

# Single-cell analysis of human MAIT cell transcriptional, functional and clonal diversity

Received: 11 January 2022

Accepted: 26 June 2023

Published online: 14 August 2023

 Check for updates

Lucy C. Garner<sup>1</sup>✉, Ali Amini<sup>1</sup>, Michael E. B. FitzPatrick<sup>1</sup>, Martin J. Lett<sup>2</sup>, Gabriel F. Hess<sup>3</sup>, Magdalena Filipowicz Sinnreich<sup>2,4</sup>, Nicholas M. Provine<sup>1,7</sup> & Paul Klenerman<sup>1,5,6,7</sup>✉

Mucosal-associated invariant T (MAIT) cells are innate-like T cells that recognize microbial metabolites through a semi-invariant T cell receptor (TCR). Major questions remain regarding the extent of human MAIT cell functional and clonal diversity. To address these, we analyzed the single-cell transcriptome and TCR repertoire of blood and liver MAIT cells and developed functional RNA-sequencing, a method to integrate function and TCR clonotype at single-cell resolution. MAIT cell clonal diversity was comparable to conventional memory T cells, with private TCR repertoires shared across matched tissues. Baseline functional diversity was low and largely related to tissue site. MAIT cells showed stimulus-specific transcriptional responses *in vitro*, with cells positioned along gradients of activation. Clonal identity influenced resting and activated transcriptional profiles but intriguingly was not associated with the capacity to produce IL-17. Overall, MAIT cells show phenotypic and functional diversity according to tissue localization, stimulation environment and clonotype.

Mucosal-associated invariant T (MAIT) cells are innate-like T cells, abundant in human blood and tissues, particularly the liver and mucosa<sup>1</sup>. MAIT cells express semi-invariant V $\alpha$ 7.2-J $\alpha$ 33/12/20 (*TRAV1-2-TRAJ33/12/20*) T cell receptors (TCRs) specific for microbial riboflavin metabolites presented by MRI (ref. 1). They can also be activated independent of their TCR by cytokines such as IL-12 and IL-18 (ref. 2). Upon activation, MAIT cells secrete type 1/17 cytokines and exhibit cytotoxic activity<sup>1</sup>.

A major outstanding question is whether human MAIT cells comprise transcriptionally and functionally distinct subsets. Alterations in frequency, phenotype and function occur in numerous human diseases, and mouse models indicate protective and pathogenic roles<sup>3</sup>. Understanding the characteristics of the MAIT cell population in health

could aid the development of therapeutics targeting specific subsets or functions in disease.

In human blood, MAIT cells are relatively homogeneous, exhibiting a predominantly CD8<sup>+</sup> effector-memory phenotype, and characteristic expression of surface molecules (for example, CD161) and transcription factors (for example, PLZF and ROR $\gamma$ t)<sup>1</sup>. However, there is some variability, for example, between CD8<sup>+</sup>, CD4<sup>-</sup>CD8<sup>-</sup> (DN) and CD4<sup>+</sup> cells<sup>4-6</sup>, and in the expression of innate immune receptors<sup>6,7</sup>. Despite universal ROR $\gamma$ t expression, <5% of human MAIT cells produce IL-17 *ex vivo*<sup>8,9</sup>. This could reflect a committed type 17 subset, as in mice<sup>10-12</sup>. Alternatively, all human MAIT cells may have the capacity to produce IL-17 under appropriate conditions. Conclusive data addressing these competing hypotheses are lacking.

<sup>1</sup>Translational Gastroenterology Unit, Nuffield Department of Medicine, University of Oxford, Oxford, UK. <sup>2</sup>Department of Biomedicine, Liver Immunology, University Hospital Basel and University of Basel, Basel, Switzerland. <sup>3</sup>Division of Visceral Surgery, Clarunis University Center for Gastrointestinal and Liver Diseases, Basel, Switzerland. <sup>4</sup>Gastroenterology and Hepatology, University Department of Medicine, Cantonal Hospital Baselland, Liestal, Switzerland.

<sup>5</sup>Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, UK. <sup>6</sup>NiHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK. <sup>7</sup>These authors contributed equally: Nicholas M. Provine, Paul Klenerman. ✉e-mail: [lucy.garner@ndm.ox.ac.uk](mailto:lucy.garner@ndm.ox.ac.uk); [paul.klenerman@ndm.ox.ac.uk](mailto:paul.klenerman@ndm.ox.ac.uk)

MAIT cell function is altered by tissue localization and stimulation. Compared with blood, gut and liver MAIT cells display an activated, tissue-resident transcriptome<sup>12–14</sup>. Genital tract<sup>15</sup> and oral mucosal<sup>16</sup> cells show type 17 skewing. MAIT cells exhibit distinct transcriptional responses to TCR and cytokine stimulation<sup>17,18</sup> and produce increased IL-17 upon sustained stimulation<sup>9</sup>. Whether functional diversity indicates the presence of multiple subsets or environment-driven plasticity remains unknown.

In addition, questions remain regarding MAIT cell TCR repertoires, including variability across tissues and donors, and the relationship between TCR usage and function. Studies variably demonstrate similar TCR repertoires across tissues<sup>19,20</sup> or differences in *TRAJ/TRBV* usage<sup>16,21,22</sup>. Diverse chain usage could have functional implications. For example, MAIT cell clonal distribution changes during human *Salmonella* Paratyphi A infection—cells transduced with TCR $\beta$  chains from expanded and contracted clonotypes show greater and lesser responses to TCR stimulation, respectively<sup>23</sup>. In vitro studies show differential activation potential dependent on clonotype<sup>24</sup> or TCR $\beta$  usage<sup>7,25</sup>. Thus, a relationship between TCR architecture and function is suggested but has not been studied systematically.

Overall, human MAIT cells show variation in phenotype, function and TCR repertoire. However, it is unknown whether they comprise multiple functionally distinct subsets, and how phenotype and function relate to TCR usage. To investigate this, we analyzed the single-cell transcriptome and TCR repertoire of human MAIT cells from matched blood and liver, as well as blood cells at rest and following TCR, cytokine or dual TCR+cytokine, stimulation. Our findings revealed a largely homogenous transcriptional program at rest, with variation linked to tissue localization. Activation triggered a plastic, stimulus-specific, effector program. The TCR repertoire was surprisingly diverse, and clonal identity influenced the transcriptome of resting and activated MAIT cells. Following dual stimulation, we identified an IL-17-expressing cluster. IL-17<sup>+</sup> cells expressed other effector molecules, such as *IFNG*, and showed similar TCR usage to IL-17<sup>-</sup> cells, suggesting they reflect an activation state rather than a bona fide MAIT17 lineage.

## Results

### MAIT cells show tissue-specific transcription and regulation

To investigate heterogeneity using an unbiased genome-wide approach, we performed single-cell RNA-sequencing (scRNA-seq) and single-cell TCR-sequencing (scTCR-seq) of sorted MAIT cells (CD3<sup>+</sup>MRI/5-OP-RU<sup>+</sup>) from matched human blood and liver (Supplementary Table 1 and Supplementary Fig. 1). Findings from an initial experiment (Exp 1;  $n = 3$  blood, 4 liver) were validated in a second experiment (Exp 2;  $n = 8$  blood, 3 liver). Conventional memory T ( $T_{\text{mem}}$ ; CD3<sup>+</sup>MRI/5-OP-RU<sup>+</sup>CCR7<sup>-</sup>) cells were analyzed in some donors (Exp 1:  $n = 2$  blood, 2 liver; Exp 2: all samples). Exp 2 included 130 oligo-conjugated antibodies for single-cell surface protein analysis.

After filtering, Exp 1 and 2 comprised 89,456 cells. MAIT and  $T_{\text{mem}}$  cells, and blood and liver cells, were transcriptionally distinct (Fig. 1a,b). Blood T cells from liver donors (who underwent surgery for removal of benign or malignant lesions) were comparable to T cells from healthy donors (Fig. 1b), suggesting our data are not reflective of disease. CD4<sup>+</sup> $T_{\text{mem}}$  cells localized to distinct clusters, while rare CD4<sup>+</sup> MAIT cells were distributed throughout the UMAP (Extended Data Fig. 1a,b). MAIT and  $T_{\text{mem}}$  cells differentially expressed 532 and 558 genes, and 37 and 33 proteins, in blood and liver, respectively (Supplementary Table 2a–d). Differences in gene and protein expression were highly correlated between tissues (Fig. 1c and Extended Data Fig. 1c–g). We defined core MAIT cell signatures of 167 genes (including *KLRB1* and *SLC4A10*) and 11 proteins (including V $\alpha$ 7.2 TCR $\alpha$  and CD161; Supplementary Table 2e).

MAIT cells from six matched blood-liver pairs (35,407 cells) comprised 11 clusters (Fig. 1d). Clusters largely contained cells from one tissue, but multiple experiments and donors (Fig. 1e,f and Extended Data Fig. 1h,i). Blood and liver cells differentially expressed 566 genes

(most upregulated in the liver) and 24 proteins (Supplementary Table 3a,b). Liver-enriched genes encoded tissue-residency markers (for example, *ITGAE*), TCR-induced transcription factors (for example, *EGR1*), effector cytokines (for example, *IFNG*) and chemokines/chemokine receptors (for example, *CXCR6*). Some genes showed uniformly higher expression in the liver compared with the blood (for example, *CD69*; Fig. 1g); others were enriched in specific clusters (for example, *CCL3*; Fig. 1h). Interestingly, 84/167 core MAIT cell genes were upregulated in the liver (for example, *RORA* and *IL23R*). Liver-upregulated proteins included CD69 and CD244 (2B4), markers of tissue residency and cell activation (Fig. 1i and Extended Data Fig. 1j,k).  $T_{\text{mem}}$  cells showed similar tissue imprinting (Supplementary Table 3c,d). We defined core liver signatures of 300 genes and eight proteins (Supplementary Table 3e). Core liver proteins included canonical markers of tissue-resident memory T ( $T_{\text{RM}}$ ) cells<sup>26</sup> (for example, CD69 and CD103) and ICAM1, required for MAIT1 retention in mouse liver<sup>12</sup>. Using gene set enrichment analysis, we demonstrated enrichment of human and mouse  $T_{\text{RM}}$  cell gene signatures<sup>27,28</sup> in liver MAIT (Fig. 1j) and  $T_{\text{mem}}$  (Extended Data Fig. 1l) cells. Other liver-enriched pathways related to cell activation, cell adhesion and inflammation (Fig. 1k and Extended Data Fig. 1m).

SCENIC<sup>29</sup> was used to discover tissue-specific MAIT cell transcription factor regulons—modules of genes predicted to be regulated by a given transcription factor. Due to batch effects, cells from Exp 1 and Exp 2 were analyzed separately. Compared with blood, liver MAIT cells showed increased activity of AP-1 (for example, FOS and JUN) and NF- $\kappa$ B (for example, NFKB1 and NFKB2) transcription factors, and the TCR-induced transcription factor EGR1 (Fig. 1l, Extended Data Fig. 1n and Supplementary Table 4a,b). RUNX3 (regulates CD8<sup>+</sup> $T_{\text{RM}}$  cell differentiation<sup>28</sup>) activity was also increased. AP-1- and NF- $\kappa$ B-regulated genes were enriched for pathways associated with T cell activation, inflammation and cytokine production (Extended Data Fig. 1o–q).

In summary, MAIT cells in blood and liver are transcriptionally distinct. MAIT and  $T_{\text{mem}}$  cells show similar adaptation to the liver environment and MAIT cell signature genes are consistent across tissues. We identify AP-1 and NF- $\kappa$ B transcription factors as central regulators of MAIT cell liver-specific gene expression.

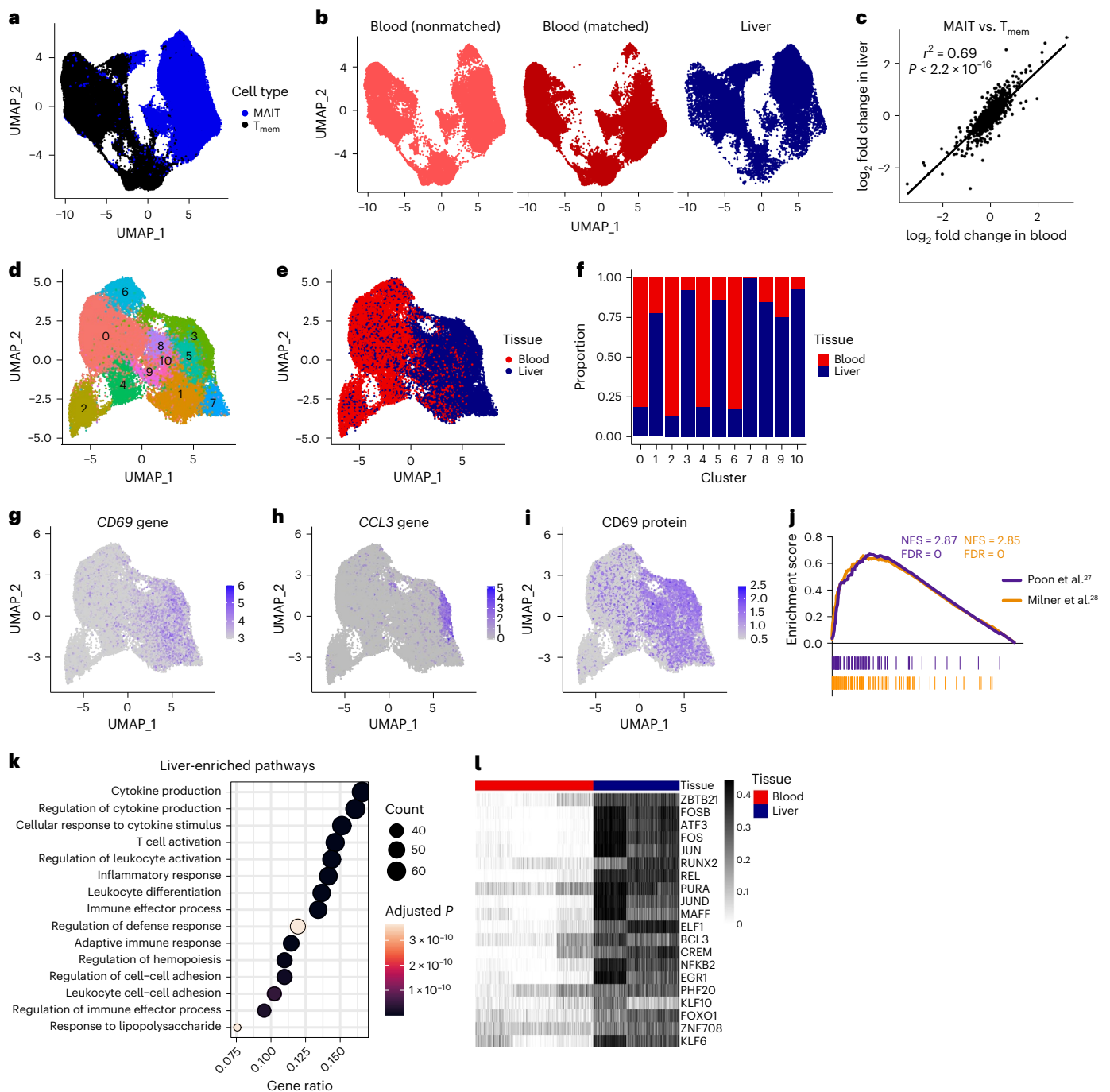
### MAIT cells have a limited TCR $\alpha$ but diverse TCR $\beta$ repertoire

We next investigated whether TCR repertoires were tissue-specific. Previous analyses were limited by scale or depth<sup>20,21,30</sup>. Our dataset of >30,000 paired TCRs from 12 donors and matched tissues provided a unique opportunity to examine TCR repertoire characteristics and diversity.

Broad characteristics of the TCR repertoire were comparable in blood and liver. *TRAJ33*, *TRAJ12* and *TRAJ20* were used by 87%, 6% and 6% of TCRs, respectively (Extended Data Fig. 2a). The CDR3 $\alpha$  region was highly restricted in length and sequence and included the canonical Tyr95 $\alpha$  residue<sup>31,32</sup> (Extended Data Fig. 2b,c). CDR3 $\alpha$  sequence, and the number of N-nucleotides and P-nucleotides, varied with *TRAJ* gene usage (Extended Data Fig. 2c–e). *TRBV* expression was diverse but biased toward *TRBV6-1*, *TRBV6-4* and *TRBV20-1* (Extended Data Fig. 2f). CDR3 $\beta$  length and sequence were highly variable (Extended Data Fig. 2g,h).

Studies of small numbers of TCR sequences suggest *TRAJ* usage could influence TCR $\alpha$  $\beta$  pairing<sup>21,30</sup>. Our data revealed increased pairing of *TRAJ12* and *TRAJ20* TCR $\alpha$  chains with *TRBV6-4* TCR $\beta$  chains compared with *TRAJ33*, while *TRAJ33* TCR $\alpha$  chains more frequently paired with *TRBV20-1* TCR $\beta$  chains (Fig. 2a).

TCR clonotypes were defined as cells with identical TCR gene segment usage and CDR3 nucleotide sequences (*TRAV1-2* TCR $\alpha$  required for MAIT cells). MAIT cells were oligoclonal, with oligoclonality comparable across donors and tissues (Fig. 2b) and with  $T_{\text{mem}}$  cells (Fig. 2c). MAIT cell clonotypes defined using only the TCR $\alpha$  chain (TCR $\alpha$  clonotypes) were more oligoclonal than those defined using



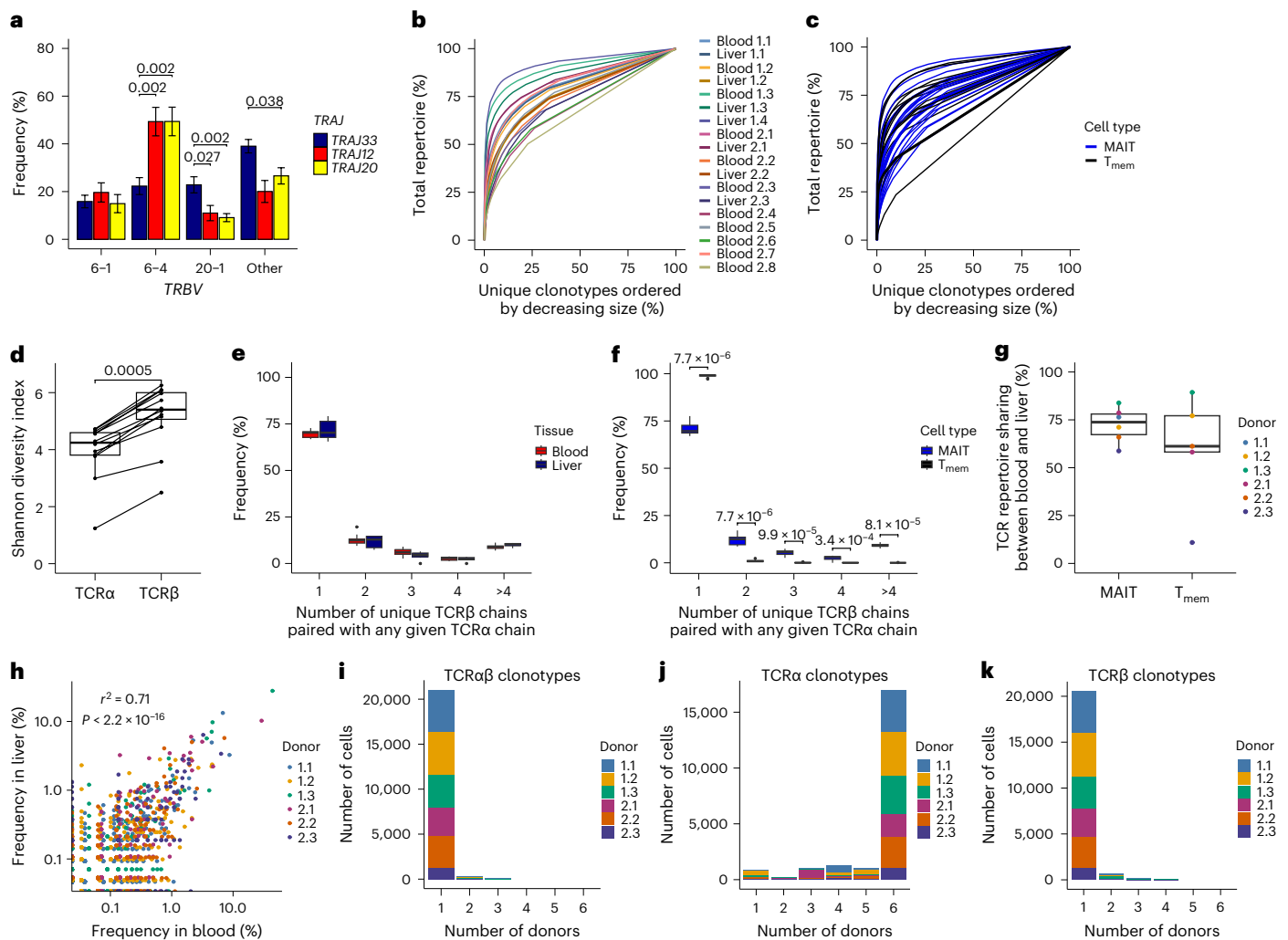
**Fig. 1 | Liver MAIT cells exhibit an activated, tissue-resident transcriptional and regulatory profile.** **a**, UMAP of blood and liver MAIT cells and conventional memory T ( $T_{\text{mem}}$ ) cells colored by cell type.  $n = 89,456$  cells from 12 donors. **b**, UMAP split by sample type, namely blood from healthy donors (nonmatched), blood from liver donors (matched) and liver. **c**, Pearson's correlation between the  $\log_2$  fold change in gene expression between MAIT and  $T_{\text{mem}}$  cells in the blood, and MAIT and  $T_{\text{mem}}$  cells in the liver. **d, e**, UMAP of matched blood and liver MAIT cells ( $n = 35,407$  cells from six donors) colored by the 11 identified clusters (**d**) or by tissue (**e**). **f**, Proportion of cells in each cluster from the blood and liver.

