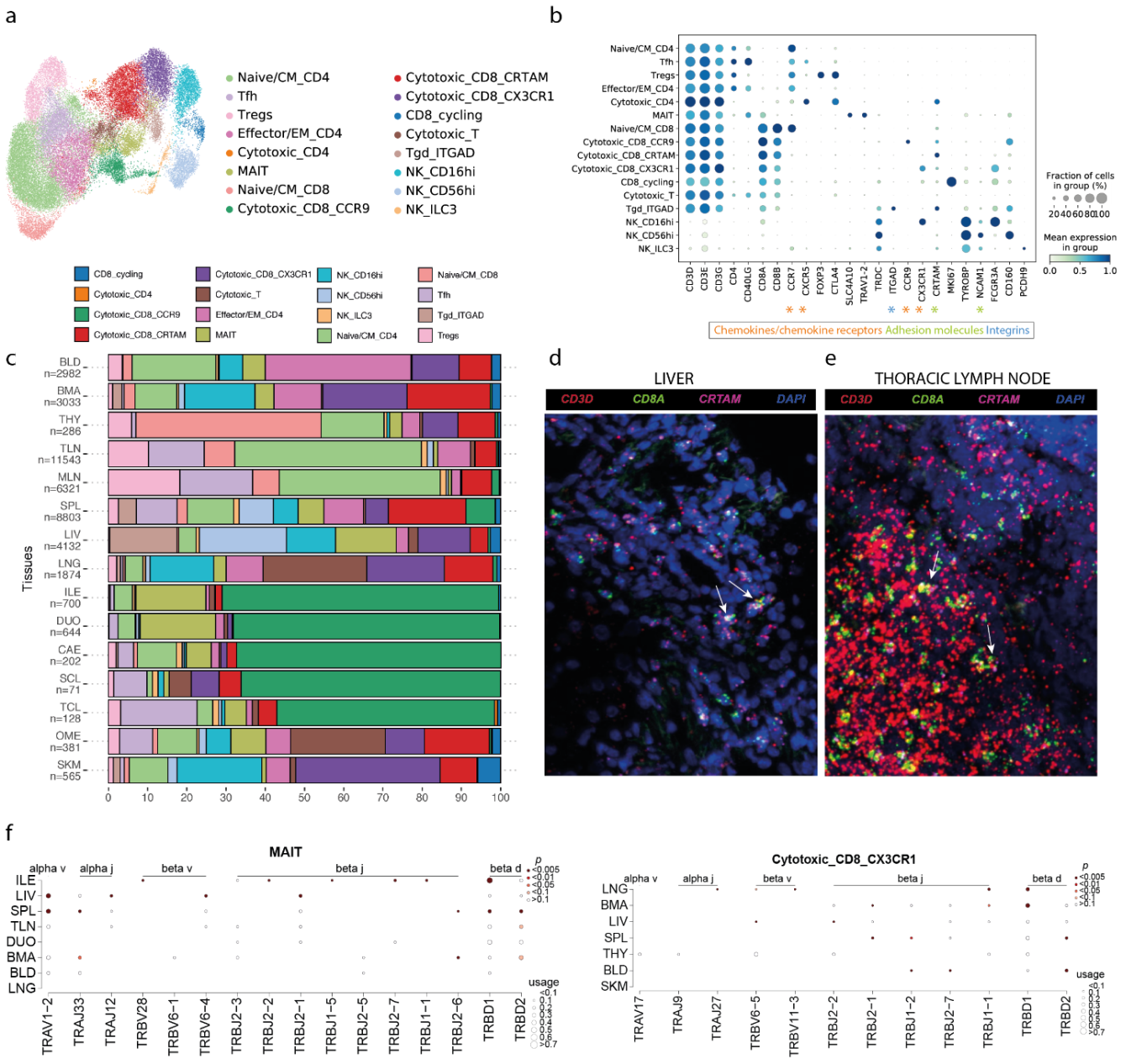
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 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. (**C**) Clustermap showing the distribution of each B cell population across the different tissues. (**D**) Stacked bar plot showing the isotype distribution per tissue within the memory B cells cluster and (**E**) the plasma cells cluster. (**F**) Violin plot of hypermutation frequency on the IgH chain across isotypes. (**G**) 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 distribution.