**g–i**, UMAPs colored by expression of *CD69* (**g**) and *CCL3* (**h**) genes or CD69 protein (**i**). **j**, Gene set enrichment analysis of liver compared with blood MAIT cells using published human and mouse tissue-resident memory T cell gene signatures. NES, normalized enrichment score. **k**, Over-representation analysis on the genes significantly upregulated in liver MAIT cells compared with blood MAIT cells. Top 15 gene ontology terms and associated Benjamini–Hochberg adjusted  $P$  values are shown. **l**, Heatmap showing activity (row-scaled AUCell scores) of the 20 most differentially active regulons (largest difference in average AUCell score) between matched blood and liver MAIT cells in Exp 1.  $n = 3$  donors.

only the TCR $\beta$  chain (TCR $\beta$  clonotypes; Fig. 2d). At the population level, MAIT cell TCR $\alpha$  chain pairing was promiscuous, with ~30% of TCR $\alpha$  chains paired with >1 unique TCR $\beta$ —generating multiple clones with identical TCR $\alpha$  chains (Fig. 2e). Conversely, most TCR $\beta$  chains paired

with a single TCR $\alpha$  (Extended Data Fig. 2i).  $T_{\text{mem}}$  cell TCR $\alpha\beta$  pairings were essentially unique (Fig. 2f and Extended Data Fig. 2j).

Overall, the MAIT cell TCR $\alpha$  repertoire is highly restricted, while the TCR $\beta$  repertoire is considerably more diverse. TCR repertoire



**Fig. 2 | MAIT cells have a restricted TCR $\alpha$  but diverse TCR $\beta$  chain, resulting in private TCR $\alpha\beta$  repertoires. a**, Percentage of *TRAJ33*, *TRAJ12* and *TRAJ20* TCR $\alpha$  chains paired with *TRBV6-1*, *TRBV6-4*, *TRBV20-1* or other TCR $\beta$  chains. Mean  $\pm$  s.e.m. is shown. **b**, Line plot colored by sample demonstrating the clonality of the TCR $\alpha\beta$  repertoire. **c**, Line plot comparing MAIT (blue) and  $T_{mem}$  (black) cell TCR $\alpha\beta$  clonality. **d**, Shannon diversity index for TCR $\alpha$  clonotypes and TCR $\beta$  clonotypes for each donor. **e, f**, TCR chain pairing at the population level. Number of unique TCR $\beta$  chains paired with any given TCR $\alpha$  chain in blood and liver MAIT cells (**e**) or MAIT and  $T_{mem}$  cells (**f**; blood and liver cells combined). **g**, Percentage of MAIT and  $T_{mem}$  cells belonging to a TCR $\alpha\beta$  clonotype shared

between matched blood and liver. **h**, Pearson's correlation between clonotype frequency in matched blood and liver. **i–k**, Number of cells from each donor belonging to TCR $\alpha\beta$  (**i**), TCR $\alpha$  (**j**) or TCR $\beta$  (**k**) functional clonotypes found in 1, 2, 3, 4, 5 or 6 (all) donors. Plots show TCR data for MAIT cells (or MAIT and  $T_{mem}$  cells in **c**, **f** and **g**) from all donors (**a–f**;  $n = 12$ ) or matched blood–liver donors (**g–k**;  $n = 6$ ). In **d–g**, boxes span the 25th–75th percentiles, the midline denotes the median and whiskers extend to  $\pm 1.5 \times IQR$ . Points in **e** and **f** indicate outliers. Two-sided Wilcoxon signed-rank test (**a** and **d**) and two-sided Wilcoxon rank-sum test (**e–g**) for all pairwise comparisons. Benjamini–Hochberg adjusted  $P$  values are shown (nonsignificant results omitted).

characteristics were similar in blood and liver, and we identified a preference for *TRAJ12/20-TRBV6-4* and *TRAJ33-TRBV20-1* pairings.

### TCR repertoires are donor-specific but shared across tissues

As TCR usage was similar across tissues, we hypothesized that blood and liver MAIT cells might show clonal overlap. On average, 72% of MAIT cells belonged to a TCR clonotype present in matched blood and liver (Fig. 2g) and clonotype frequency correlated between tissues (Fig. 2h). Clonal sharing was similarly high for  $T_{mem}$  cells (Fig. 2g). The extent of MAIT cell TCR repertoire overlap between tissues correlated with TCR capture rate, suggesting our data underestimates blood–liver sharing (Extended Data Fig. 2k).

We next examined repertoire overlap between donors, with clonotypes defined using CDR3 amino acid sequences (functional clonotypes). Despite their semi-invariant TCR, 98% of MAIT cells belonged

to a donor-specific clonotype, with no clonotypes shared between all six donors (Fig. 2i).  $T_{mem}$  cell functional clonotypes showed no overlap between donors (Extended Data Fig. 2l).

Given the restricted MAIT cell TCR $\alpha$  repertoire, we reasoned that functional clonotypes defined using the TCR $\alpha$  chain only (functional TCR $\alpha$  clonotypes) may show high overlap between donors. Supporting this hypothesis, the six donors shared 27 functional TCR $\alpha$  clonotypes comprising 79% (16,973/21,421) of MAIT cells (Fig. 2j). In contrast, 1.1% (196/18,564) and  $<0.1\%$  of  $T_{mem}$  cells belonged to a functional TCR $\alpha$  clonotype found in two and three donors, respectively (Extended Data Fig. 2m). Functional TCR $\beta$  clonotypes were largely donor-specific for MAIT and  $T_{mem}$  cells (Fig. 2k and Extended Data Fig. 2n).

Thus, distinct from  $T_{mem}$  cells, the MAIT cell TCR $\alpha$  repertoire is public. In contrast, the TCR $\beta$  chain is markedly more private and is what governs the uniqueness of individual MAIT cell TCR repertoires.



### Within-tissue transcriptional heterogeneity is limited

We next explored within-tissue heterogeneity.

Blood MAIT cells comprised nine clusters (Fig. 3a and Extended Data Fig. 3a). Transcriptional diversity between clusters was low, with few genes displaying cluster-specific expression (Fig. 3b and Supplementary Table 5a). Apart from the three clusters discussed below, cluster markers were not indicative of specific functions or known T cell differentiation states. Using mouse MAIT1 and MAIT17 (refs. 33,34) or human Th1 and Th17 (ref. 35) gene signatures, we were unable to identify human MAIT1 and MAIT17 subsets (Fig. 3c and Extended Data Fig. 3b). *CCL4* was upregulated in cluster 6 and interferon-stimulated genes in cluster 8 (Fig. 3b and Extended Data Fig. 3c), perhaps indicating some degree of basal cell activation. Cells in cluster 2 appeared primed for cytotoxicity with increased expression of granzysin and granzymes (Fig. 3d and Extended Data Fig. 3d). This cluster did not simply indicate cell activation, as *GZMB* and *GZMH* (lowly expressed in resting MAIT cells<sup>36</sup>) were only expressed by a small percentage of cells.

Liver MAIT cells comprised ten clusters (Fig. 3e and Extended Data Fig. 3e) with modest transcriptional differences (Extended Data Fig. 3f and Supplementary Table 5b). Clusters did not correspond to known differentiation states such as MAIT1 and MAIT17 (Extended Data Fig. 3g). As in blood, there was a *GNLV*-expressing cluster (cluster 6; Fig. 3f), while remaining clusters expressed different activation- or stress-induced molecules (Fig. 3g). Most genes showed a gradient of expression across clusters.

Despite reported phenotypic, functional and/or transcriptional differences<sup>4–6</sup>, CD8<sup>+</sup>, DN and CD4<sup>+</sup> MAIT cells did not comprise separate clusters in blood or liver (Fig. 3h,i and Extended Data Fig. 3h–k), and differentially expressed few genes and proteins (Supplementary Table 6a–h). Previous bulk RNA-seq data indicated higher TCR repertoire diversity in CD8<sup>+</sup> relative to DN MAIT cells<sup>5</sup>. After downsampling to equalize numbers of CD8<sup>+</sup> and DN cells per donor, we identified a small increase in the frequency of unique clonotypes among CD8<sup>+</sup> MAIT cells (Fig. 3j), but no difference in *TRAJ/TRBV* chain usage (Extended Data Fig. 3l,m), and an equivalent Shannon diversity index (Fig. 3k). Therefore, consistent with minor transcriptional differences, CD8<sup>+</sup> and DN MAIT cells show similar TCR usage.

Overall, contrasting with mice, human MAIT cells show limited transcriptional heterogeneity within tissues and do not comprise distinct MAIT1 and MAIT17 subsets, or subsets defined by coreceptor expression.

### TCR clonotypes show variable bias in cluster localization

We next examined whether the limited transcriptional heterogeneity within tissues correlated with clonal identity. Given the donor-specific private TCR repertoire, we separately clustered each sample, then used the exact multinomial test to determine whether clonotypes were nonrandomly distributed across clusters.

There was a range of associations between clonotype and cluster, both within and between donors. Some clonotypes predominantly localized in a single cluster (Fig. 4a and Extended Data Fig. 4a). Some showed a subtle but still significant bias in cluster localization (Fig. 4b and Extended Data Fig. 4a). Others were randomly distributed

(Fig. 4c and Extended Data Fig. 4a). Bias in cluster distribution was more frequently significant for larger clonotypes (Fig. 4d), suggesting nonsignificant results for some smaller clonotypes may reflect a lack of statistical power.

For a given clonotype, the extent of bias in cluster localization was not necessarily concordant in blood and liver (Fig. 4a,e and Extended Data Fig. 4a,b). This may be due in part to different clonotype sizes in the two tissues. Some clonotypes had a stable transcriptional phenotype across tissues. For example, a *TRAV1-2/TRAJ12* clonotype from donor 1.2 preferentially localized to the *GNLV*-expressing cluster in blood (Fig. 4f,g and Extended Data Fig. 4c) and liver (Fig. 4h,i and Extended Data Fig. 4d). Therefore, the transcriptional profile of resting MAIT cells is influenced by their clonal identity.

### MAIT cell functional diversity is stimulus-specific

Since we did not identify subsets of resting MAIT cells, we investigated whether functional subsets were present following activation. CD8<sup>+</sup> T cells were left unstimulated or stimulated with MR1/5-OP-RU (TCR) or IL-12 + IL-18 (cytokine). After 20 h, CD8<sup>+</sup> MAIT cells were sorted (CD26<sup>+</sup>CD161<sup>hi</sup>Vα7.2<sup>+</sup> for unstimulated and cytokine-stimulated, and CD26<sup>+</sup>CD161<sup>hi</sup> for TCR-stimulated due to TCR downregulation) for scRNA-seq and scTCR-seq (Exp 3; Supplementary Fig. 2), an approach we termed functional RNA-sequencing (fRNA-seq).

fRNA-seq revealed stimulus-specific transcriptional responses (Fig. 5a and Supplementary Table 7a,b). MAIT cells (27,305 cells) comprised nine clusters—these were present in all donors but largely stimulus-specific (Fig. 5b,c and Extended Data Fig. 5a). Consistent with their homogeneous resting transcriptome, unstimulated cells predominantly localized in one cluster. TCR-stimulated cells localized in clusters 1 and 4. Cells in cluster 1 were more activated than those in cluster 4, displaying increased expression of chemokines and cytokines including *CCL4*, *TNF* and *CSF2* (Fig. 5d,e, Extended Data Fig. 5b–d and Supplementary Table 8). Clusters 2, 3 and 5 largely comprised cytokine-stimulated cells and appeared to indicate different degrees of cell activation. Expression of activation markers (for example, *IL2RA*) and effector molecules (for example, *GZMB*) was low in cluster 3, but high in cluster 5 (Fig. 5d,e, Extended Data Fig. 5b,e and Supplementary Table 8). Cells in cluster 2 expressed high levels of *IFNG* but less *GZMB* than cells in cluster 5. Interferon-stimulated genes were uniquely expressed in cluster 7 (mostly cytokine-stimulated; Fig. 5d and Extended Data Fig. 5b).

As in Exp 1 and 2, we identified a *GNLV*-expressing cluster (cluster 6; Fig. 5d and Extended Data Fig. 5f) that contained cells from all three conditions (Fig. 5c). The cells did not express other cytotoxic molecules (for example, *GZMB*) or markers of activation, inconsistent with our initial hypothesis that *GNLV*-expressing cells at rest are primed for cytotoxicity.

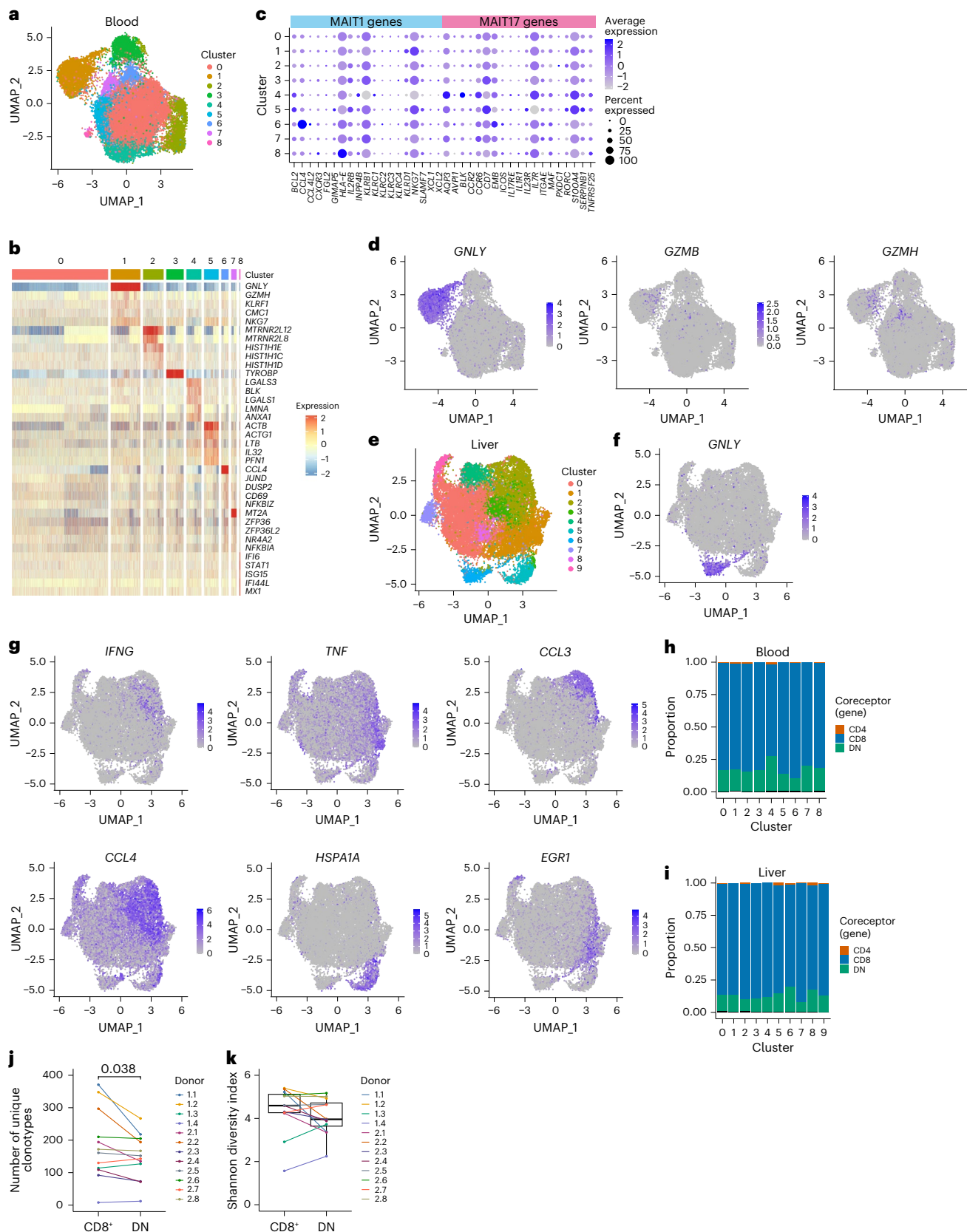
It was suggested that CD56 expression identifies a MAIT cell subset with enhanced cytokine responsiveness<sup>7</sup>. Following cytokine stimulation, CD56 (*NCAMI*)-expressing MAIT cells showed increased *IFNG* production relative to their nonexpressing counterparts (Extended Data Fig. 5g). However, as CD56 expression was qualitatively increased following cytokine stimulation (Extended Data Fig. 5h), we could not establish the usefulness of CD56 as a baseline indicator of functional

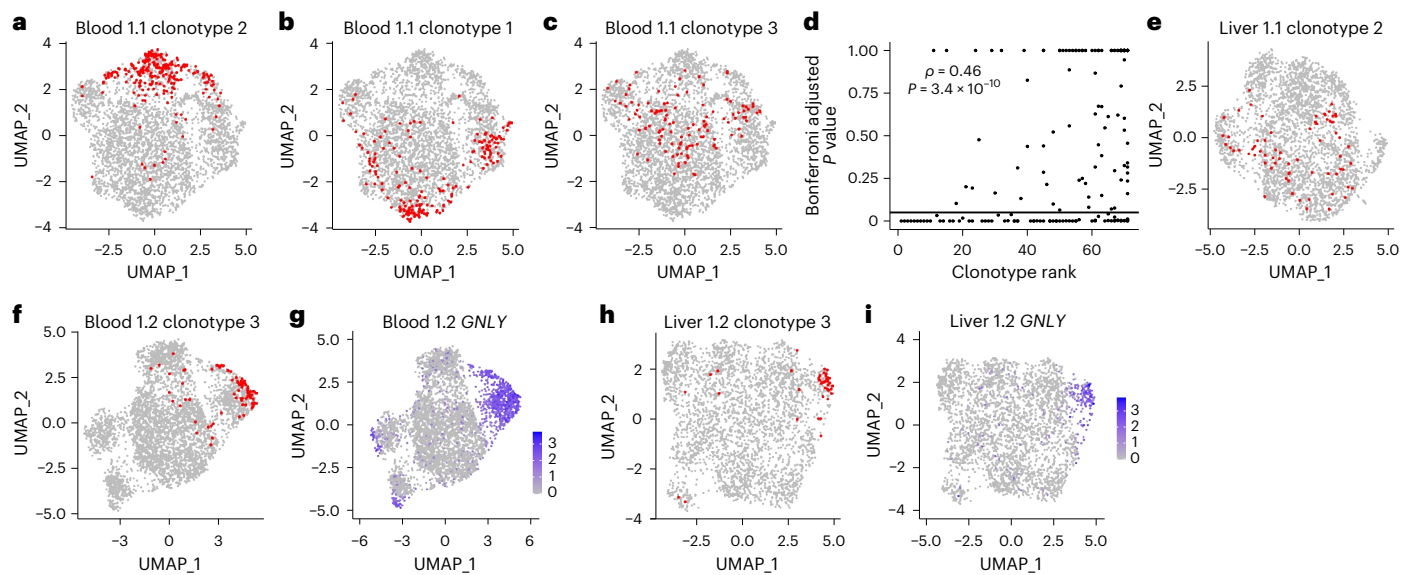
**Fig. 3 | MAIT cells within the blood and liver show minimal transcriptional heterogeneity.** **a**, UMAP of blood MAIT cells from matched blood–liver donors ( $n = 6$ ) colored by the nine identified clusters. **b**, Heatmap showing row-scaled log-transformed normalized expression of the top five or all (if <5) marker genes for each blood MAIT cell cluster. **c**, Expression of MAIT1 and MAIT17 genes in blood MAIT cell clusters. Dot color indicates the level of gene expression and dot size indicates the percentage of cells expressing the gene. **d**, UMAPs of blood MAIT cells colored by expression of *GNLV*, *GZMB* and *GZMH*. **e**, UMAP of liver MAIT cells from matched blood–liver donors ( $n = 6$ ) colored by the ten identified clusters. **f**, UMAP of liver MAIT cells colored by expression of *GNLV*. **g**, UMAPs

of liver MAIT cells colored by expression of *IFNG*, *TNF*, *CCL3*, *CCL4*, *HSPA1A* and *EGRI*. **h,i**, Proportion of CD4<sup>+</sup>, CD8<sup>+</sup> and DN cells in each blood (**h**) and liver (**i**) cluster. Coreceptor identity defined by the expression of *CD4*, *CD8A* and *CD8B* genes (Methods). **j**, Number of unique clonotypes in CD8<sup>+</sup> and DN MAIT cells from each donor ( $n = 12$ ; CD8<sup>+</sup> cell number within each donor downsampled to match the number of DN cells). **k**, Shannon diversity index for TCRαβ clonotypes in CD8<sup>+</sup> and DN MAIT cells from each donor ( $n = 12$ ). Boxes span the 25th–75th percentiles, the midline denotes the median and whiskers extend to  $\pm 1.5 \times$  IQR. Two-sided Wilcoxon signed-rank test in **j** and **k** (nonsignificant results omitted).

potential. To address this, we stimulated sorted CD56<sup>-</sup> and CD56<sup>+</sup> MAIT cells with IL-12 + IL-18 for 20 h, and measured IFN $\gamma$ , granzyme B, perforin and CD94 expression by flow cytometry. There was a trend toward increased expression of all tested markers in activated CD56<sup>+</sup> cells relative to CD56<sup>-</sup> cells (Extended Data Fig. 5i–l). Thus, CD56

expression correlates with MAIT cells primed for cytokine responsiveness. However, CD56<sup>+</sup> cells did not comprise a transcriptionally distinct cluster of MAIT cells in resting blood or liver (Extended Data Fig. 5m,n). Further experiments are necessary to understand the overall impact of CD56 expression on MAIT cell biology.





**Fig. 4 | TCR $\alpha\beta$  clonotypes are variably associated with transcriptional clusters.** **a–c**, UMAPs of blood MAIT cells from donor 1.1 showing in red cells from a single TCR $\alpha\beta$  clonotype. Plots show clonotype 2 (**a**), clonotype 1 (**b**) and clonotype 3 (**c**). Clonotype 1 is the largest clonotype from a donor, clonotype 2 is the second largest and so on. **d**, Spearman's rank correlation between clonotype size (rank) and the Bonferroni adjusted *P* value for association between clonotype and cluster (exact multinomial test, performed for clonotypes from

all donors ( $n = 12$ ) with  $\geq 20$  cells). **e**, UMAP of liver MAIT cells from donor 1.1 with cells from clonotype 2 indicated in red (the same clonotype as in **a**). **f, g**, UMAPs of blood MAIT cells from donor 1.2 with cells from clonotype 3 indicated in red (**f**) or expression of *GNLY* shown in blue (**g**). **h, i**, UMAPs of liver MAIT cells from donor 1.2 with cells from clonotype 3 indicated in red (**h**) or expression of *GNLY* shown in blue (**i**).

### Pseudotime analysis reveals linear activation trajectories

As MAIT cell clusters captured cells at different stages of activation, we further explored transcriptional responses to stimulation using pseudotime analysis. The Slingshot<sup>37</sup> algorithm identified a branching trajectory with a single branch point close to unstimulated cells (Fig. 5f), suggesting MAIT cells become transcriptionally distinct early following TCR and cytokine stimulation. Results were validated using SCORPIUS<sup>38</sup> (Extended Data Fig. 6a,b).

Through random forest regression, we identified the genes most important for predicting cell pseudotime on the TCR and cytokine trajectories (Supplementary Table 9). Gene importance was highly correlated between the two trajectories (Fig. 5g), with nine of the top 20 genes overlapping, including *IL2RA* (CD25) and *TNFRSF18* (GITR; both upregulated) and *IL7R* (downregulated). However, several notable genes were important primarily for one trajectory (Fig. 5g–i). *IFNG* and *IL26* were specific to the cytokine trajectory, while *CCL3*, *CCL4* and *TNFRSF9* (4-1BB) showed greater importance for the TCR trajectory. Protein expression, as measured by flow cytometry, was consistent with gene expression (Extended Data Fig. 6c–k)—IFN $\gamma$  and CD40L were more strongly induced by cytokine stimulation and CCL4, TNF, 4-1BB and CD25 by TCR stimulation. Granzyme B was similarly induced by both stimuli.

### Regulation of TCR- and cytokine-induced transcription

Using SCENIC<sup>29</sup>, we identified 159 high-confidence regulons regulating shared and stimulus-specific gene expression. Although global changes in transcription factor activity relative to unstimulated cells were highly correlated for TCR- and cytokine-stimulated cells (Fig. 6a), some regulons showed markedly different activity between conditions.

Relative to unstimulated cells, 65 regulons had increased activity: 11 were TCR-specific, 22 were cytokine-specific and 32 were shared (Fig. 6b). TCR-specific regulons included TCR-induced transcription factors (EGR1, EGR2, NR4A1), and VDR, CREM and STAT5A, which have varied roles in regulating Th17 differentiation and IL-17 production<sup>39–41</sup> (Fig. 6b and Extended Data Fig. 7a). Cytokine-specific regulons

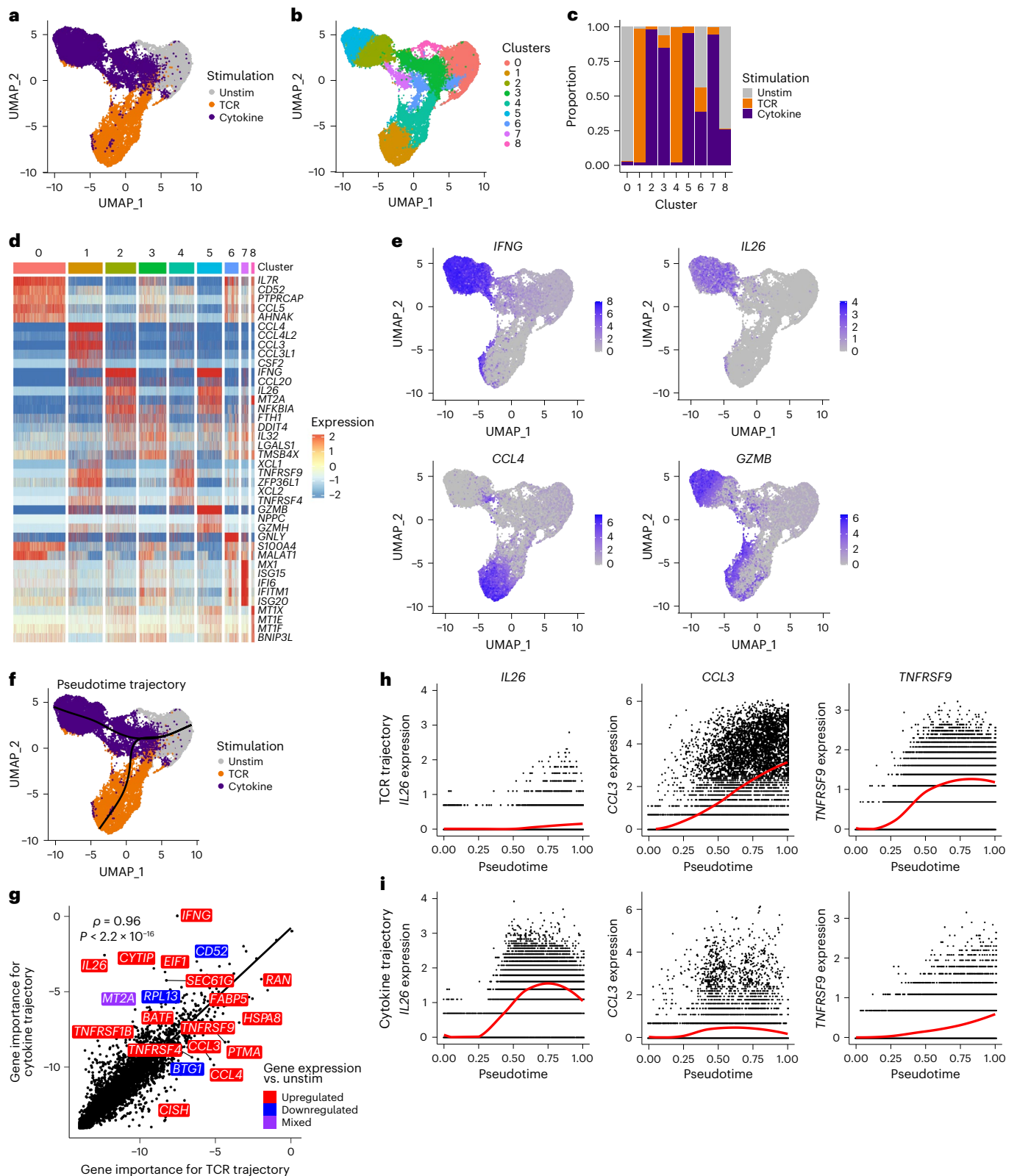
included STAT1, interferon regulatory factors, XBP1 and IKZF1 (Ikaros; Fig. 6b and Extended Data Fig. 7b). Ikaros regulates activated conventional CD8<sup>+</sup> T cell responsiveness to IL-12 (ref. 42).

The 32 regulons with increased activity upon TCR and cytokine stimulation were also all differentially active between the two conditions. TBX21 activity was most similar between TCR- and cytokine-stimulated cells—its target genes were enriched for interleukin, Toll-like receptor and NF- $\kappa$ B signaling pathways (Extended Data Fig. 7c). As expected, NFATC1 and STAT4 showed increased activity in TCR- and cytokine-stimulated cells, respectively. We identified new candidate regulators of stimulus-specific MAIT cell functions, namely HIVEP3 for TCR-stimulated cells, and BATF, BCL6 and HIF1A for cytokine-stimulated cells (Fig. 6c–f). HIVEP3 is essential for the development of innate-like T cells including MAIT cells<sup>43</sup>. Among its many roles, BATF promotes effector CD8<sup>+</sup> T cell differentiation through the upregulation of key transcription factors, cytokine receptors and signaling molecules<sup>44</sup>.

As with upregulated genes, regulon activity varied across stimulated cells. Most regulons progressively increased in activity over pseudotime (Fig. 6g,h). However, activity of some regulons peaked early and subsequently declined, for example, EGR1 and NFKB1 on the TCR trajectory, and STAT1 and IRF1 on the cytokine trajectory. Several early-activated regulons regulated later-activated transcription factors. STAT1 was a predicted regulator of *TBX21*. *HIVEP3* was a predicted target of EGR1, supporting its TCR-specific activation and function.

To identify candidate regulators of MAIT cell function, we examined effector gene localization within SCENIC regulons. Unsupervised hierarchical clustering identified two clusters of effector genes (Fig. 6i). Cluster 1 comprised genes preferentially induced by TCR signaling, namely *CSF2*, *TNF*, *CCL3* and *CCL4*, suggesting similar regulation. Genes in cluster 1 were regulated by EGR1, NFATC1 and MYC. In addition, HIVEP3 was a predicted regulator of *CSF2*, *CCL3* and *CCL4*. Cluster 2 comprised both cytokine-specific genes (for example, *IL26*) and genes induced by both stimuli (for example, *GZMB*). As expected, *IFNG* was





**Fig. 5 | TCR- and cytokine-activated MAIT cells follow distinct linear trajectories.** **a, b**, UMAPs of MAIT cells from all donors colored by stimulation condition (**a**) or the nine identified clusters (**b**).  $n = 27,305$  cells from three donors. **c**, Proportion of cells in each cluster from the three stimulation conditions. **d**, Heatmap showing row-scaled log-transformed normalized expression of the top five marker genes for each cluster. **e**, UMAPs colored by expression of *IFNG*, *IL26*, *CCL4* and *GZMB*. **f**, UMAP of MAIT cells from all donors with the branching pseudotime trajectory identified using Slingshot shown in

black. **g**, Spearman's rank correlation between gene importance ( $\log_2 1/\text{gene importance rank}$ ) on SCORPIUS TCR and cytokine trajectories. Labels indicate the most differentially important genes, ten with higher importance on the TCR trajectory and ten with higher importance on the cytokine trajectory. Colors indicate whether gene expression was upregulated (red), downregulated (blue) or mixed (purple; upregulated in TCR and downregulated in cytokine or vice versa) relative to unstimulated cells. **h, i**, Expression of *IL26*, *CCL3* and *TNFRSF9* along SCORPIUS TCR (**h**) and cytokine (**i**) trajectories.



regulated by STAT4, but was also in the BATF regulon, suggesting BATF could contribute to enhanced *IFNG* production in cytokine-stimulated MAIT cells compared with TCR-stimulated MAIT cells.

Overall, TCR- and cytokine-stimulated MAIT cells exhibit shared and stimulus-specific regulation. Our data reveal new candidate regulators of TCR- and cytokine-specific responses and their predicted target genes.

### Clonal identity influences MAIT cell activation potential

TCR clonotypes showed varied associations with resting transcriptional clusters (Fig. 4) and published data suggest functional differences linked to TCR $\beta$  usage<sup>7,23,25</sup>. Using fRNA-seq, we investigated whether MAIT cell activation potential (pseudotime position) correlated with TCR $\beta$  usage or clonal identity. Within donors, activation capacity was significantly associated with *TRBV* usage, but there was considerable variability among cells with the same *TRBV* gene (Fig. 7a,b and Extended Data Fig. 8a,b). *TRBV* pseudotimes did not correlate between donors (Fig. 7a,b and Extended Data Fig. 8c,d) or between TCR and cytokine trajectories (Fig. 7c), indicating no intrinsic difference in the activation potential of *TRBV* genes.

Therefore, we hypothesized that observed variation could reflect clonal differences. Consistent with this, activation capacity differed between clonotypes on the TCR and cytokine trajectory (Fig. 7d,e and Extended Data Fig. 8e,f). High variability within clonotypes (Fig. 7d,e) suggested additional major influences on MAIT cell activation capacity. Clonotype pseudotimes were not correlated in response to TCR and cytokine stimulation (Fig. 7f), and there was no consistent association between clonotype size and responsiveness to stimulation (Fig. 7g–i). Therefore, larger clones are not intrinsically more functional than smaller clones. Given that variation between clonotypes was observed on the cytokine trajectory as well as the TCR trajectory, differences in clonotype functionality may not solely be associated with the strength of TCR-ligand binding.

### IL-17<sup>-</sup> and IL-17<sup>+</sup> cells overlap in function and TCR usage

MAIT1 and MAIT17 subsets were not detected in blood or liver. However, due to minimal IL-17 production, our stimulation experiment (Exp 3) did not allow us to address the source and functionality of IL-17-producing human MAIT cells. Therefore, our second stimulation experiment (Exp 4) included a dual TCR+cytokine condition that induces enhanced IL-17 production<sup>9</sup>. Isolated T cells were left unstimulated or stimulated for 20 h with MRI/5-OP-RU (TCR), IL-12 + IL-18 (cytokine) or both, before MAIT cell sorting (Supplementary Fig. 3). For the TCR+cytokine condition, we performed an additional 3 d (68 h) stimulation, previously shown to increase IL-17 production<sup>9</sup>.

We analyzed 96,867 and 42,765 MAIT cells from three donors at 20 h and 68 h, respectively. After 20 h, unstimulated and stimulated cells were phenotypically and transcriptionally distinct (Supplementary Table 10a–f), with TCR+cytokine-stimulated cells localizing between TCR and cytokine single-stimulated cells on the UMAP (Fig. 8a and Extended Data Fig. 9a). Likewise, TCR+cytokine-stimulated cells were distinct from unstimulated cells at 68 h (Fig. 8b, Extended Data Fig. 9b and Supplementary Table 10g,h).

At 20 h, *IFNG* and *IL26* were primarily upregulated following cytokine stimulation; *CCL3*, *CCL4* and *TNFRSF9* (and the corresponding 4-1BB protein) were specific to the TCR condition; and *GZMB* was similarly upregulated by both stimuli, consistent with Exp 3 (Extended Data Fig. 9c–e and Supplementary Table 10a,c). TCR+cytokine stimulation upregulated both TCR- and cytokine-specific genes (Supplementary Table 10e,g). In general, gene and protein expression were highly correlated in the single and dual stimulation conditions (Fig. 8c,d and Extended Data Fig. 9f,g). Notably, the expression of *IL17F* was significantly increased following dual relative to single stimulation (Fig. 8c,d). IL-17-expressing cells comprised a distinct cluster at both timepoints (Fig. 8e,f and Extended Data Fig. 9h–k). Three-day stimulation induced a higher fraction of IL-17<sup>+</sup> MAIT cells compared with 1 d stimulation (Fig. 8g). *IL17F* was expressed by all cells in the cluster, while a small percentage produced *IL17A* (Extended Data Fig. 9i,k).

To investigate whether IL-17-expressing MAIT cells comprise a distinct subset, we examined transcriptional differences between IL-17<sup>-</sup> and IL-17<sup>+</sup> MAIT cells following dual stimulation. At 20 h, IL-17<sup>-</sup> and IL-17<sup>+</sup> cells differentially expressed 23 genes (Supplementary Table 11a). Along with *IL17A* and *IL17F*, IL-17-expressing cells showed increased expression of *CCR6* and *CCL20* but reduced *GZMB* and *KLRD1*. Nevertheless, IL-17<sup>+</sup> cells expressed high levels of *GZMB* and *KLRD1* compared with unstimulated cells (Extended Data Fig. 10a). Expression of *IFNG*, *IL26*, *CCL3* and other effector molecules was comparable in IL-17<sup>-</sup> and IL-17<sup>+</sup> cells (Extended Data Fig. 9c). At 68 h, IL-17<sup>-</sup> and IL-17<sup>+</sup> cells differentially expressed 28 genes (Supplementary Table 11c), but again differences in effector gene expression were small and only five differentially expressed genes overlapped at the two timepoints. Protein analysis revealed similar findings—IL-17<sup>-</sup> and IL-17<sup>+</sup> cells differentially expressed three and six proteins at 20 h and 68 h, respectively (Supplementary Table 11b,d).

The similar transcriptional profiles of IL-17<sup>-</sup> and IL-17<sup>+</sup> cells, and the increased frequency of IL-17-expressing cells at 68 h relative to 20 h, suggested that IL-17<sup>+</sup> cells may represent a functional state obtainable by all MAIT cells under appropriate stimulation conditions. To investigate this, we compared the TCR repertoire of IL-17<sup>-</sup> and IL-17<sup>+</sup> cells. Clonotype abundance strongly correlated between the two groups at both timepoints (Fig. 8h,i). The number of unique clonotypes among IL-17-expressing cells was increased at 68 h compared with 20 h (Fig. 8j), indicating that new cells become IL-17<sup>+</sup>. However, there was no difference in the number of unique clonotypes among IL-17<sup>-</sup> and IL-17<sup>+</sup> cells (Fig. 8k) or in the Shannon diversity index (Extended Data Fig. 10b) (data downsampled to ensure equivalent numbers of IL-17<sup>-</sup> and IL-17<sup>+</sup> cells within each donor at a given timepoint). Moreover, *TRAJ* and *TRBV* usage was comparable (Extended Data Fig. 10c,d).