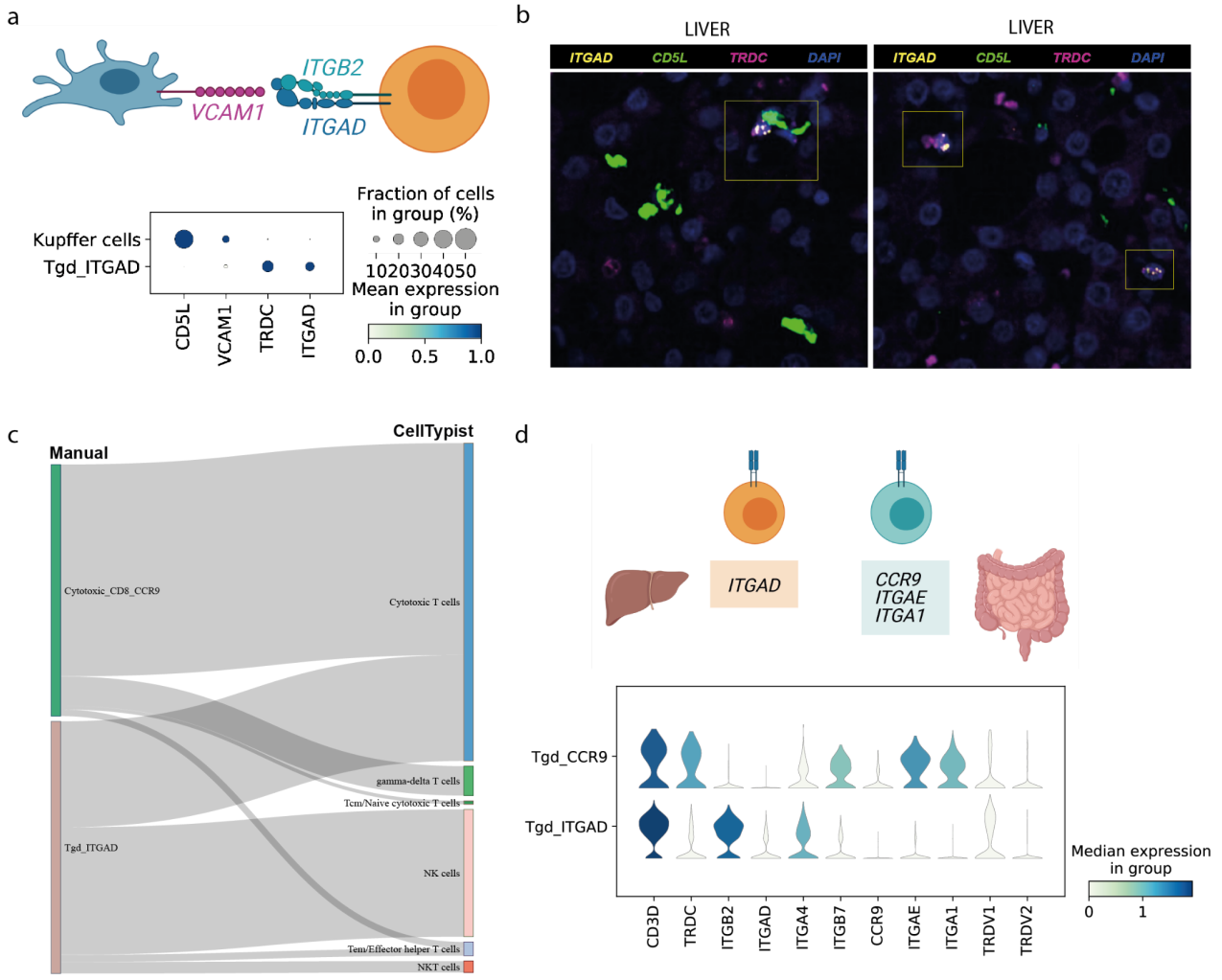
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 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 type. (**B**) Dot plot for expression of marker genes of the identified immune populations. Color represents maximum-normalized mean expression of cells expressing marker genes. (**C**) Stacked barplot showing the distribution of each T cell or ILC population across the different tissues. (**D**,**E**) smFISH visualisation of *CD3D, CD8A* and *CRTAM* transcripts in a liver tissue section, validating this tissue-resident memory CD8+ T cell population in the liver and thoracic lymph node. (**F**) Dot plots denoting the TRA and TRB V(D)J usage across tissues for MAIT cells (left) and *CX3CR1*+ CD8+ T cells (right). 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 distribution.