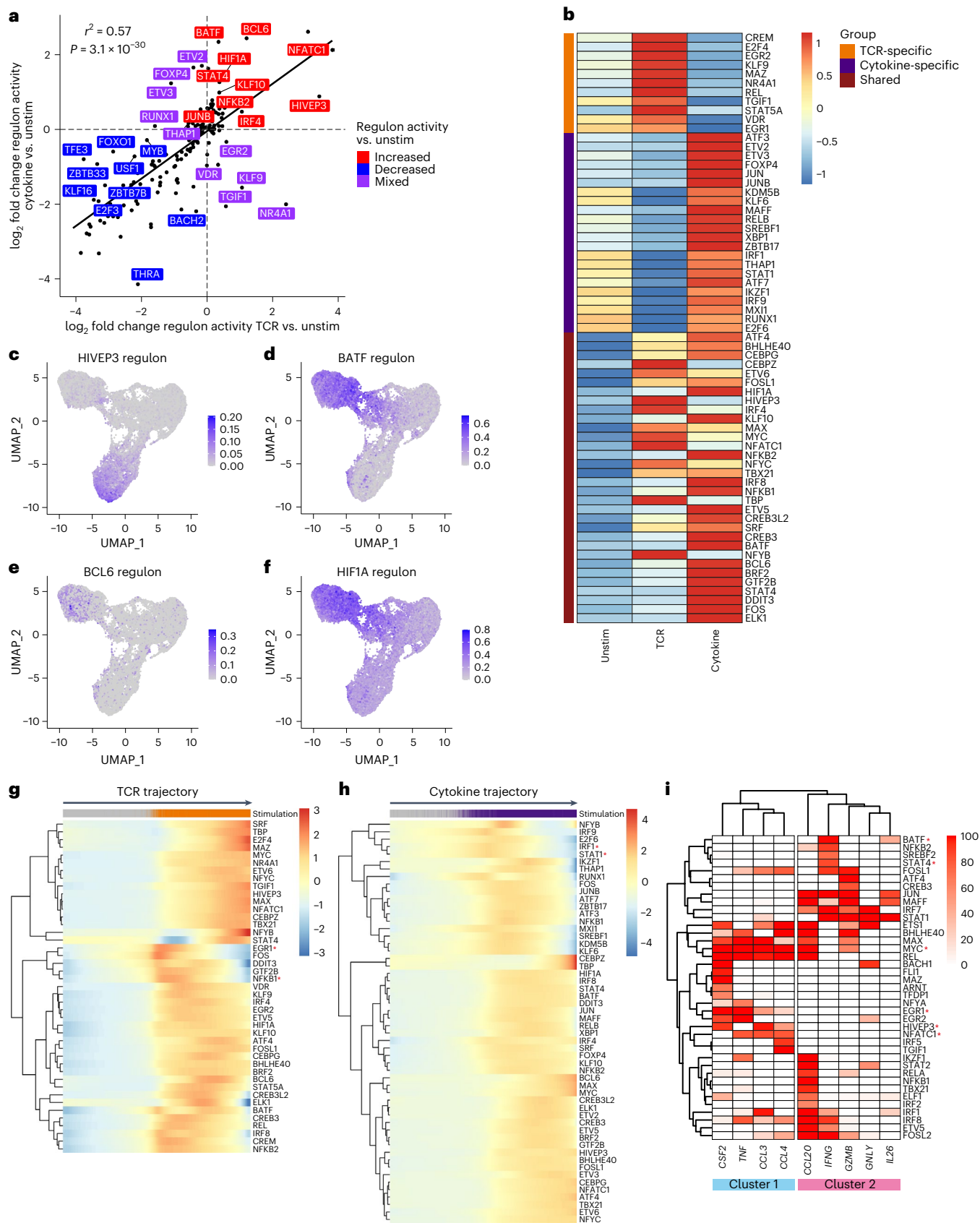
To examine the regulation of IL-17 gene expression in MAIT cells, we generated a bulk ATAC-seq dataset ( $n = 3$  donors) comprising three blood CD8<sup>+</sup> T cell subsets: naïve T cells, MAIT cells and T<sub>mem</sub> cells. As expected given their expression by resting MAIT cells<sup>1</sup>, the type 17-associated genes *RORC*, *IL23R* and *CCR6* showed increased accessibility in MAIT cells compared with T<sub>mem</sub> and naïve T cells (Extended Data Fig. 10e–g). In addition, MAIT cells showed increased accessibility of peaks associated with *IL17A* and *IL17F* (Fig. 8l,m).

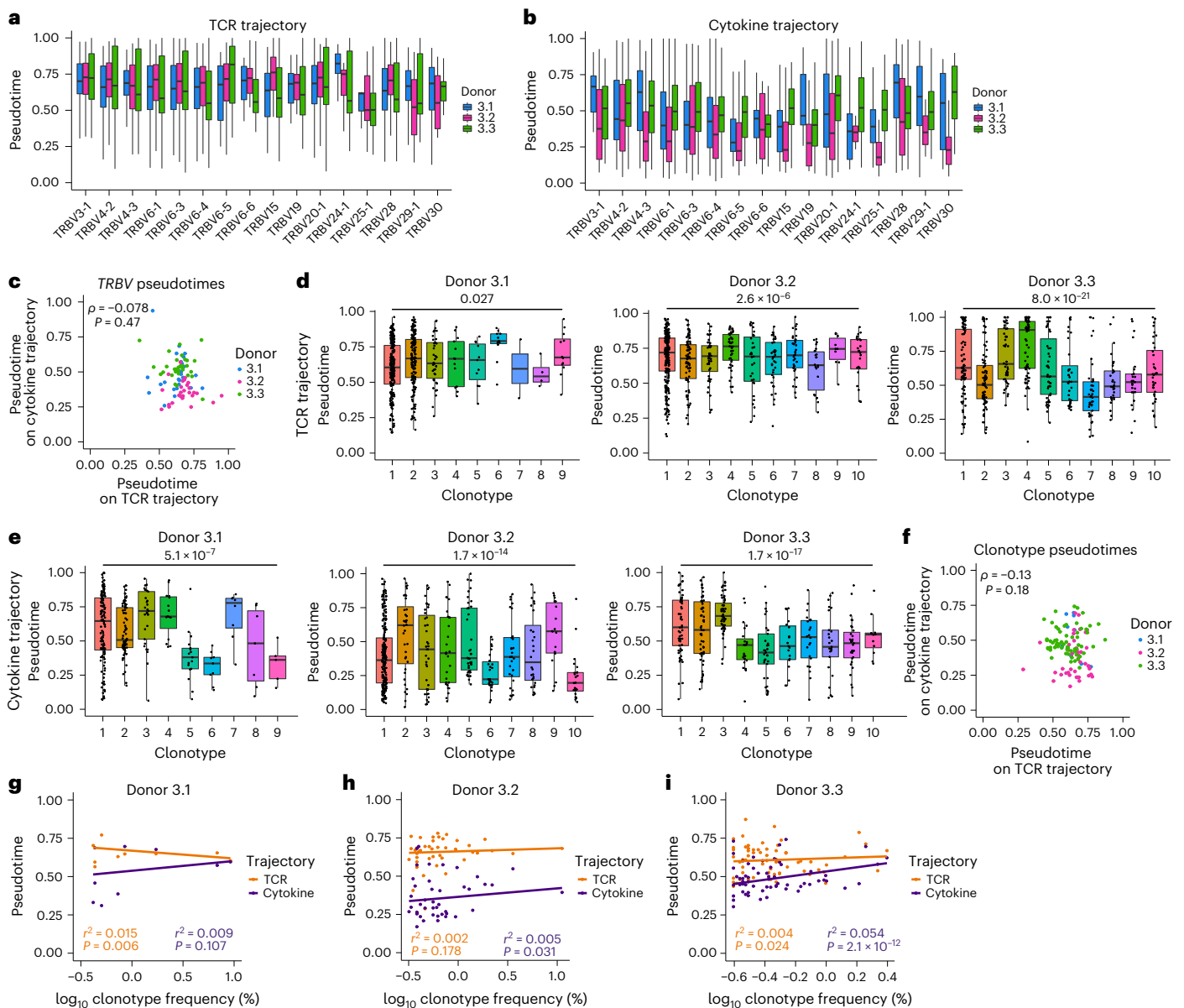
**Fig. 6 | Transcriptional regulation of TCR- and cytokine-stimulated MAIT cells exhibits shared and distinct properties.** **a**, Pearson's correlation between the log<sub>2</sub> fold change in regulon activity (AUCell scores) between TCR-stimulated and unstimulated MAIT cells, and cytokine-stimulated and unstimulated MAIT cells. Labels show the regulons with the largest difference in log<sub>2</sub> fold change relative to unstimulated cells between the TCR and cytokine trajectory, ten with increased (red), ten with decreased (blue) and ten with mixed (purple; increased in TCR and decreased in cytokine or vice versa) activity following stimulation. **b**, Heatmap showing the activity (row-scaled average AUCell scores) of TCR-specific (orange), cytokine-specific (purple) and shared (maroon) upregulated regulons in each stimulation condition. **c–f**, UMAPs colored by the

activity of HIVEP3 (**c**), BATF (**d**), BCL6 (**e**) and HIF1A (**f**) regulons. **g**, Heatmap showing regulon activity (smoothed AUCell scores) over pseudotime on the SCORPIUS TCR trajectory for regulons upregulated upon TCR stimulation. Gray, unstimulated cells; orange, TCR-stimulated cells. **h**, Heatmap showing regulon activity (smoothed AUCell scores) over pseudotime on the SCORPIUS cytokine trajectory for regulons upregulated upon cytokine stimulation. Gray, unstimulated cells; purple, cytokine-stimulated cells. **i**, Regulation of select MAIT cell effector genes. Heatmap is colored by the percent occurrence of each gene within each transcription factor regulon. High-confidence regulons predicted to regulate at least one of the genes in >50% of pySCENIC runs are included. Red asterisks in **g–i** indicate regulons mentioned in the text.

The two differentially accessible *IL17A* peaks were located at the promoter and in the upstream intergenic region, while the two *IL17F* peaks were in the downstream intergenic region. Compared with the *GZMB* promoter peak (Fig. 8n), the *IL17A* promoter peak was of

lower magnitude, while *IL17F* lacked a peak at the promoter. Reduced accessibility of *IL17A* and *IL17F* promoters may explain delayed IL-17 secretion relative to rapid granzyme B upregulation following MAIT cell activation.





**Fig. 7 | Clonal identity influences MAIT cell activation potential.** **a, b**, Box plots split by donor ( $n = 3$ ) showing pseudotime values on SCORPIUS TCR (**a**) and cytokine (**b**) trajectories for MAIT cells expressing different *TRBV* gene segments. Kruskal–Wallis test  $P$  values for **a** are  $8.5 \times 10^{-21}$ ,  $2.2 \times 10^{-9}$  and  $1.3 \times 10^{-15}$ , and for **b** are 0.045, 0.045 and  $2.2 \times 10^{-20}$  for donors 3.1, 3.2 and 3.3, respectively. **c**, Spearman's rank correlation between average *TRBV* pseudotimes on SCORPIUS TCR and cytokine trajectories. **d, e**, Pseudotime values for cells from the largest ten clonotypes in each donor or all clonotypes containing  $\geq 20$  cells ( $n = 9, 10$  and 10 clonotypes for donors 3.1, 3.2 and 3.3, respectively) on SCORPIUS TCR (**d**) and

cytokine (**e**) trajectories. Kruskal–Wallis test was performed for all clonotypes containing  $\geq 20$  cells. **f**, Spearman's rank correlation between average clonotype pseudotimes on SCORPIUS TCR and cytokine trajectories. **g–i**, Pearson's correlation between  $\log_{10}$  clonotype frequency and pseudotime on SCORPIUS TCR and cytokine trajectories for donors 3.1 (**g**), 3.2 (**h**) and 3.3 (**i**). Plots show stimulated cells only (**a–i**), *TRBV* gene segments with a frequency of  $>1\%$  in any donor (**a–c**) and clonotypes containing  $\geq 20$  cells (**d–i**). In **a, b, d** and **e**, boxes span the 25th–75th percentiles, the midline denotes the median and whiskers extend to  $\pm 1.5 \times \text{IQR}$ .

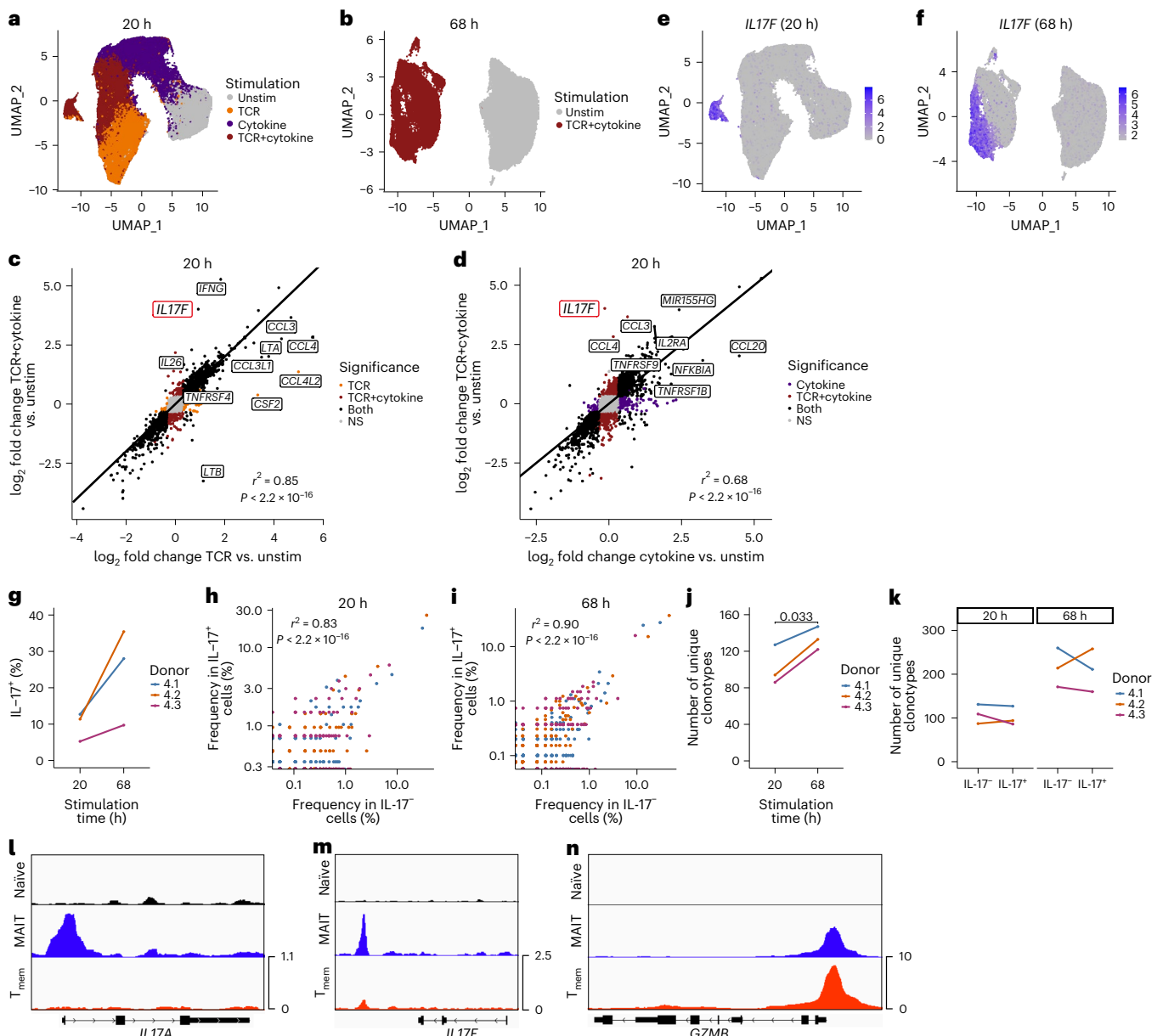
Our data suggest that increased numbers of MAIT cells acquire the capacity to produce IL-17 over time, perhaps due to a requirement for chromatin remodeling, and that aside from IL-17 production, IL-17<sup>+</sup> and IL-17<sup>+</sup> MAIT cells show similar transcriptional and functional profiles. This is in stark contrast to resting mouse MAIT cells that comprise distinct MAIT1 and MAIT17 subsets<sup>10–12,33,34</sup>.

## Discussion

Our single-cell data from blood and liver, and TCR- and/or cytokine-stimulated MAIT cells, suggest that human MAIT cells comprise a single, highly adaptable, cell population. Transcriptional plasticity is governed

by tissue localization, clonal identity and activation state (influenced by type and duration of stimulation). Despite their semi-invariant TCR and shared antigen specificity, diverse TCR $\beta$  usage results in private MAIT cell TCR repertoires. This may have important functional consequences, as the clonal identity of an individual MAIT cell influenced its resting and activated transcriptional profile.

Liver MAIT cells were transcriptionally distinct from blood, expressing genes and proteins associated with activation and tissue residency. Basal activation could reflect responses to microbial ligands transported from the gut to the liver via the hepatic portal vein<sup>45</sup>. Liver residency is consistent with bulk RNA-seq



**Fig. 8 | IL-17<sup>-</sup> and IL-17<sup>+</sup> MAIT cells are functionally and clonally related.**

**a**, UMAP of 20 h-stimulated MAIT cells colored by stimulation condition.  $n = 96,867$  cells from three donors. **b**, UMAP of 68 h-stimulated MAIT cells colored by stimulation condition.  $n = 42,765$  cells from three donors. **c**, Pearson's correlation between the  $\log_2$  fold change in gene expression between TCR-stimulated and unstimulated MAIT cells, and TCR+cytokine-stimulated and unstimulated MAIT cells (20 h stimulation). Labels highlight selected genes that were differentially regulated by the two stimuli. Point colors indicate whether the gene was significantly differentially expressed in response to TCR stimulation only (orange), TCR+cytokine stimulation only (maroon), both (black) or neither (NS; gray). **d**, Pearson's correlation between the  $\log_2$  fold change in gene expression between cytokine-stimulated and unstimulated MAIT cells, and TCR+cytokine-stimulated and unstimulated MAIT cells (20 h stimulation). Labels highlight selected genes that were differentially regulated by the two stimuli. Point colors indicate whether the gene was significantly differentially expressed in response to cytokine stimulation only (purple), TCR+cytokine stimulation only

(maroon), both (black) or neither (NS; gray). **e, f**, UMAPs of 20 h-stimulated (**e**) and 68 h-stimulated (**f**) MAIT cells colored by expression of *IL17F*. **g**, Percentage of cells within the IL-17-expressing cluster following 20 h or 68 h stimulation. **h, i**, Pearson's correlation between clonotype frequency in IL-17<sup>-</sup> (cells within all other clusters) and IL-17<sup>+</sup> (cells within the IL-17<sup>+</sup> cluster) TCR+cytokine-stimulated MAIT cells at 20 h (**h**) and 68 h (**i**). **j**, Number of unique clonotypes detected within the IL-17<sup>-</sup> cluster following 20 h or 68 h stimulation. Cell numbers for each donor were downsampled to ensure equal numbers of TCR+cytokine-stimulated cells at the two timepoints. **k**, Number of unique clonotypes within IL-17<sup>-</sup> (cells within all other clusters) and IL-17<sup>+</sup> (cells within the IL-17<sup>+</sup> cluster) TCR+cytokine-stimulated MAIT cells following 20 h or 68 h stimulation. Cell numbers for each donor were downsampled to ensure equal numbers of IL-17<sup>-</sup> and IL-17<sup>+</sup> cells at a given timepoint. **l–n**, Representative ATAC-seq tracks showing *IL17A* (**l**), *IL17F* (**m**) and *GZMB* (**n**) gene loci in naive T (black), MAIT (blue) and T<sub>mem</sub> (red) cells.  $n = 3$  donors in **a–k**. Two-sided paired *t*-test was performed in **g, j** and **k** (nonsignificant results omitted).

analysis of human liver MAIT cells and with mouse parabiosis experiments<sup>12</sup>. However, MAIT cell frequency and TCR usage are similar in human thoracic duct and matched blood<sup>19</sup>, and TCR repertoires overlapped in blood and liver. In human intestinal and uterine

transplantation, tissue MAIT cells are largely recipient-derived at >1-year posttransplantation<sup>46,47</sup>. Therefore, the extent of human MAIT cell tissue residency requires further examination. A small fraction of liver MAIT cells may be circulating cells in transit through



the liver. However, this is unlikely to have meaningfully impacted our conclusions.

Mouse MAIT cells comprise developmentally, transcriptionally and functionally distinct MAIT1 and MAIT17 subsets<sup>10–12,33,34</sup>. In contrast, human MAIT cells displayed low baseline transcriptional heterogeneity in blood and liver, and clusters were not indicative of MAIT1 and MAIT17 cells or other T cell polarization states. Resting MAIT cell clusters did not clearly associate with the clusters identified following activation. Thus, mRNA-seq adds an important dimension to the analysis of T cell biology. TCR+cytokine stimulation induced IL-17 in a fraction of MAIT cells, but IL-17<sup>+</sup> and IL-17<sup>-</sup> cells similarly expressed other effector molecules and had overlapping TCR repertoires. Therefore, we hypothesize that all human MAIT cells have the capacity to produce IL-17 under appropriate conditions.

TCR and cytokine stimulation induced distinct responses, underpinned by altered regulatory networks. Activated MAIT cells did not comprise discrete functional lineages but were distributed along stimulus-specific activation trajectories. HIVEP3, a TCR-specific transcription factor (regulating *CCL3* and *CCL4*), and BATF, a cytokine-specific transcription factor (regulating *IFNG* and *IL26*), were new predicted regulators of MAIT cell function. A limitation of this analysis is that high-confidence regulons were not identified for ROR $\gamma$ t and PLZF, perhaps due to relatively poor gene detection.

Basic TCR repertoire characteristics were consistent with prior studies<sup>21,30</sup>. Surprisingly, the extent of MAIT cell clonality was comparable with T<sub>mem</sub> cells and individuals displayed largely private TCR $\alpha\beta$  repertoires. This challenges the paradigm of MAIT cells as a clonally-restricted population with large numbers of public TCRs and validates previous studies with small cell numbers or bulk TCR repertoire data<sup>20,48</sup>. However, the TCR $\alpha$  chain, key for ligand recognition<sup>31,32</sup>, was highly shared between individuals. TCR $\alpha\beta$  clonotypes overlapped considerably in blood and liver. While consistent with shared TCR $\beta$  usage in matched blood and lymph<sup>19</sup>, differential *TRAJ/TRBV* usage was identified in studies without matched blood and tissue (breast<sup>22</sup>, kidney and intestine<sup>21</sup>). Identification of private TCR repertoires highlights the importance of matched samples for accurately comparing TCR usage across tissues.

We identified an association between the clonal identity and transcriptome of individual MAIT cells. We hypothesize that the clonotype-cluster association in resting blood and liver reflects differences in the basal activation of clones. Differential activation capacity dependent on TCR clonotype is consistent with altered clonal distribution following *Salmonella* infection<sup>23</sup> and increased clonality with age<sup>20</sup>. However, activation capacity was not correlated with clonotype size. This appears to contrast with the superior proliferation of MAIT cells expressing the most abundant V $\beta$  segments upon in vitro *Escherichia coli* stimulation<sup>7</sup>. Discordant results could reflect differences in experimental approach or the absence of a direct correlation between activation kinetics and proliferative potential. Further study is necessary to understand the driving factors and functional consequences of clonal differences in activation capacity.

A recent paper<sup>49</sup> and a preprint<sup>50</sup> present human MAIT cell scRNA-seq data that are relevant to our findings. However, our study is unique in several key regards, namely the inclusion of TCR data—allowing clonotype to be linked to function—and detailed characterization of responses to multiple stimuli and at multiple timepoints. This is critical for understanding the diversity of MAIT cell functions, including IL-17 production. Consistent with our study, Chandra et al.<sup>50</sup> reported a tissue residency signature in lung MAIT cells and failed to identify a MAIT17 subset. Vorkas et al.<sup>49</sup> analyzed blood MAIT cells following 15 h and 7 d TCR stimulation (direct ex vivo analysis or following cytokine stimulation was not performed). Based on the identification of 12 clusters and gene expression differences between CD4- and CD8-expressing cells (although these did not form separate clusters), the authors concluded that human MAIT cells comprise

multiple subsets. However, our trajectory and clonality analysis (based on TCR-confirmed MAIT cells) suggest a continuum of response to stimulation by a single population.

In conclusion, we present a genome-wide single-cell characterization of the transcriptome and TCR repertoire of blood and liver, and resting and activated, human MAIT cells. Our data indicate largely private TCR repertoires, highly shared between matched blood and liver. MAIT cells showed stimulus-specific transcriptional responses, and we identified candidate regulators of the TCR- and cytokine-specific response. While human MAIT cells produce IL-17 following TCR+cytokine stimulation, IL-17<sup>-</sup> cells have a similar TCR repertoire and effector profile to IL-17<sup>+</sup> cells, suggesting they do not comprise a bona fide MAIT17 subset. CD4/CD8 coreceptor expression was not associated with distinct transcriptional states. At rest and following activation, MAIT cell clones show subtle differences in transcriptional profile and functional capacity, which may have important biological consequences, particularly in the context of suboptimal stimulation. Our data provide new insights into human MAIT cell biology, relevant to related innate-like subsets, and a comprehensive resource for further MAIT cell studies in health and disease.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41590-023-01575-1>.

## References

- Garner, L. C., Klenerman, P. & Provine, N. M. Insights into mucosal-associated invariant T cell biology from studies of invariant natural killer T cells. *Front. Immunol.* **9**, 1478 (2018).
- Ussher, J. E. et al. CD161<sup>+</sup>CD8<sup>+</sup> T cells, including the MAIT cell subset, are specifically activated by IL-12+IL-18 in a TCR-independent manner. *Eur. J. Immunol.* **44**, 195–203 (2014).
- Godfrey, D. I., Koay, H. F., McCluskey, J. & Gherardin, N. A. The biology and functional importance of MAIT cells. *Nat. Immunol.* **20**, 1110–1128 (2019).
- Kurioka, A. et al. Shared and distinct phenotypes and functions of human CD161<sup>+</sup> V $\alpha$ 7.2<sup>+</sup> T cell subsets. *Front. Immunol.* **8**, 1031 (2017).
- Dias, J. et al. The CD4<sup>+</sup>CD8<sup>-</sup> MAIT cell subpopulation is a functionally distinct subset developmentally related to the main CD8<sup>+</sup> MAIT cell pool. *Proc. Natl Acad. Sci. USA* **115**, E11513–E11522 (2018).
- Gherardin, N. A. et al. Human blood MAIT cell subsets defined using MR1 tetramers. *Immunol. Cell Biol.* **96**, 507–525 (2018).
- Dias, J., Leeansyah, E. & Sandberg, J. K. Multiple layers of heterogeneity and subset diversity in human MAIT cell responses to distinct microorganisms and to innate cytokines. *Proc. Natl Acad. Sci. USA* **114**, E5434–E5443 (2017).
- Dusseaux, M. et al. Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161<sup>hi</sup> IL-17-secreting T cells. *Blood* **117**, 1250–1259 (2011).
- Cole, S. et al. Interleukin (IL)-12 and IL-18 synergize to promote MAIT cell IL-17A and IL-17F production independently of IL-23 signaling. *Front. Immunol.* **11**, 585134 (2020).
- Rahimpour, A. et al. Identification of phenotypically and functionally heterogeneous mouse mucosal-associated invariant T cells using MR1 tetramers. *J. Exp. Med.* **212**, 1095–1108 (2015).
- Koay, H. F. et al. A three-stage intrathymic development pathway for the mucosal-associated invariant T cell lineage. *Nat. Immunol.* **17**, 1300–1311 (2016).
- Salou, M. et al. A common transcriptomic program acquired in the thymus defines tissue residency of MAIT and NKT subsets. *J. Exp. Med.* **216**, 133–151 (2018).

13. Slichter, C. K. et al. Distinct activation thresholds of human conventional and innate-like memory T cells. *JCI Insight*. **1**, e86292 (2016).
14. Lamichhane, R. et al. Human liver-derived MAIT cells differ from blood MAIT cells in their metabolism and response to TCR-independent activation. *Eur. J. Immunol.* **51**, 879–892 (2021).
15. Gibbs, A. et al. MAIT cells reside in the female genital mucosa and are biased towards IL-17 and IL-22 production in response to bacterial stimulation. *Mucosal Immunol.* **10**, 35–45 (2016).
16. Sobkowiak, M. J. et al. Tissue-resident MAIT cell populations in human oral mucosa exhibit an activated profile and produce IL-17. *Eur. J. Immunol.* **49**, 133–143 (2018).
17. Leng, T. et al. TCR and inflammatory signals tune human MAIT cells to exert specific tissue repair and effector functions. *Cell Rep.* **28**, 3077–3091 (2019).
18. Lamichhane, R. et al. TCR- or cytokine-activated CD8<sup>+</sup> mucosal-associated invariant T cells are rapid polyfunctional effectors that can coordinate immune responses. *Cell Rep.* **28**, 3061–3076 (2019).
19. Voillet, V. et al. Human MAIT cells exit peripheral tissues and recirculate via lymph in steady state conditions. *JCI Insight*. **3**, e98487 (2018).
20. Loh, L. et al. Human mucosal-associated invariant T cells in older individuals display expanded TCRαβ clonotypes with potent antimicrobial responses. *J. Immunol.* **204**, 1119–1133 (2020).
21. Lepore, M. et al. Parallel T-cell cloning and deep sequencing of human MAIT cells reveal stable oligoclonal TCRβ repertoire. *Nat. Commun.* **5**, 3866 (2014).
22. Zumwalde, N. A., Haag, J. D., Gould, M. N. & Gumperz, J. E. Mucosal associated invariant T cells from human breast ducts mediate a Th17-skewed response to bacterially exposed breast carcinoma cells. *Breast Cancer Res.* **20**, 111 (2018).
23. Howson, L. J. et al. MAIT cell clonal expansion and TCR repertoire shaping in human volunteers challenged with *Salmonella* Paratyphi A. *Nat. Commun.* **9**, 253 (2018).
24. Gold, M. C. et al. MR1-restricted MAIT cells display ligand discrimination and pathogen selectivity through distinct T cell receptor usage. *J. Exp. Med.* **211**, 1601–1610 (2014).
25. Eckle, S. B. G. et al. A molecular basis underpinning the T cell receptor heterogeneity of mucosal-associated invariant T cells. *J. Exp. Med.* **211**, 1585–1600 (2014).
26. Kumar, B. V. et al. 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).
27. Poon, M. M. L. et al. Tissue adaptation and clonal segregation of human memory T cells in barrier sites. *Nat. Immunol.* **24**, 309–319 (2023).
28. Milner, J. J. et al. Runx3 programs CD8<sup>+</sup> T cell residency in non-lymphoid tissues and tumours. *Nature* **552**, 253–257 (2017).
29. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
30. Reantragoon, R. et al. Antigen-loaded MR1 tetramers define T cell receptor heterogeneity in mucosal-associated invariant T cells. *J. Exp. Med.* **210**, 2305–2320 (2013).
31. Reantragoon, R. et al. Structural insight into MR1-mediated recognition of the mucosal associated invariant T cell receptor. *J. Exp. Med.* **209**, 761–774 (2012).
32. Patel, O. et al. Recognition of vitamin B metabolites by mucosal-associated invariant T cells. *Nat. Commun.* **4**, 2142 (2013).
33. Koay, H. F. et al. A divergent transcriptional landscape underpins the development and functional branching of MAIT cells. *Sci. Immunol.* **4**, eaay6039 (2019).
34. Legoux, F. et al. Molecular mechanisms of lineage decisions in metabolite-specific T cells. *Nat. Immunol.* **20**, 1244–1255 (2019).
35. Radens, C. M., Blake, D., Jewell, P., Barash, Y. & Lynch, K. W. Meta-analysis of transcriptomic variation in T-cell populations reveals both variable and consistent signatures of gene expression and splicing. *RNA* **26**, 1320–1333 (2020).
36. Kurioka, A. et al. MAIT cells are licensed through granzyme exchange to kill bacterially sensitized targets. *Mucosal Immunol.* **8**, 429–440 (2015).
37. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genom.* **19**, 477 (2018).
38. Cannoodt, R. et al. SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. Preprint at *bioRxiv* <https://doi.org/10.1101/079509> (2016).
39. Rauen, T., Hedrich, C. M., Tenbrock, K. & Tsokos, G. C. cAMP responsive element modulator: a critical regulator of cytokine production. *Trends Mol. Med.* **19**, 262–269 (2013).
40. Kongsbak, M., Levring, T. B., Geisler, C. & von Essen, M. R. The vitamin D receptor and T cell function. *Front. Immunol.* **4**, 148 (2013).
41. Laurence, A. et al. Interleukin-2 signaling via STAT5 constrains T helper 17 cell generation. *Immunity* **26**, 371–381 (2007).
42. Clambey, E. T. et al. The Ikaros transcription factor regulates responsiveness to IL-12 and expression of IL-2 receptor alpha in mature, activated CD8 T cells. *PLoS ONE* **8**, e57435 (2013).
43. Krovi, S. H. et al. Thymic iNKT single cell analyses unmask the common developmental program of mouse innate T cells. *Nat. Commun.* **11**, 6238 (2020).
44. Kurachi, M. et al. The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8<sup>+</sup> T cells. *Nat. Immunol.* **15**, 373–383 (2014).
45. Lett, M. J. et al. Stimulatory MAIT cell antigens reach the circulation and are efficiently metabolised and presented by human liver cells. *Gut* **71**, 2526–2538 (2022).
46. FitzPatrick, M. E. B. et al. Human intestinal tissue-resident memory T cells comprise transcriptionally and functionally distinct subsets. *Cell Rep.* **34**, 108661 (2021).
47. Bister, J. et al. Human endometrial MAIT cells are transiently tissue resident and respond to *Neisseria gonorrhoeae*. *Mucosal Immunol.* **14**, 357–365 (2020).
48. Youssef, G. B. et al. Ontogeny of human mucosal-associated invariant T cells and related T cell subsets. *J. Exp. Med.* **215**, 459–479 (2018).
49. Vorkas, C. K. et al. Single-cell transcriptional profiling reveals signatures of helper, effector, and regulatory MAIT cells during homeostasis and activation. *J. Immunol.* **208**, 1042–1056 (2022).
50. Chandra, S. et al. Transcriptomes and metabolism define mouse and human MAIT cell heterogeneity. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.12.20.473182> (2021).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

## Methods

### Data generation

**Liver tissue collection and processing (Exp 1 and 2).** Liver tissue ( $n = 7$ ) and matched blood ( $n = 6$ ) were obtained from patients undergoing liver resection at the Churchill Hospital, Oxford, UK and the University Hospital Basel, Basel, Switzerland (Supplementary Table 1). Patients had no chronic liver disease, active excess alcohol consumption (>14 g per day), infection, immunosuppression or family history of liver disease.

Disease-free liver tissue was collected from the resection margin, cut into small pieces with a scalpel and ground through a 70  $\mu\text{m}$  cell strainer. Cells were washed with R10 (RPMI-1640 (Sigma-Aldrich), 10% FBS (Sigma-Aldrich), 1% penicillin–streptomycin (Thermo Fisher Scientific); 931g, 10 min, 4 °C) and mononuclear cells isolated by density gradient centrifugation on a discontinuous 35%/70% Percoll (GE Healthcare) gradient (931g, 20 min, 21 °C, no brake). Mononuclear cells were collected from the interface and washed with R10 (596g, 10 min, 4 °C). Residual red blood cells were lysed with ACK for 3–5 min. Cells were washed twice (596g, 10 min, 4 °C) and cryopreserved (90% FBS, 10% DMSO (Sigma-Aldrich)) in liquid nitrogen.

**Ethics statement.** Samples were obtained with written informed consent through the Oxford Gastrointestinal Illnesses Biobank (REC ref. 16/YH/0247) or under Ethikkommission Nordwest- und Zentralschweiz (EKNZ) numbers EKNZ-2014-362, EKNZ-2016-01188 and EKNZ-2019-02118.

**Peripheral blood mononuclear cell (PBMC) isolation.** PBMCs were isolated from fresh whole blood by density gradient centrifugation (Lymphoprep, Axis-Shield) at 931g for 30 min with no brake. Cells were cryopreserved in liquid nitrogen and thawed in complete medium (R10, 1X nonessential amino acids (Thermo Fisher Scientific), 1 mM sodium pyruvate (Thermo Fisher Scientific), 10 mM HEPES (pH 7.0–7.5; Thermo Fisher Scientific), 50  $\mu\text{M}$   $\beta$ -mercaptoethanol (Thermo Fisher Scientific)) on the day of use.

**Stimulation of isolated CD8<sup>+</sup>/CD3<sup>+</sup> T cells for scRNA-seq and scTCR-seq (Exp 3 and 4) or activation marker/cytokine validation.** Pierce streptavidin-coated high-capacity flat-bottom 96-well plates (Thermo Fisher Scientific) were coated with 50  $\mu\text{l}$  biotinylated MR1/5-OP-RU monomer (NIH Tetramer Core Facility) at 10  $\mu\text{g}$  per ml in PBS (Sigma-Aldrich) overnight at 4 °C. Cryopreserved PBMCs were thawed in complete medium. CD8<sup>+</sup> T cells were isolated using CD8 MicroBeads (Exp 3; Miltenyi Biotec) and CD3<sup>+</sup> T cells using the REAlease CD3 MicroBead Kit (Exp 4 and validation experiments; Miltenyi Biotec) following the manufacturer's instructions. Isolated CD8<sup>+</sup>/CD3<sup>+</sup> T cells were washed in complete medium and resuspended at  $1 \times 10^7$  cells per ml. One million (20 h stimulation) or 500,000 (68 h stimulation) cells were added per well to the appropriate 96-well plates (MR1/5-OP-RU-coated plate for TCR and TCR+cytokine stimulation, round-bottom plate for unstimulated and cytokine stimulation). IL-12 (50 ng ml<sup>-1</sup>; R&D Systems) and IL-18 (50 ng ml<sup>-1</sup>; R&D Systems) were added for cytokine stimulation;  $\alpha\text{CD}28$  (1  $\mu\text{g}$  ml<sup>-1</sup>; clone: CD28.2; BioLegend) for TCR stimulation; IL-12, IL-18 and  $\alpha\text{CD}28$  for TCR+cytokine stimulation; and complete medium for unstimulated cells (final volume 200  $\mu\text{l}$  per well). Cells were incubated for 20 h or 68 h at 37 °C, 5% CO<sub>2</sub>. For intracellular cytokine staining, brefeldin A (BioLegend) and monensin (BioLegend) were added for the final 4 h.

**Tetramer staining (Exp 1 and 2).** Biotinylated human MR1/5-OP-RU and MR1/6-FP monomers were provided by the NIH Tetramer Core Facility. Tetramers were generated using streptavidin-PE (high concentration) or streptavidin-BV421 (both BioLegend) following the NIH Tetramer Core Facility protocol. Tetramer staining was performed for 40 min

at 21 °C in FACS buffer (PBS, 0.5% BSA (Sigma-Aldrich), 1 mM EDTA (Sigma-Aldrich)).

**Surface staining and cell sorting for scRNA-seq and scTCR-seq (Exp 1–4).** TotalSeq-C hashtag antibodies (BioLegend) were used in Exp 2 and 4. Hashtag antibody dilutions were prepared according to the manufacturer's instructions. Namely, antibody vials were centrifuged at 10,000g, 30 s, 4 °C, before antibody dilution in FACS buffer. Diluted hashtags were centrifuged at 14,000g, 10 min, 4 °C. Cells were incubated in Human TruStain FcX (BioLegend) for 10 min at 4 °C before the addition of diluted hashtag antibodies (0.2  $\mu\text{g}$  per well) for 10 min at 4 °C. Surface fluorochrome-conjugated antibodies were added without washing off the hashtag antibodies. Surface staining was performed in Brilliant Stain Buffer Plus (BD Biosciences) for 30 min at 4 °C. Cells were washed twice in PBS with 0.5% BSA, resuspended in presort buffer (PBS, 1% BSA, 25 mM HEPES) containing 3–5 nM SYTOX Green Nuclear Acid Stain (Thermo Fisher Scientific) and incubated for 20 min at 4 °C. Cells were sorted on a BD FACSAria III with an 85  $\mu\text{m}$  nozzle. Sorted cells were collected in RPMI-1640, 10% FBS, 25 mM HEPES, or HBSS (Thermo Fisher Scientific), 50% FBS, 25 mM HEPES. Sort purity was >99%. For Exp 2 and 4, sorted cells were stained with the TotalSeq-C Human Universal Cocktail V1.0 (BioLegend) according to the manufacturer's instructions. Staining reagents are listed in Supplementary Table 12.