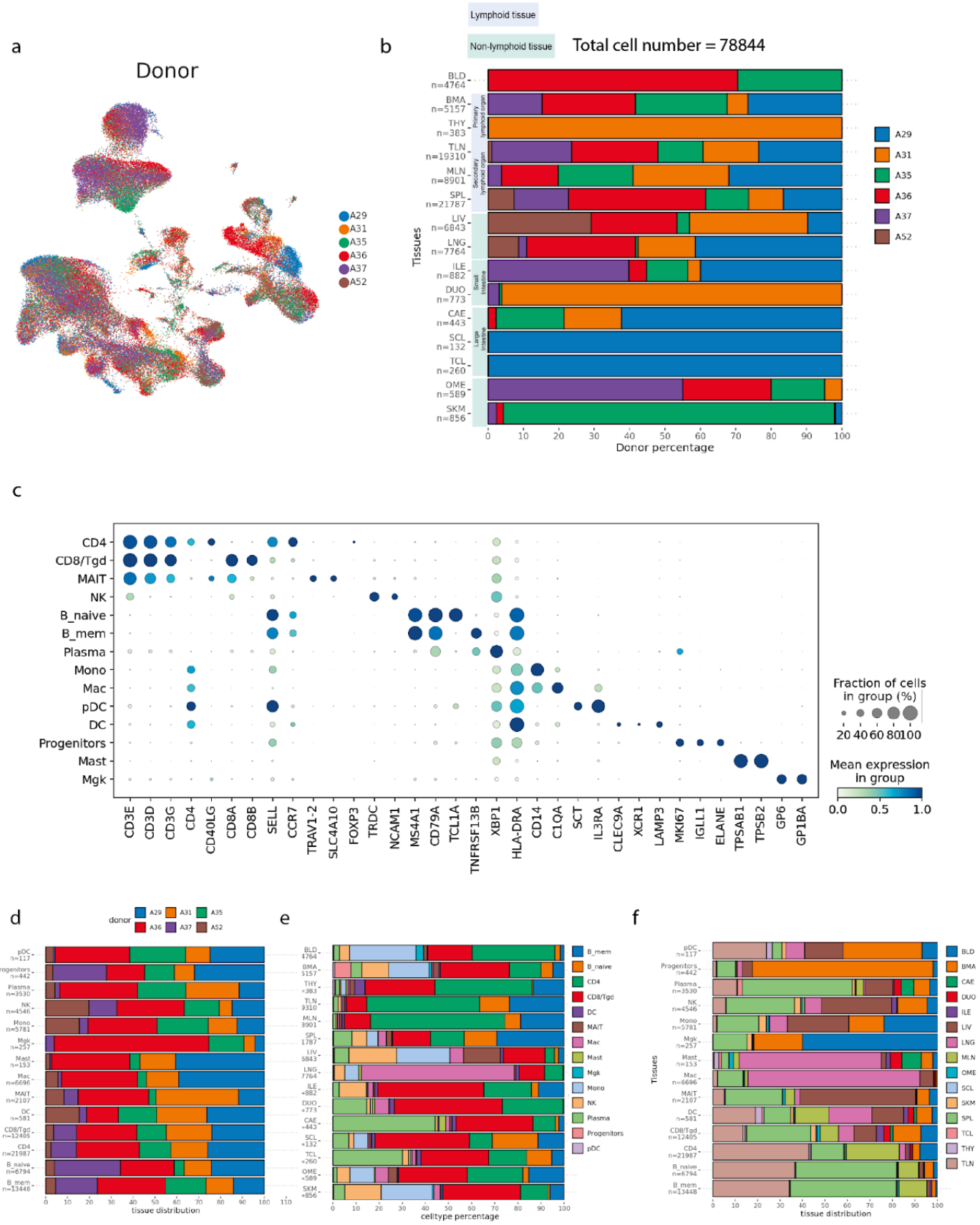
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 5. Two distinct subsets of γδ T cells across human tissues. (**A,B**) smFISH visualisation of *CD5L* (marking Kupffer cells), *TRDC* and *ITGAD* transcripts in a liver tissue section, supporting the existence of these cell populations and their interaction. (**C**) UMAP visualization of the γδTCR sequencing. (**D**) Stacked barplot showing proportion of T and ILC subsets according to γδTCR chain pairing information. (**E**) *CCR9*-expressing γδ T cells identified using CellTypist. (**F**) Dot plot for expression of marker genes of the identified γδ T cell subsets. Color represents maximum-normalized mean expression of cells expressing marker genes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