**Stimulation of CD56<sup>-</sup> and CD56<sup>+</sup> MAIT cells.** CD3<sup>+</sup> T cells were isolated using the REAlease CD3 MicroBead Kit following the manufacturer's instructions. Surface antibody and live/dead (SYTOX Green Nuclear Acid Stain) staining was performed as above, then CD56<sup>-</sup> and CD56<sup>+</sup> MAIT cells (V $\alpha$ 7.2<sup>+</sup>CD161<sup>hi</sup>) were sorted on a BD FACSAria III with an 85  $\mu\text{m}$  nozzle. Sorted cells were collected in HBSS, 50% FBS, 25 mM HEPES, then centrifuged at 400g, 5 min, 21 °C and incubated overnight at 37 °C, 5% CO<sub>2</sub>. Rested cells were washed in complete medium, plated in a 96-well round-bottom plate and stimulated with IL-12 (50 ng ml<sup>-1</sup>) and IL-18 (50 ng ml<sup>-1</sup>) at 37 °C, 5% CO<sub>2</sub> for 20 h, with the addition of brefeldin A and monensin for the final 4 h.

**Surface marker and intracellular cytokine staining for flow cytometry.** Surface staining was performed in Brilliant Stain Buffer Plus for 30 min at 4 °C. Stained cells were washed twice in FACS buffer. For intracellular cytokine staining, cells were fixed in Cytofix/Cytoperm (BD Biosciences) for 20 min at 4 °C, then washed twice in 1X Perm/Wash (BD Biosciences). Intracellular staining was performed in 1X Perm/Wash for 30 min at 4 °C. Cells were acquired on a BD LSR II flow cytometer with BD FACSDiva Software (v8.0.1). Staining reagents are listed in Supplementary Table 12.

**10x Genomics library generation and sequencing.** Sequencing libraries were generated using 10x Genomics Chromium Single Cell V(D)J Reagent Kits (v1.0 Chemistry; Exp 1 and 3) or 10x Genomics Chromium Next GEM Single Cell 5' Reagent Kits v2 (Dual Index; Exp 2 and 4) following the manufacturer's instructions. For Exp 1 and 3, cells were loaded onto the Chromium Controller (10x Genomics) at a concentration of  $1 \times 10^6$  cells per ml, with 6,000–8,000 cells loaded per channel. For Exp 2 and 4, 17,750–30,000 cells were loaded per channel. Library quality and concentration were assessed using a TapeStation (Agilent) and Qubit 2.0–4 Fluorometer (Thermo Fisher Scientific), respectively. Library generation for Exp 1 and 3 was performed at the Oxford Genomics Centre (Wellcome Centre for Human Genetics, University of Oxford), and for Exp 2 and 4 was performed in-house. Libraries were sequenced on an Illumina HiSeq 4000 (Exp 1) or Illumina NovaSeq 6000 (Exp 2–4) at the Oxford Genomics Centre. Sequencing depths were as follows: Exp 1–39,013–46,998 reads per cell for scRNA-seq, 10,323–30,883 reads per cell for scTCR-seq; Exp 2–76,710–89,378 reads per cell for scRNA-seq, 19,673–41,612 reads per cell for TotalSeq-C feature barcoding antibodies, 5,192–7,065 reads per cell for scTCR-seq; Exp



3–75,638–88,871 reads per cell for scRNA-seq, 12,517–27,197 reads per cell for scTCR-seq; Exp 4–53,120–102,777 reads per cell for scRNA-seq, 9,333–17,717 reads per cell for TotalSeq-C feature barcoding antibodies, 2,904–14,636 reads per cell for scTCR-seq.

**ATAC-seq library generation and sequencing.** Naïve T cells (CD8<sup>+</sup>CD45RO<sup>+</sup>CCR7<sup>+</sup>), MAIT cells (CD8<sup>+</sup>CCR7<sup>+</sup>MR1/5-OP-RU<sup>+</sup>) and T<sub>mem</sub> cells (CD8<sup>+</sup>CCR7<sup>+</sup>MR1/5-OP-RU<sup>+</sup>) were sorted from CD8-enriched (CD8 MicroBeads) PBMCs ( $n = 3$  donors, 50,000 cells per population). ATAC-seq was performed as previously described<sup>51</sup>. Briefly, cells were pelleted at 500g for 10 min at 4 °C and the supernatant was removed. Cells were resuspended in 50 µl cold lysis and transposition mix (25 µl TD buffer, 2.5 µl TDE1 (Illumina, FC-121-1030; product discontinued), 22 µl nuclease-free H<sub>2</sub>O (Thermo Fisher Scientific), 0.5 µl 1% digitonin (Promega)) and incubated for 30 min at 37 °C with agitation at 300 rpm (Thermo-Shaker TS-100, Biosan). Transposed DNA was purified using the Qiagen MinElute Reaction Cleanup Kit and eluted in 13 µl elution buffer (10 mM Tris–HCl, pH 8). Purified library fragments were PCR amplified for 11 cycles with barcoded primers using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs). Amplified DNA was purified using the Qiagen MinElute PCR Purification Kit (23 µl elution volume) and PCR primer contamination was removed using SPRI beads (5 min dry time, 15 µl elution volume; Agencourt AMPure XP PCR Purification, Beckman Coulter). Fragment size distribution was analyzed using a 2100 Bioanalyzer (Agilent) with the High Sensitivity DNA Kit. Libraries were quantified using the KAPA Library Quantification Kit (Roche). Paired-end sequencing (40 bp) was performed on an Illumina NextSeq 500 using the High Output v2 Kit (75 cycles). Libraries were sequenced to a depth of 215–271 million paired-end reads per sample.

## Data analysis

**10x Genomics raw data processing.** FASTQ files were generated from BCL files using Illumina bcl2fastq. For Exp 1 and 3, FASTQ files for gene expression and TCR data were processed using Cell Ranger (v3.0.1–3.0.2; <https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger>) count and vj pipelines, respectively. For Exp 2 and 4, FASTQ files for all modalities were processed using the Cell Ranger (v7.0.1) multi pipeline. For TCR analysis, the filtered\_contig\_annotations.csv file was filtered to retain only high-confidence, full-length, productive contigs corresponding to TCR $\alpha$  or TCR $\beta$  chains.

**Hashtag demultiplexing.** Hashtag demultiplexing (Exp 2 and 4) was performed using the consensus calling approach from cellhashR<sup>52</sup> (v1.0.3) with the following methods: BFF<sub>cluster</sub><sup>52</sup>, BFF<sub>raw</sub><sup>52</sup>, GMM-Demux<sup>53</sup>, MULTI-seq<sup>54</sup>, Seurat HTODemux<sup>55</sup> and DropletUtils hashedDrops<sup>55</sup>.

**Quality control.** Quality control was performed separately for cells from each channel of the Chromium Controller. Filtered feature-barcode matrices from Cell Ranger count/multi were imported into R using Seurat (v4.0.3–4.3.0)<sup>56</sup>. Cells with low unique molecular identifier counts, low gene counts and/or a high percentage of mitochondrial reads, were removed. For Exp 1 and 2, cells labeled as empty droplets or damaged cells by DropletQC<sup>57</sup> (v0.0.0.9000) were removed (damaged cells in Exp 4 were also removed). For Exp 2 and 4, only cells called as consensus singlets by hashtag demultiplexing were retained. Cells with two TCR $\alpha$  and two TCR $\beta$  chains, or more than two TCR $\alpha$  and/or TCR $\beta$  chains, were assumed to be doublets and discarded. TCR and BCR genes were removed to ensure downstream clustering analysis was not influenced by TCR or BCR chain usage.

**Normalization, integration, dimensionality reduction and clustering (Exp 1 and 2).** For combined analysis of Exp 1 and 2, data from each donor were normalized separately using sctransform<sup>58</sup> (v0.3.5). Highly variable genes (HVGs) were defined as the 3,000 genes with the largest

residual variance following variance stabilizing transformation. Cells from different donors were integrated using Seurat<sup>56</sup>. Integration features ( $n = 3,000$ ) were selected using matched blood and liver samples, with STACAS<sup>59</sup> (v2.0.1) blacklisted genes subsequently removed. For anchor finding, dimensionality reduction was performed using canonical correlation analysis (MAIT cells only) or reciprocal principal component analysis (PCA; MAIT and T<sub>mem</sub> cells combined). The number of dimensions used for identifying and weighting anchors was selected empirically by performing integration with multiple input dimensions and evaluating downstream clustering results. Following integration, dimensionality reduction was performed using PCA. Scree plots were used to determine how many PCs to use for UMAP generation and clustering. Cell clusters were identified using Seurat's graph-based clustering approach. Briefly, a shared nearest neighbor graph was constructed using dimensionally-reduced data, and then clusters were determined by optimizing the standard modularity function (Louvain algorithm).

**Normalization, dimensionality reduction, batch correction and clustering (Exp 3 and 4).** Per experiment and timepoint, data from all donors combined were normalized using sctransform<sup>58</sup> (v0.3.2–0.3.5). HVGs were defined as the 3,000 genes with the largest residual variance following variance stabilizing transformation (in Exp 4, STACAS<sup>59</sup> blacklisted genes were removed from HVGs). Dimensionality reduction was performed using PCA. Batch correction for donor was performed using Harmony<sup>60</sup> (v0.1.1) with 50 input PCs ( $\theta = 2, \lambda = 1$ ). Scree plots were used to determine how many PCs to use for UMAP generation and clustering. Cell clusters were identified using Seurat's graph-based clustering approach.

**Differential gene expression analysis.** Differential gene expression analysis between clusters was performed using MAST<sup>61</sup> (v1.18.0–1.24.1; FindMarkers function from Seurat) with cellular detection rate as a covariate. Cluster markers were defined as genes with significantly increased expression in one cluster relative to the average of all other clusters (fold change > 1.25 and adjusted  $P < 0.05$  based on Bonferroni correction using all genes in the dataset). Differential gene expression analysis between conditions (for example, tissues, coreceptors and stimuli) was performed using MAST<sup>61</sup> with cellular detection rate and donor as covariates. Genes with a fold change > 1.25 and a Bonferroni adjusted  $P < 0.05$  were defined as significantly differentially expressed. Input data were log-transformed normalized counts generated by global-scaling normalization (NormalizeData function from Seurat).

**Coreceptor assignment (Exp 1 and 2).** Cells were defined as CD8<sup>+</sup>, DN or CD4<sup>+</sup> based on normalized coreceptor gene or protein (measured using TotalSeq-C antibodies) expression. For assignment based on gene expression, cells were defined as CD8<sup>+</sup> if  $CD8A > 0$  and/or  $CD8B > 0$ , CD4<sup>+</sup> if  $CD4 > 0$ , and DN if  $CD8A, CD8B$  and  $CD4$  were undetected. For assignment based on protein expression (Exp 2 only), cells were defined as CD8<sup>+</sup> if  $CD8 > 0.3$ , CD4<sup>+</sup> if  $CD4 > 2.5$  and  $CD8 < 0.2$ , and DN if neither CD8<sup>+</sup> nor CD4<sup>+</sup>. Thresholds were selected empirically by examining histograms of normalized count data.

**Pseudotime analysis.** Pseudotime analysis was performed using Slingshot<sup>37</sup> (v2.0.0) and SCORPIUS<sup>38</sup> (v1.0.8). UMAP coordinates and Seurat cluster labels (0.1 resolution) were provided as input to Slingshot, with the main unstimulated cluster specified as the start of the trajectory. Normalized expression values (sctransform) were provided as input to SCORPIUS. Two separate SCORPIUS trajectories were generated from unstimulated and TCR-stimulated cells, and unstimulated and cytokine-stimulated cells. Gene importance along SCORPIUS trajectories was determined using random forest regression (gene\_importances function, num\_permutations = 10). Differential gene importance was calculated by taking the ratio of gene importance ranks on the TCR and cytokine trajectories (higher rank as numerator). Genes defined



as differentially important had an importance false discovery rate (FDR) < 0.05 and were within the top 150 most highly ranked genes for either the TCR or cytokine trajectory.

**Transcription factor regulon analysis.** Transcription factor regulons were identified using SCENIC<sup>29,62</sup> (pySCENIC v0.11.2–0.12.1). Briefly, the raw expression matrix was filtered to retain genes expressed in >1% of cells and with a count >3 × 0.01 × number of cells. Modules comprising transcription factors and coexpressed genes were generated using GRNBoost2, then pruned to remove indirect targets lacking enrichment for the corresponding transcription factor motif (cisTarget). This resulted in a set of transcription factor regulons. Due to stochasticity in gene regulatory network inference using GRNBoost2, each pySCENIC run can identify a different number of regulons, as well as different target genes for each transcription factor. Thus, pySCENIC was run 100 times. High-confidence regulons were defined as regulons that occurred in >80% of runs and that contained at least five high-confidence target genes. High-confidence target genes were those found within a regulon in >80% of runs. Cells were scored for the activity of each high-confidence regulon (including only high-confidence target genes) using AUCell (v1.16.0–1.20.1). Regulons differentially active between tissues or stimulation conditions were determined using MAST<sup>61</sup> with donor as a covariate. Regulons with a Bonferroni adjusted  $P < 0.01$  were defined as differentially active. Smoothed regulon activity scores (AUCell scores) over SCORPIUS trajectories were generated by loess regression (loess function from the stats R package).

**Gene set enrichment analysis (GSEA) and over-representation analysis.** GSEA<sup>63</sup> (v4.3.2) for published mouse and human T<sub>RM</sub> cell gene signatures<sup>27,28</sup> was performed using pseudobulk gene counts (normalized gene counts summed for all cells within a sample) with 1,000 gene permutations. Over-representation analysis for gene ontology (GO) terms and Reactome pathways was performed using clusterProfiler<sup>64</sup> (v4.7.1) and ReactomePA<sup>65</sup> (v1.36.0), respectively. Gene symbols were converted to Entrez IDs using the Bioconductor org.Hs.eg.db annotation package (v3.13.0–3.16.0). Background genes were defined as genes expressed (count >0) in ≥1% of cells (or for Exp 3, genes expressed in at least five cells). Redundant enriched GO terms were removed using the simplify function from clusterProfiler.

**Gene lists.** MAIT1 and MAIT17 gene signatures were generated by overlapping MAIT1 and MAIT17 genes from two published scRNA-seq datasets<sup>33,34</sup>. Mouse gene symbols were converted to human gene symbols using the biomaRt R package (getLDS function; v2.54.1) with the ENSEMBL\_MART\_ENSEMBL BioMart database and the hsapiens\_gene\_ensembl and mmusculus\_gene\_ensembl datasets. Human Th1 and Th17 gene signatures were generated by combining genes from the NanoString nCounter Human Immunology V2 Panel Gene List (<https://nanosttring.com/support-documents/ncounter-human-immunology-v2-panel-gene-list>) and a meta-analysis published by Radens et al.<sup>35</sup>. Interferon-stimulated genes were obtained from Schoggins and Rice<sup>66</sup>.

**Nucleotide and functional TCR clonotypes.** Tables of TCRα and TCRβ usage for each cell in a sample were combined to generate one table per donor for nucleotide clonotype calling, and one table for all donors for functional clonotype calling. MAIT cells were required to have a *TRAVI-2* TCRα chain and at least one TCRβ chain. T<sub>mem</sub> cells were required to have at least one TCRα chain and at least one TCRβ chain. Cells expressing *TRAVI-2* paired with *TRAJ33*, *TRAJ12* or *TRAJ20* and with a 12 amino acid CDR3α region were assumed to be contaminating MAIT cells and were removed before T<sub>mem</sub> cell clonotype calling.

Nucleotide clonotypes (TCRαβ) were defined as cells with identical TCR gene segment usage, and CDR3α and CDR3β nucleotide sequences. TCRα clonotypes and TCRβ clonotypes were defined as cells with identical TCRα segment usage and CDR3α sequences, or

identical TCRβ segment usage and CDR3β sequences, respectively. TCRαβ clonotypes were numbered according to size, with clonotype 1 being the largest, clonotype 2 being the second largest and so on. Clonotypes of identical size were randomly ordered for numbering. TCRαβ clonotypes were assigned ranks in a similar manner, but clonotypes of identical size were given the same rank.

Functional clonotypes (TCRαβ) were defined as cells with at least one identical TCRα and TCRβ chain amino acid sequence (gene segment usage and CDR3 sequences). Functional TCRα and functional TCRβ clonotypes were defined as cells with at least one matching TCRα or TCRβ chain amino acid sequence, respectively. Given the presence of TCR dropout, functional clonotypes were permitted to contain a mixture of cells with one or two TCRα or TCRβ chains, providing all detected chains matched those within the clonotype.

**TCR analyses.** TCR analyses were performed only for cells with a defined TCR clonotype. The Shannon diversity index was calculated using the diversity function from the vegan R package (v2.6.4). To test for an association between clonotype (clonotypes from  $n = 12$  donors with ≥20 cells) and cluster, a multinomial test was performed using the EMT R package (v1.3)–MonteCarlo = FALSE when the number of distinct possible outcomes (events) <  $1 \times 10^6$ , else MonteCarlo = TRUE with ntrial =  $10 \times$  events or  $1 \times 10^8$  (whichever smaller).  $P$  values were Bonferroni adjusted for the number of clonotypes tested per donor. Sequence logos were generated using ggseqlogo (v0.1). The overall height of the stacked letters at each position indicates the sequence conservation, while the relative abundance of each amino acid is indicated by the height of individual letters within the stack. Acidic bases are shown in red, basic residues in blue, hydrophobic in black and polar in green. The number of N-nucleotides and P-nucleotides in 36-nucleotide CDR3α sequences was determined using IMGT/JunctionAnalysis<sup>67</sup> (v2.3.0).

**ATAC-seq.** Read quality was checked using FastQC (v0.11.5; Babraham Bioinformatics) and adapter sequences were removed using Trimmomatic<sup>68</sup> (v0.36). Reads were mapped to hg38 using Bowtie 2 (ref. 69; --very-sensitive -X 2000 --no-mixed --no-discordant --no-unal; v2.3.4.1). BAM files from multiple sequencing lanes were merged using samtools<sup>70</sup> (v1.6) and duplicate reads were removed using Picard MarkDuplicates (v2.15.0). BAM files were filtered to remove reads mapping to the mitochondrial genome (samtools) and blacklisted regions (bedtools<sup>71</sup> v2.26.0). Blacklisted regions comprised the ENCODE blacklist<sup>72</sup> and a custom ATAC-seq blacklist generated by J. Buenrostro<sup>73</sup>. Read start sites were adjusted to correspond to the center of the transposase binding site—reads on the forward strand were offset by +5 bp and reads on the reverse strand were offset by -4 bp using deepTools<sup>74</sup> alignmentSieve (v3.1.0). Cut sites were identified as the 5' ends of forward and reverse reads. Peak calling was performed with MACS2 (ref. 75; --nomodel -p 0.1 -f BAMPE --call-summits; v2.1.1) and an optimal peak list defined using the irreproducibility discovery rate framework<sup>76</sup>. Cut sites within peaks were quantified using featureCounts<sup>77</sup> (v1.6.0). Peaks were annotated to genes using HOMER<sup>78</sup> annotatePeaks.pl (v4.8). Differentially accessible peaks (fold change > 2, FDR < 0.05) were identified using edgeR<sup>79</sup> (v3.24.3). BAM files were converted into BigWig files using deepTools<sup>74</sup> bamCoverage (--normalizeUsing CPM --ignoreForNormalization chrM chrX chrY --binSize 1; v3.5.1).

**Statistics and reproducibility.** Statistical tests are listed in the relevant sections of the Methods and/or in the figure legends. Standard statistical tests were performed in R (v4.1.1–4.2.0) using the stats or rstatix (v0.7.2) packages. All tests were two-sided. Initial blood-liver (Exp 1) and stimulation (Exp 3) scRNA-seq experiments were followed by validation experiments (Exp 2 and 4)—the findings of these were highly concordant. Genes of interest from Exp 1 and 3 were validated at the protein level by CITE-seq (Exp 2 and 4) and flow cytometry.

No statistical method was used to predetermine sample size. No samples were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Plots.** Most plots were generated using the ggplot2 R package (v3.3.4–3.4.2). FACS plots were generated in FlowJo (v10.8.1; BD Biosciences). Heatmaps were generated using pheatmap (v1.0.12) or ComplexHeatmap (v2.14.0). GSEA plots were generated in GraphPad Prism (v9.5.1; GraphPad Software, LLC). ATAC-seq traces were generated using Integrative Genomics Viewer (v2.16.0).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Sequencing data generated in this study have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO SuperSeries accession number [GSE194189](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194189). The Bioconductor org.Hs.eg.db annotation package and ENSEMBL\_MART\_ENSEMBL BioMart database are publicly available. Source data are provided with this paper.