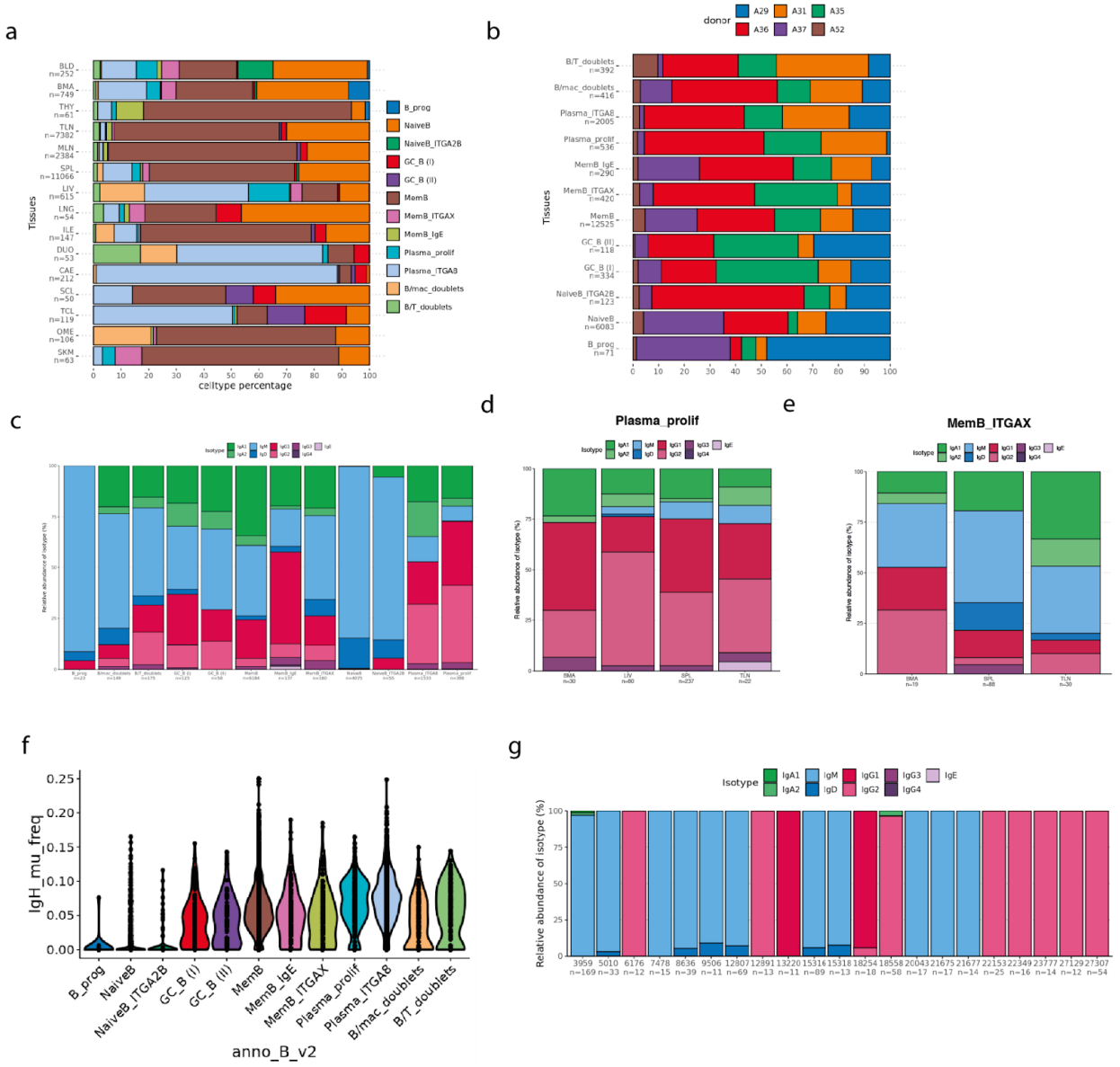
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplementary Figure 1. (A,B) UMAP visualization showing the tissue and donor distribution. **(C)** Stacked barplots showing the number and percentage of immune cell types per donor. **(D)** Dot plot for expression of marker genes of the identified immune populations. Color represents maximum-normalized mean expression of cells expressing marker genes. **(E,F)** Stacked barplots showing the number and percentage of immune cell types per tissue and cells from a given cell type across tissues.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

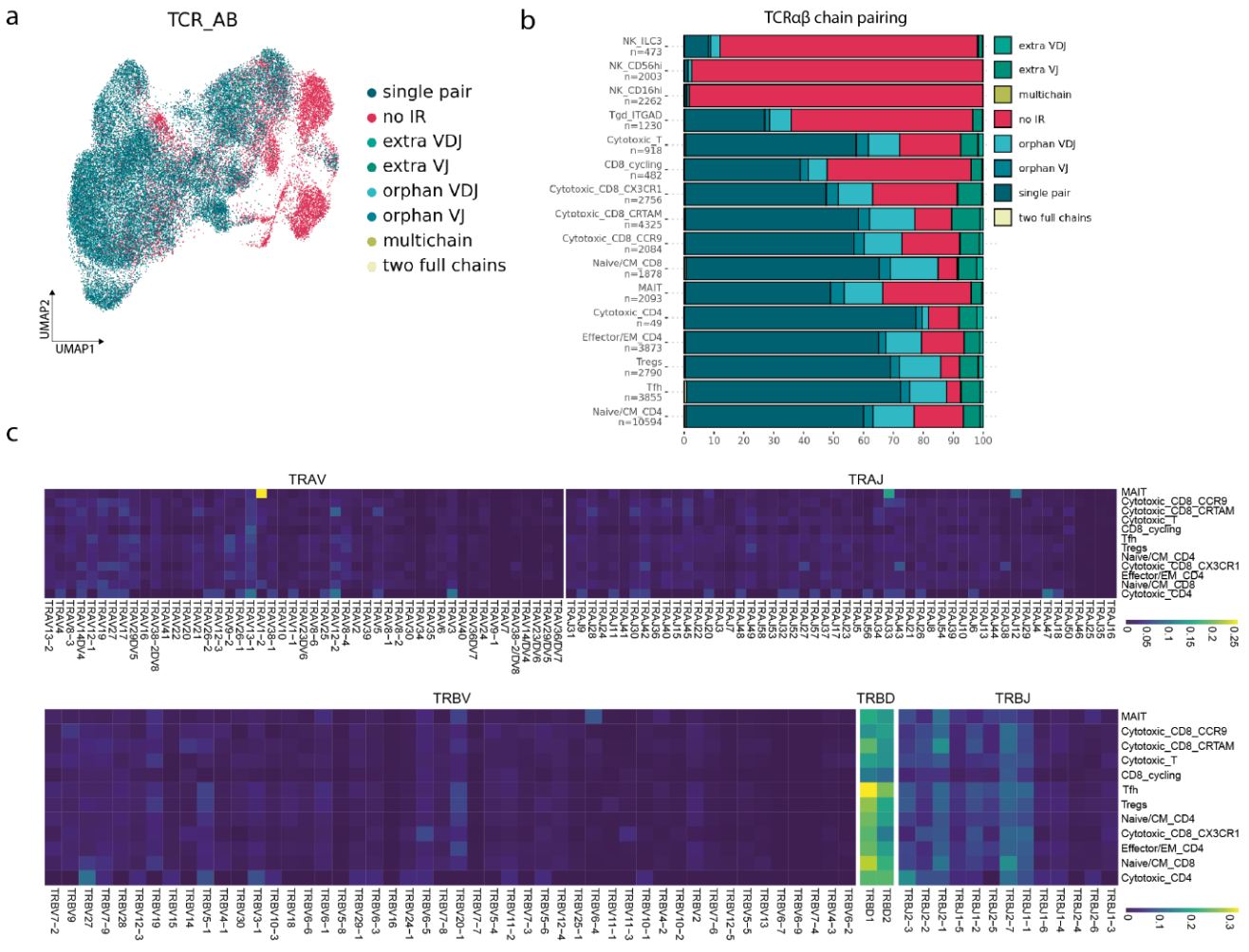
Supplementary Figure 2. Stacked barplots showing the proportion of myeloid cells obtained (A) per tissue and cellular subtype and (B) per cellular subtype and donor. The number of cells per category is identified on the left side of the plot.