### Code availability

No custom software was generated for this manuscript. The software used for all analyses is listed in the relevant sections of the Methods.

### References

51. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
52. Boggy, G. J. et al. BFF and cellhashR: analysis tools for accurate demultiplexing of cell hashing data. *Bioinformatics* **38**, 2791–2801 (2022).
53. Xin, H. et al. GMM-Demux: sample demultiplexing, multiplet detection, experiment planning, and novel cell-type verification in single cell sequencing. *Genome Biol.* **21**, 188 (2020).
54. McGinnis, C. S. et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
55. Stoeckius, M. et al. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
56. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
57. Muskovic, W. & Powell, J. E. DropletQC: improved identification of empty droplets and damaged cells in single-cell RNA-seq data. *Genome Biol.* **22**, 329 (2021).
58. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
59. Andreatta, M. & Carmona, S. J. STACAS: Sub-type anchor correction for alignment in Seurat to integrate single-cell RNA-seq data. *Bioinformatics* **37**, 882–884 (2021).
60. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
61. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
62. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
63. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
64. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb.)* **2**, 100141 (2021).
65. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
66. Schoggins, J. W. & Rice, C. M. Interferon-stimulated genes and their antiviral effector functions. *Curr. Opin. Virol.* **1**, 519–525 (2011).
67. Giudicelli, V. & Lefranc, M.-P. IMGT/JunctionAnalysis: IMGT standardized analysis of the V-J and V-D-J junctions of the rearranged immunoglobulins (IG) and T cell receptors (TR). *Cold Spring Harb. Protoc.* **2011**, 716–725 (2011).
68. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
69. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
70. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
71. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
72. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
73. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
74. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
75. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
76. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
77. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
78. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
79. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

### Acknowledgements

We thank biobankers of the Translational Gastroenterology Unit and surgeons of Oxford University Hospitals NHS Foundation Trust for collecting patient samples, especially M. Silva for facilitating this; S. Slevin for liver processing; K. Lynch for liver patient identification and sample processing; M. Salio for the plate-bound MR1 stimulation protocol; H. Ferry for sorting; D. Sims and D. Agarwal for discussions regarding computational analysis; and the NIH Tetramer Core Facility for providing MR1 monomers. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome grant ref. 203141/A/16/Z) for the generation and initial processing of the sequencing data. Analysis was performed using computer systems at the MRC WIMM Centre for Computational Biology. L.C.G. and A.A. are supported by Wellcome (109028/Z/15/Z and 216417/Z/19/Z, respectively). M.E.B.F. is supported by Beyond

Celiac and the Academy of Medical Sciences. M.F.S. is supported by the Swiss National Science Foundation (PZ00P3\_167828 and PZ00P3\_189490), Goldschmidt-Jacobson Stiftung and Uniscientia Stiftung. N.M.P. is supported by an Oxford-UCB Postdoctoral Fellowship. P.K. is supported by Wellcome (222426/Z/21/Z), the NIH (U19 I082360), the NIHR Oxford Biomedical Research Centre and an NIHR Senior Fellowship.

### Author contributions

L.C.G., N.M.P. and P.K. designed the project and the experiments. L.C.G., A.A., M.E.B.F., M.J.L., G.F.H., M.F.S., N.M.P. and P.K. recruited patients and/or performed liver tissue processing. L.C.G. performed all other experiments and all data analysis. L.C.G. wrote the manuscript. All authors critically reviewed and approved the manuscript.

### Competing interests

P.K. has acted as a consultant to UCB, Biomunex, AstraZeneca and Infinitopes. N.M.P. has acted as a consultant to Infinitopes. The other authors declare no competing interests.

### Additional information

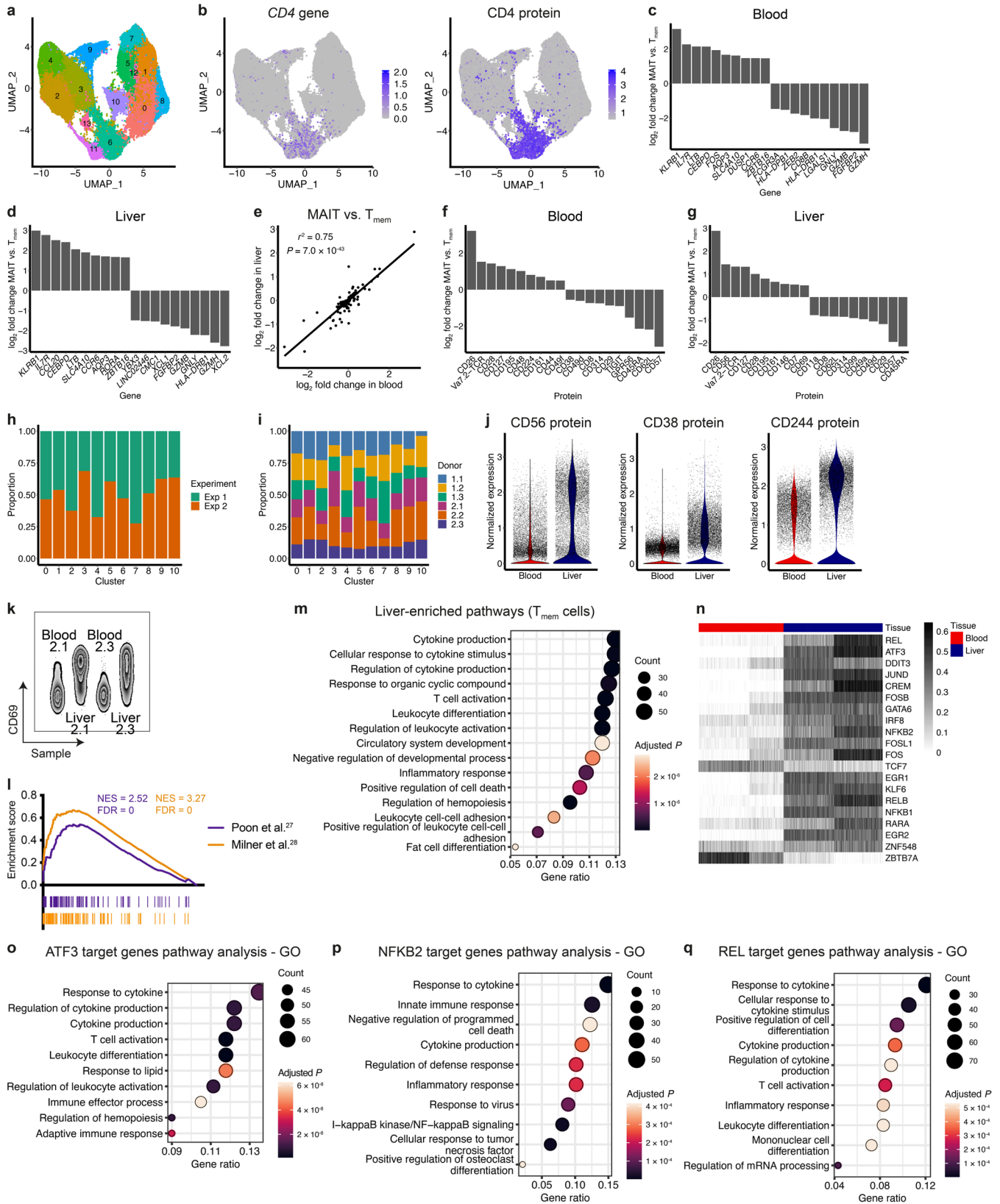
**Extended data** is available for this paper at <https://doi.org/10.1038/s41590-023-01575-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41590-023-01575-1>.

**Correspondence and requests for materials** should be addressed to Lucy C. Garner or Paul Klenerman.

**Peer review information** *Nature Immunology* thanks Johan Sandberg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: L.A. Dempsey, in collaboration with the *Nature Immunology* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

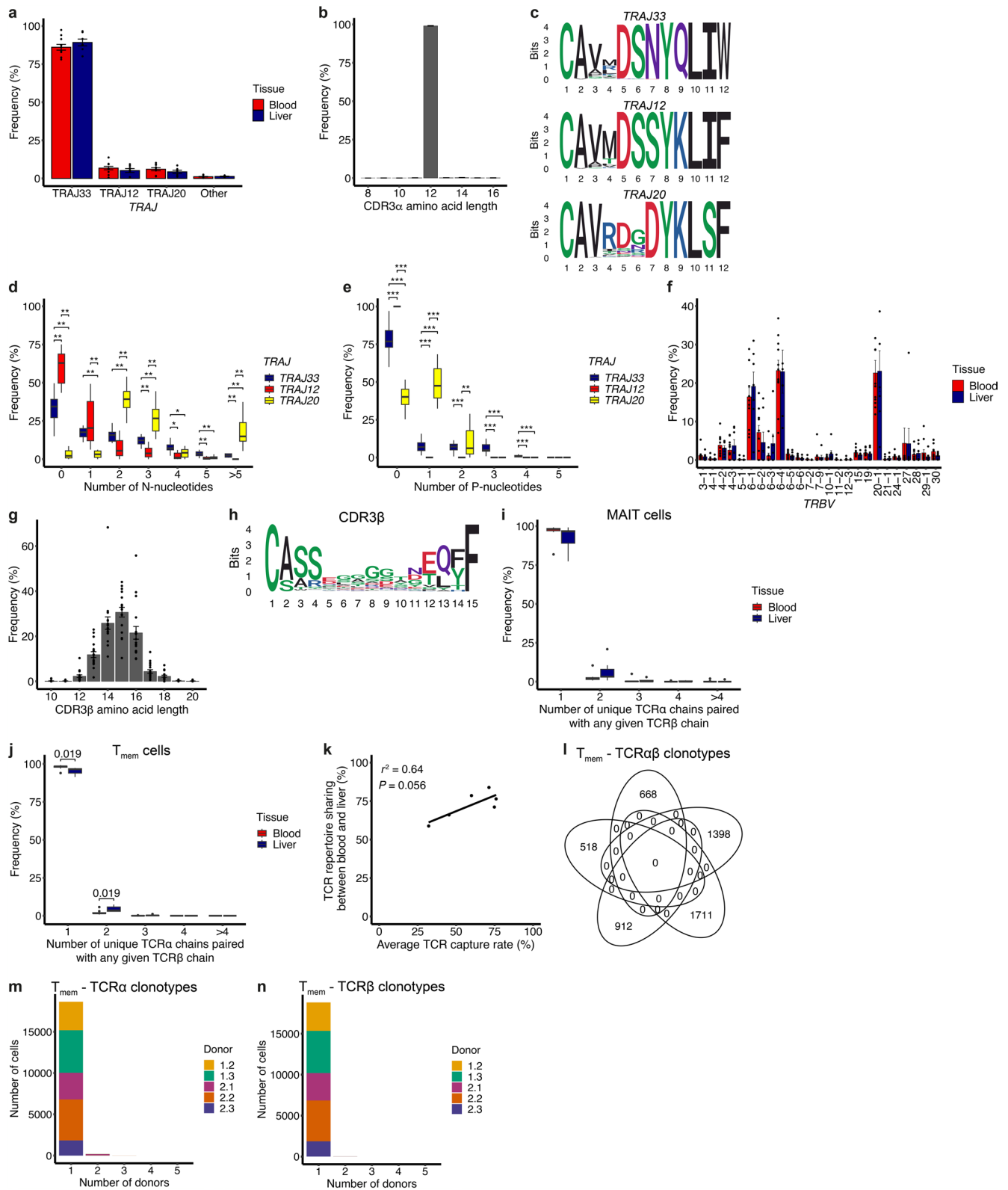


Extended Data Fig. 1 | See next page for caption.



**Extended Data Fig. 1 | MAIT and T<sub>mem</sub> cells, and blood and liver MAIT cells, exhibit distinct phenotypes and transcriptomes.** **a**, UMAP of blood and liver MAIT cells and conventional memory T (T<sub>mem</sub>) cells colored by the 14 identified clusters. *n* = 89,456 cells from 12 donors (11 blood, 7 liver). **b**, UMAPs of blood and liver MAIT and T<sub>mem</sub> cells colored by CD4 gene (left) and protein (right) expression. **c,d**, Bar plots showing the log<sub>2</sub> fold change in expression of the top ten genes upregulated in MAIT cells compared with T<sub>mem</sub> cells, and vice versa, in the blood (**c**) and liver (**d**). **e**, Pearson's correlation between the log<sub>2</sub> fold change in protein expression between MAIT and T<sub>mem</sub> cells in the blood, and MAIT and T<sub>mem</sub> cells in the liver. **f,g**, Bar plots showing the log<sub>2</sub> fold change in expression of the top ten proteins upregulated in MAIT cells compared with T<sub>mem</sub> cells, and vice versa, in the blood (**f**) and liver (**g**). **h,i**, Proportion of cells in each cluster (analysis of MAIT cells from six matched blood-liver donors) from Exp 1 and Exp 2 (**h**) and from each donor (**i**). **j**, Expression of CD56, CD38 and CD244 (2B4)

proteins in blood and liver MAIT cells. **k**, Flow cytometry plot showing CD69 expression on blood and liver MAIT cells from two representative donors. **l**, Gene set enrichment analysis of liver compared with blood T<sub>mem</sub> cells using published human and mouse tissue-resident memory T cell gene signatures. NES, normalized enrichment score. **m**, Over-representation analysis on the genes significantly upregulated in liver T<sub>mem</sub> cells compared with blood T<sub>mem</sub> cells. Top 15 gene ontology (GO) terms and associated Benjamini–Hochberg adjusted *P* values are shown. **n**, Heatmap showing activity (row-scaled AUCell scores) of the 20 most differentially active regulons (largest difference in average AUCell score) between matched blood and liver MAIT cells in Exp 2. *n* = 3 donors. **o–q**, Over-representation analysis on predicted ATF3 (**o**), NFκB2 (**p**) and REL (**q**) target genes. Top ten GO terms and associated Benjamini–Hochberg adjusted *P* values are shown.

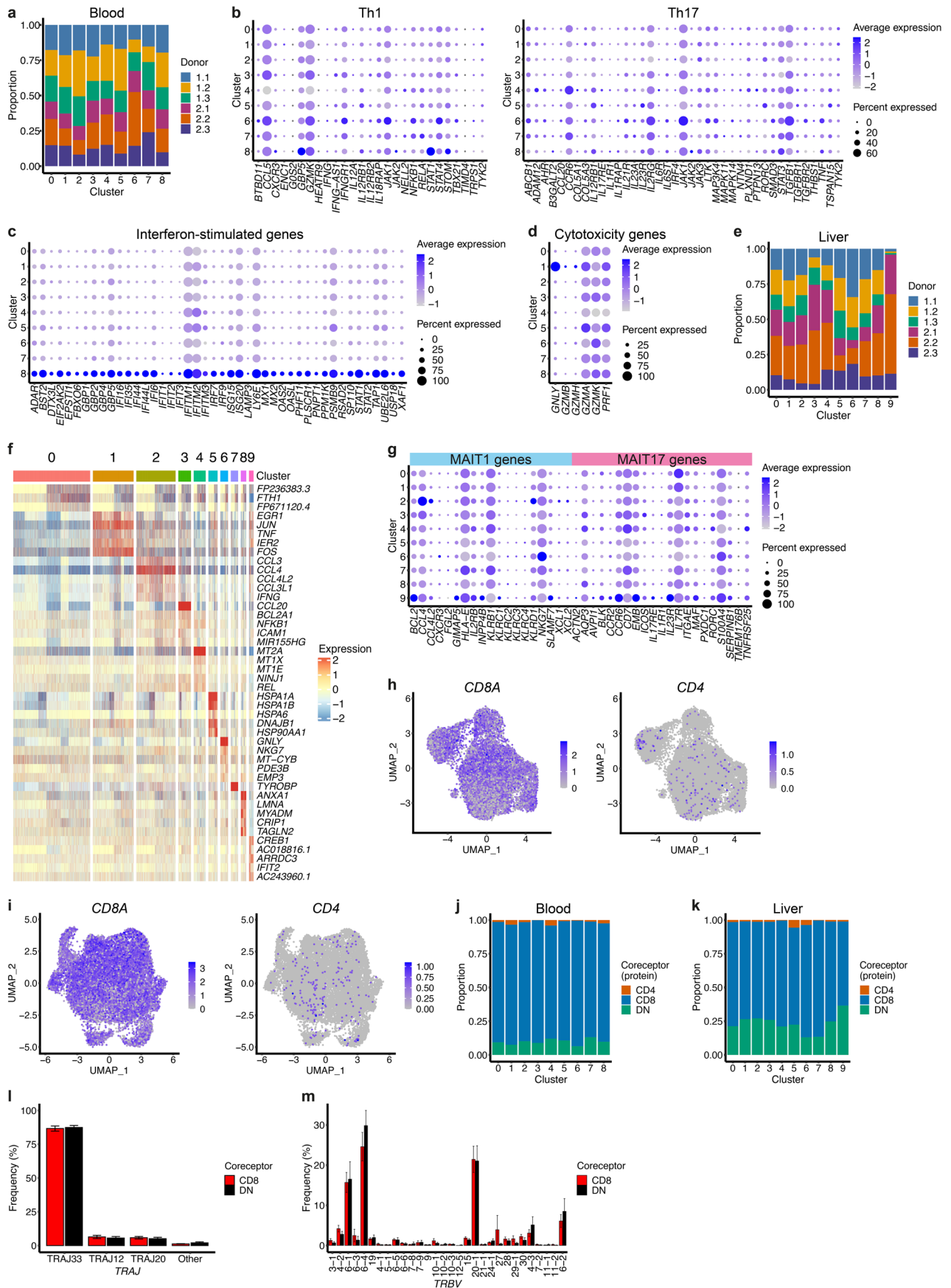


Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | MAIT cells have a restricted TCR $\alpha$  but diverse TCR $\beta$  chain, resulting in private TCR $\alpha\beta$  repertoires.** **a**, Proportion of blood and liver cells expressing *TRAJ33*, *TRAJ12*, *TRAJ20* and other *TRAJ* gene segments. **b**, Distribution of CDR3 $\alpha$  amino acid lengths. **c**, Sequence logos generated from all *TRAJ33*, *TRAJ12* or *TRAJ20* CDR3 $\alpha$  amino acid sequences of length 12 ( $n = 26,529$ ,  $1,852$  and  $1,451$  sequences, respectively). **d,e**, Frequency of N-nucleotides (**d**) and P-nucleotides (**e**) in *TRAJ33*, *TRAJ12* and *TRAJ20* TCRs from 12 donors. **f**, Proportion of blood and liver MAIT cells expressing different *TRBV* gene segments. Plot includes *TRBV* gene segments with a frequency  $>1\%$  in at least one sample. **g**, Distribution of CDR3 $\beta$  amino acid lengths. **h**, Sequence logo generated from all MAIT cell CDR3 $\beta$  amino acid sequences of length 15 ( $n = 9,300$  sequences). **i,j**, TCR chain pairing at the population level. Number of unique TCR $\alpha$  chains paired with any given TCR $\beta$  chain in blood and liver MAIT

(**i**;  $n = 11$  blood, 7 liver samples) or  $T_{\text{mem}}$  (**j**;  $n = 10$  blood, 5 liver samples) cells. **k**, Pearson's correlation between the average TCR capture rate (percentage of cells with a paired *TRAV1-2* TCR) for a donor ( $n = 6$ ) and percentage MAIT cell TCR repertoire sharing between matched blood and liver. **l**, Venn diagram showing the number of TCR $\alpha\beta$  clonotypes shared between the five  $T_{\text{mem}}$  cell donors. **m,n**, Number of  $T_{\text{mem}}$  cells from each donor belonging to TCR $\alpha$  clonotypes (**m**) or TCR $\beta$  clonotypes (**n**) found in 1, 2, 3, 4, or 5 (all) donors. **a-c** and **f-i** show data from  $n = 18$  samples (11 blood, 7 liver), 12 donors. Data in **a, b, f, g** are presented as mean  $\pm$  s.e.m. In **d, e, i, j**, boxes span the 25th–75th percentiles, the midline denotes the median and whiskers extend to  $\pm 1.5 \times$  IQR. Points in **i** and **j** indicate outliers. Two-sided Wilcoxon rank-sum test (**a, f, i, j**) and two-sided Wilcoxon signed-rank test (**d, e**) for all pairwise comparisons. Benjamini-Hochberg adjusted  $P$  values are shown (nonsignificant results omitted). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

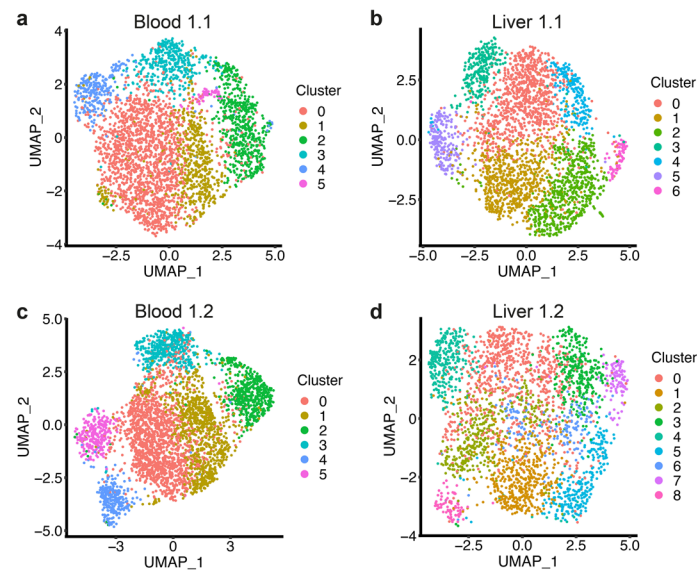




Extended Data Fig. 3 | See next page for caption.