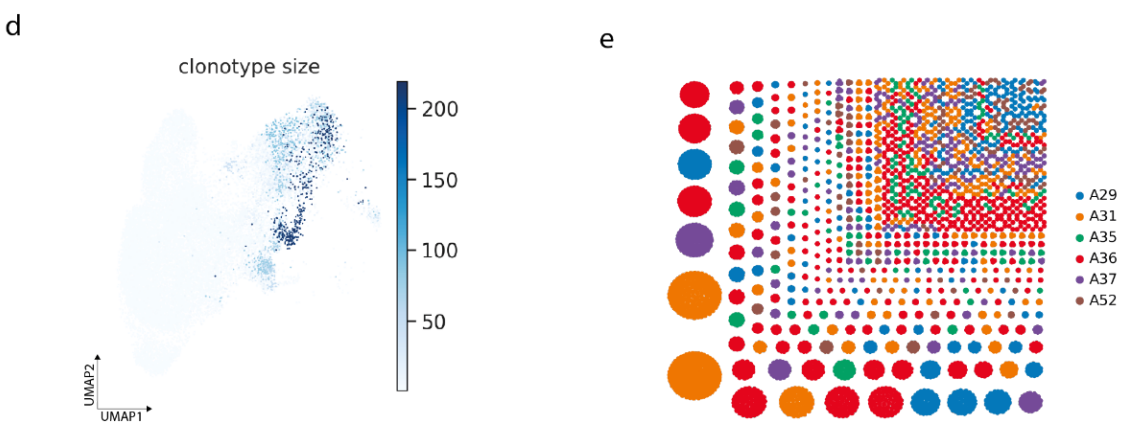
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

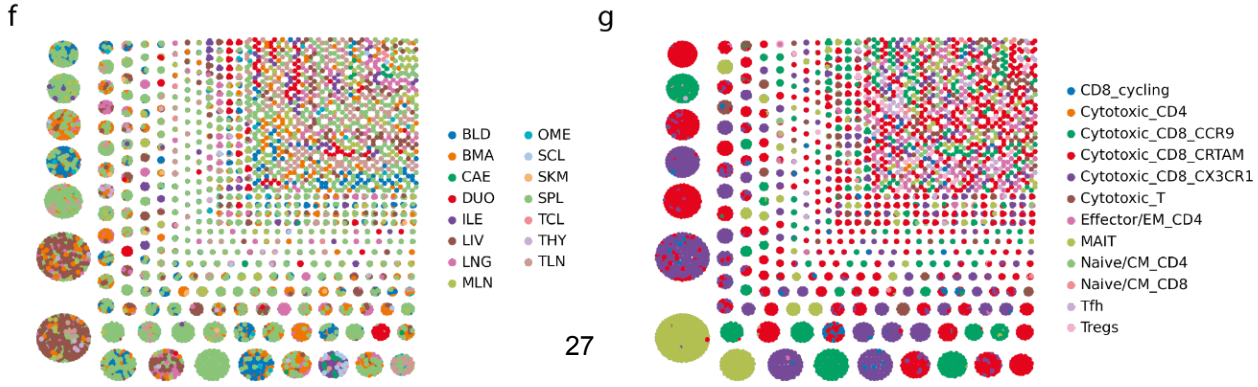


401 **Supplementary Figure 3.** (**A**) Stacked barplot showing the proportion of B cell subtypes by tissue. (**B**)
402 Stacked barplot showing the donor distribution of each B cell subtypes. Stacked bar plot showing the
403 isotype distribution across B cell subsets (**C**), within the plasmablasts across tissues (**D**) and within the
404 *ITGAX*+ memory B cells across tissues (**E**). (**F**) Violin plot showing hypermutation frequency across b
405 cell subsets. (**G**) Isotype distribution across 21 expanded clonotypes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



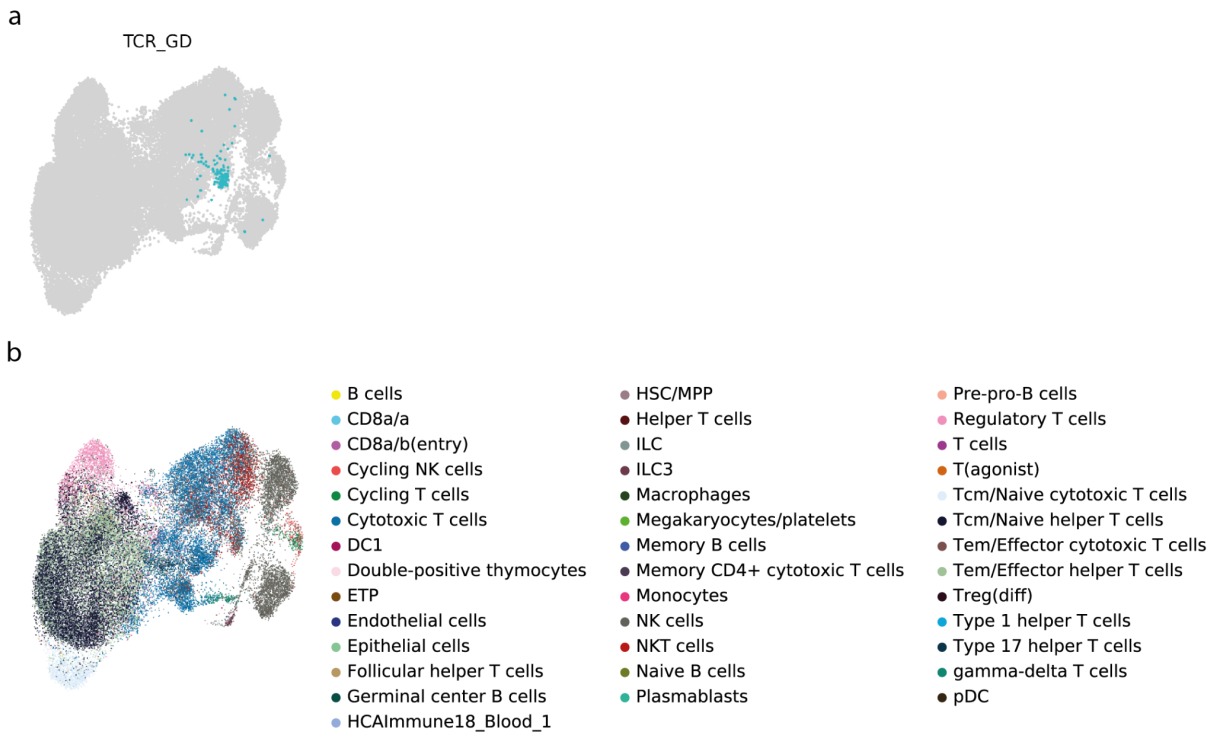




bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplementary Figure 4. (**A**) UMAP showing TCRab chain pairing. (**B**) Proportion of TCRαβ chain pairing by subset. (**C**) Heatmap showing relative usage of V(D)J genes across T cell subsets. (**D**) UMAP showing clonal expansion of the TCRab T cells. (**E, F, G**) Clonotype network color coded by donor, tissue and T cell subset.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



Supplementary Figure 5. (**A**) UMAP showing cells with a pair of productive γδ TCR chains. Data from two spleen samples. (**B**) UMAP showing the cell type labels predicted by CellTypist on the T and ILC compartment.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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)	Antibiotics within 2 weeks of death	Steroids
A29	F	65-70	ICH	Y	2	30-35	CMV+/EBV+/TOXO-	N/K	<1	Nil	N
A31	M	50-55	ICH	Y	8	30-35	CMV+/EBV+/TOXO-	Y	<1	Co-amoxiclav, Tazocin	N
A35	M	60-65	ICH	N	2	20-25	CMV-/EBV+/TOXO-	Y	>9	Gentamicin, Flucloxacillin	Y Dexamethasone
A37	F	55-60	ICH	N	3	20-25	CMV-/EBV+/TOXO-	Y	>9	Co-amoxiclav	N
A36	M	70-75	ICH	N	5	25-30	CMV-/EBV+/TOXO+	Y	<2	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	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

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 37ºC 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 5000 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 1000 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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Doublet detection

Doublet detection was performed using the scrublet algorithm (https://github.com/AllonKleinLab/scrublet,(*44*)) with percolation as previously described in a per sample basis(*45*). Briefly, scrublet scores were obtained per cell. The percolation step was performed on over-clustered data using the *scanpy.tl.louvain* function from the scanpy package. Scrublet was run, obtaining per-cell doublet scores. The standard Seurat-inspired Scanpy processing pipeline was performed up to the clustering stage, using default parameters. 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. Clusters with a scrublet score above 0.6 were flagged as potential doublet clusters and removed from further downstream biological 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 (scanpy.pp.filter_gene_dispersion), data feature scaling (scanpy.pp.scale), PCA analysis (scanpy.pp.pca, from variable genes), batch-balanced neighbourhood graph building (scanpy.pp.bbknn) and Louvain graph-based clustering (scanpy.tl.leiden, clustering resolution manually adjusted) performed using the python package scanpy (version 1.6.0). 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

CellTypist

Full details on the data collection, processing, curation, model training, and testing of the CellTypist pipeline can be found in the **Supplementary Note**.

scTCR-seq downstream analysis

VDJ sequence information was extracted from the output file "filtered_contig_annotations.csv" using the scirpy package(*46*). 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 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. 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*, *47*), with all sequences from a given patient grouped together for analysis. AssignGenes.py(*48*) and IgBLAST(*49*) were used to reannotate IgH sequences prior to correction of ambiguous V gene assignments using TIgGER (v1.0.0)(*50*). Clonally-related IgH sequences were identified using DefineClones.py with a nearest neighbour distance threshold of 0.15 before running CreateGermlines.py (ChangeO)(*51*) to infer germline sequences for each clonal family and calculate somatic hypermutation frequencies with observedMutations (Shazam)(*51*). IgH diversity analyses were performed using the rarefyDiversity and testDiversity of Alakazam (v1.0.2; (*51*)). 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Single molecule FISH

Samples were either snap frozen in chilled isopentane (-40^{0}C for striated muscle, -70^{0}C 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), Hs-TRDC (433678), Hs-ITGAD-C2 (881498-C2), Hs-CD5L-C3 (850518-C3). 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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. M. D. Young, T. J. Mitchell, F. A. Vieira Braga, M. G. B. Tran, B. J. Stewart, J. R. Ferdinand, G. Collord, R. A. Botting, D.-M. Popescu, K. W. Loudon, R. Vento-Tormo, E. Stephenson, A. Cagan, S. J. Farndon, M. Del Castillo Velasco-Herrera, C. Guzzo, N. Richoz, 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, P. H. Maxwell, S. Lindsay, A. Y. Warren, G. D. Stewart, N. Sebire, N. Coleman, M. Haniffa, S. A. Teichmann, M. Clatworthy, S. Behjati, Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science*. **361**, 594–599 (2018).

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. 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).

24. M. Di Rosa, G. Malaguarnera, C. De Gregorio, F. D'Amico, M. C. Mazzarino, L. Malaguarnera, Modulation of chitotriosidase during macrophage differentiation. *Cell Biochem. Biophys.* **66**, 239–247 (2013).

25. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

genotype and smoking habit. *J. Allergy Clin. Immunol.* **122**, 944–950.e3 (2008).

26. 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).

27. 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).

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

29. 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).

30. 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).

31. 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).

32. E. Hammarlund, A. Thomas, I. J. Amanna, L. A. Holden, O. D. Slayden, B. Park, L. Gao, M. K. Slifka, Plasma cell survival in the absence of B cell memory. *Nat. Commun.* **8**, 1781 (2017).

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

34. 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).

35. E. Mass, I. Ballesteros, M. Farlik, F. Halbritter, P. Günther, L. Crozet, C. E. Jacome-Galarza, K. Händler, J. Klughammer, Y. Kobayashi, E. Gomez-Perdiguero, J. L. Schultze, M. Beyer, C. Bock, F. Geissmann, Specification of tissue-resident macrophages during organogenesis. *Science*. **353** (2016), doi:10.1126/science.aaf4238.

36. Z. Bian, Y. Gong, T. Huang, C. Z. W. Lee, L. Bian, Z. Bai, H. Shi, Y. Zeng, C. Liu, J. He, J. Zhou, X. Li, Z. Li, Y. Ni, C. Ma, L. Cui, R. Zhang, J. K. Y. Chan, L. G. Ng, Y. Lan, F. Ginhoux, B. Liu, Deciphering human macrophage development at single-cell resolution. *Nature*. **582**, 571–576 (2020).

37. J. R. Fergusson, M. D. Morgan, M. Bruchard, L. Huitema, B. A. Heesters, V. van Unen, J. P. van Hamburg, N. N. van der Wel, D. Picavet, F. Koning, S. W. Tas, M. S. Anderson, J. C. Marioni, G. A. Holländer, H. Spits, Maturing Human CD127+ CCR7+ PDL1+ Dendritic Cells Express AIRE in the Absence of Tissue Restricted Antigens. *Front. Immunol.* **9**, 2902 (2018).

38. 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).

39. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.28.441762; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

trigger dendritic cell mediated allergic inflammation by producing TSLP. *Nat. Immunol.* **3**, 673–680 (2002).

40. 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).

41. 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).

42. 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).

43. 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).

44. 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).

45. 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).

46. 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).

47. 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.

48. 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.

49. 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).

50. 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).

51. 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).