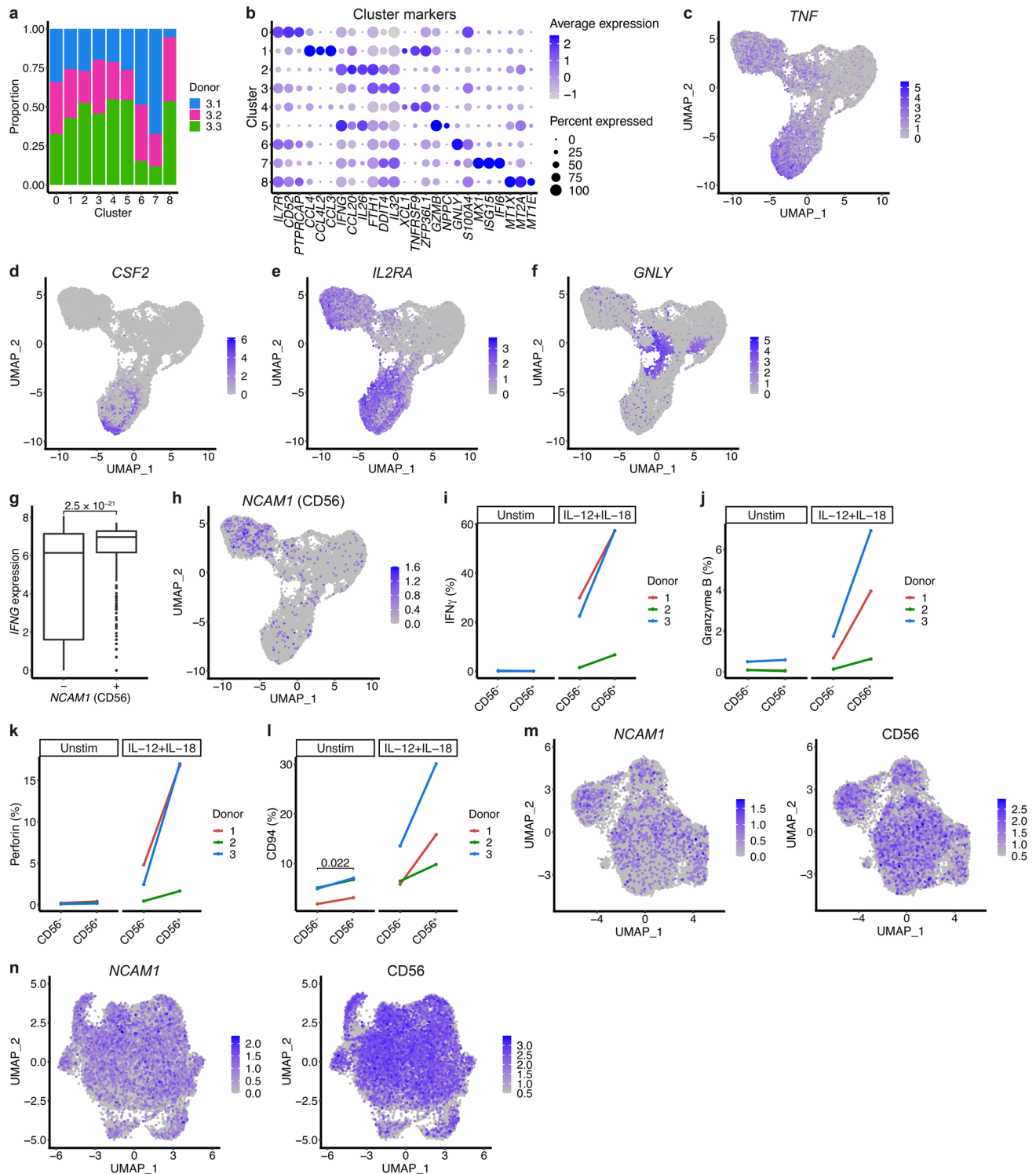
**Extended Data Fig. 3 | MAIT cells within the blood and liver show minimal transcriptional heterogeneity.** **a**, Proportion of cells in each blood MAIT cell cluster from each donor ( $n = 6$ ). **b-d**, Expression of Th1 and Th17 (**b**), interferon-stimulated (**c**) and cytotoxicity (**d**) genes in blood MAIT cell clusters. Dot color indicates the level of gene expression and dot size indicates the percentage of cells expressing the gene. **e**, Proportion of cells in each liver MAIT cell cluster from each donor ( $n = 6$ ). **f**, Heatmap showing row-scaled log-transformed normalized expression of the top five or all (if  $< 5$ ) marker genes for each liver MAIT cell cluster. **g**, Expression of MAIT1 and MAIT17 genes in liver MAIT cell clusters. Dot color indicates the level of gene expression and dot size indicates

the percentage of cells expressing the gene. **h,i**, UMAPs of blood (**h**) and liver (**i**) MAIT cells colored by expression of *CD8A* (left) and *CD4* (right).  $CD4^+$  cells positioned in front of  $CD4^-$  cells to allow better visibility. **j,k**, Proportion of  $CD4^+$ ,  $CD8^+$  and DN cells in each blood (**j**) and liver (**k**) cluster. Coreceptor identity defined based on the expression of CD4 and CD8 proteins (Methods). **l,m**, Frequency of  $CD8^+$  and DN MAIT cells expressing the indicated *TRAJ* (**l**) and *TRBV* (**m**) genes.  $n = 12$  donors. *TRBV* genes expressed in  $> 1\%$  of  $CD8^+$  or DN cells from any donor are included. Mean  $\pm$  s.e.m. is shown. Two-sample Wilcoxon signed-rank test for all pairwise comparisons (nonsignificant results omitted).



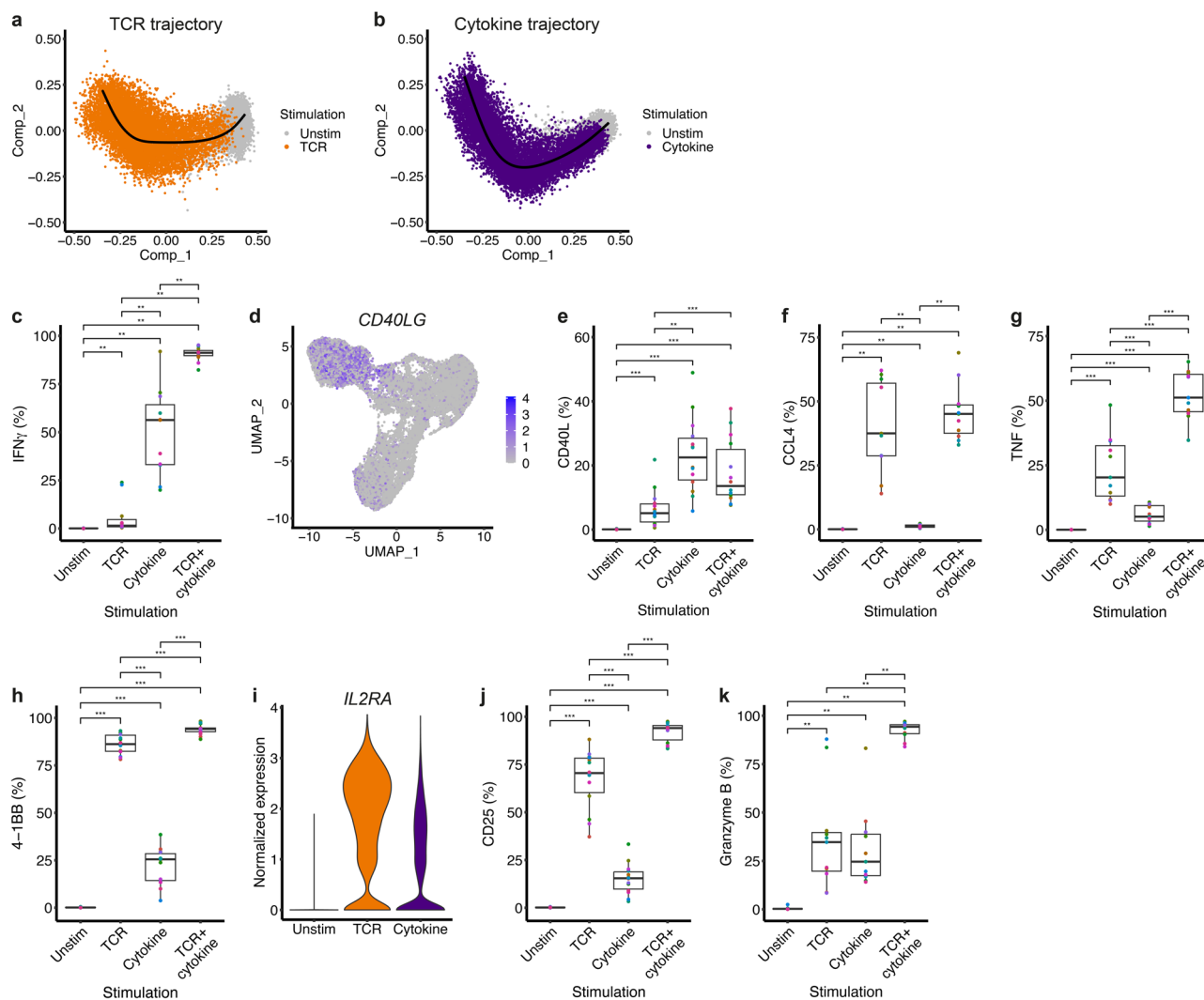
**Extended Data Fig. 4 | Clusters identified in individual blood or liver samples. a-d,** Individual sample UMAPs colored by the identified clusters—blood 1.1 (a), liver 1.1 (b), blood 1.2 (c), liver 1.2 (d).





**Extended Data Fig. 5 | Gene and protein expression by TCR- and cytokine-stimulated MAIT cells.** **a**, Proportion of cells from each donor ( $n = 3$ ) in each cluster (unstimulated, TCR-stimulated and cytokine-stimulated MAIT cells combined). **b**, Expression of the top three marker genes per cluster. Dot color indicates the level of gene expression and dot size indicates the percentage of cells expressing the gene. **c-f**, UMAPs colored by expression of *TNF* (**c**), *CSF2* (**d**), *IL2RA* (**e**) and *GNLY* (**f**). **g**, Expression of *IFNG* in cytokine-stimulated MAIT cells negative ( $n = 11,122$  cells) and positive ( $n = 387$  cells; log-transformed normalized expression  $> 0$ ) for the expression of *NCAM1* (CD56). Boxes span the 25th–75th

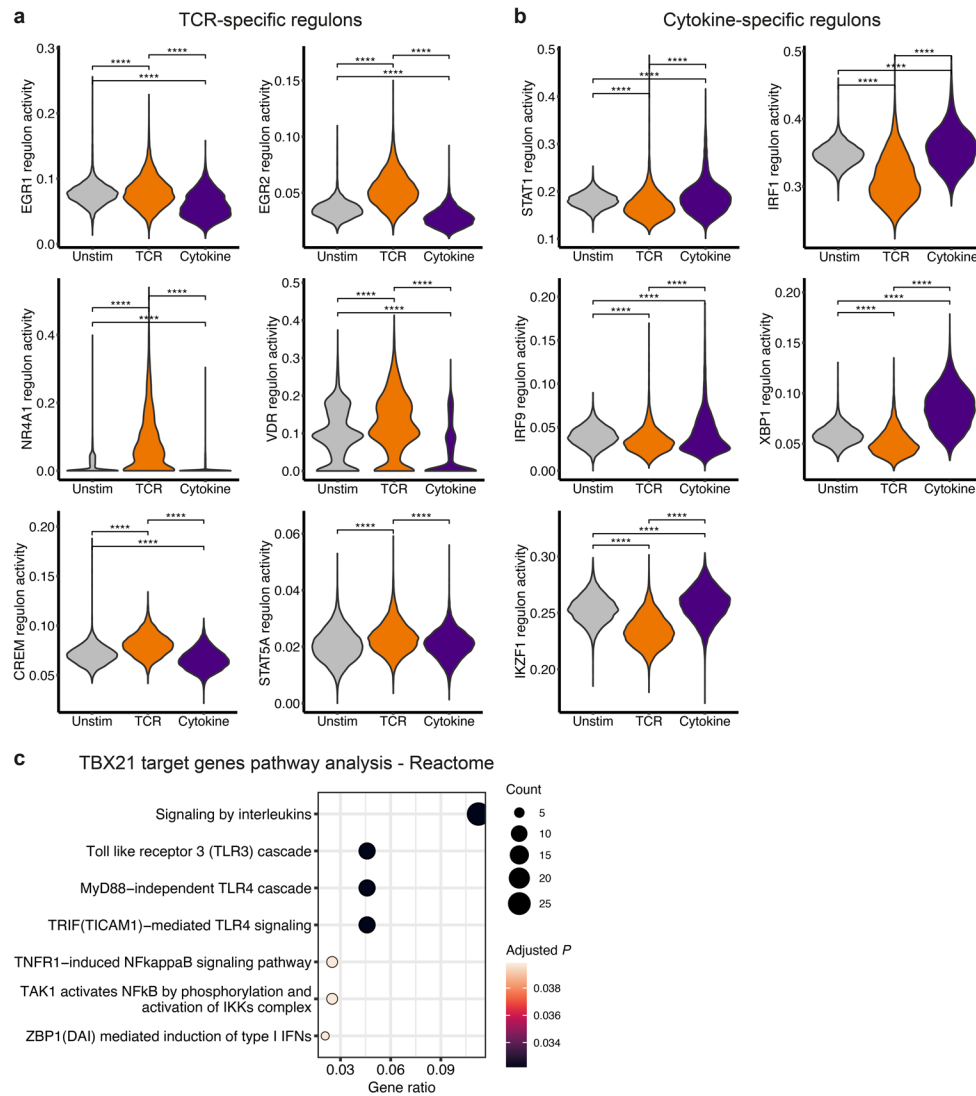
percentiles, the midline denotes the median and whiskers extend to  $\pm 1.5 \times \text{IQR}$ . Points indicate outliers. Two-sample Wilcoxon rank-sum test. **h**, UMAP colored by expression of *NCAM1* (CD56). **i-i**, Percentage of sorted CD56<sup>-</sup> and CD56<sup>+</sup> MAIT cells expressing IFN $\gamma$  (**i**), granzyme B (**j**), perforin (**k**) and CD94 (**l**) when left unstimulated or stimulated with IL-12 + IL-18 for 20 h (as measured by flow cytometry).  $n = 3$  donors. Two-sided paired  $t$ -test between CD56<sup>-</sup> and CD56<sup>+</sup> cells in both conditions (nonsignificant results omitted). **m, n**, CD56 gene (left) and protein (right) expression in blood (**m**) and liver (**n**) MAIT cells. Protein expression measured in Exp 2 only.



**Extended Data Fig. 6 | MAIT cell activation trajectories and validation of stimulus-specific and shared activation markers and cytokines.**

**a, b**, Multidimensional scaling plots of unstimulated and TCR-stimulated MAIT cells (**a**), or unstimulated and cytokine-stimulated MAIT cells (**b**). SCORPIUS TCR (**a**) and cytokine (**b**) trajectories are shown in black.  $n = 3$  donors. **c, e-h, j, k**, Percentage of MAIT cells expressing IFN $\gamma$  (**c**), CD40L (**e**), CCL4 (**f**), TNF (**g**), 4-1BB (**h**), CD25 (**j**) and granzyme B (**k**) when left unstimulated or stimulated with plate-bound MRI/5-OP-RU (TCR), IL-12 + IL-18 (cytokine) or both (TCR+cytokine) for 20 h. Protein expression was measured by flow cytometry on all MAIT cells from

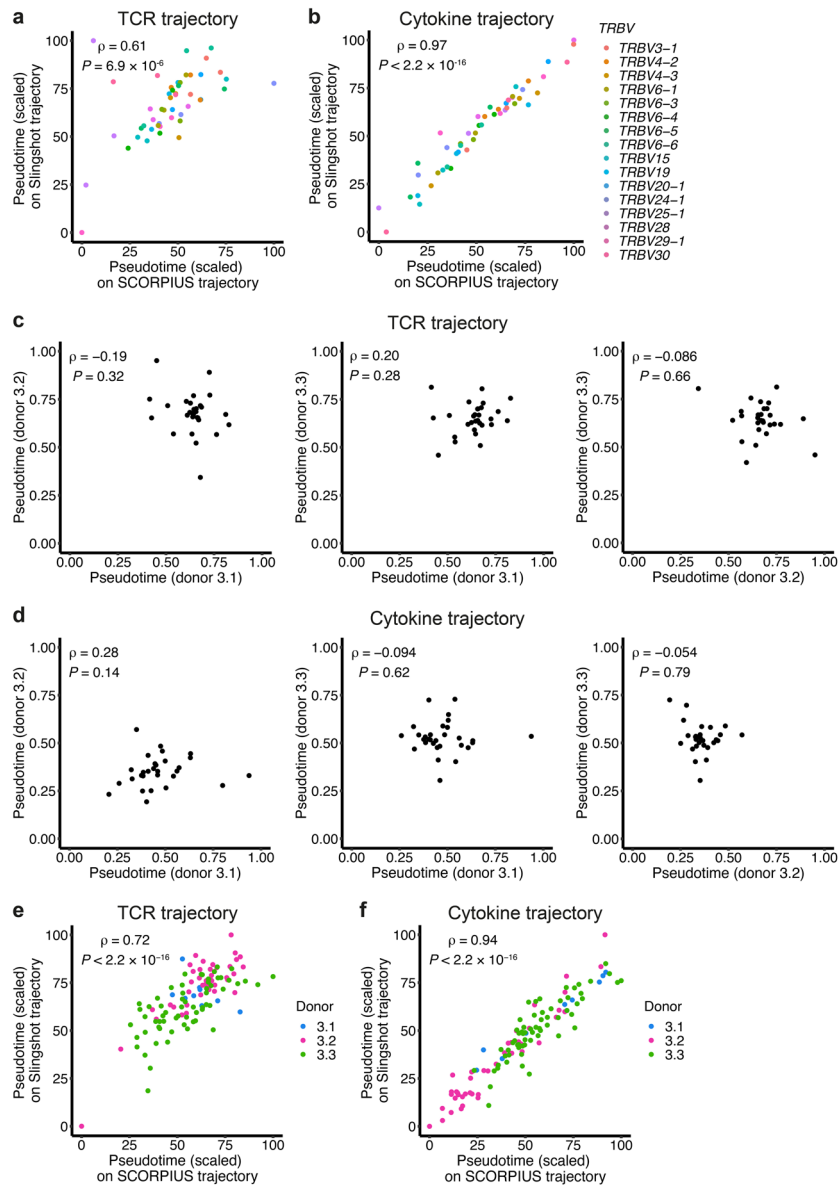
11 donors in two independent experiments (**c, f, g, k**) or CD4<sup>+</sup> MAIT cells from 14 donors in three independent experiments (**e, h, j**). Boxes span the 25th–75th percentiles, the midline denotes the median and whiskers extend to  $\pm 1.5 \times$  IQR. Two-sample Wilcoxon signed-rank test. Benjamini-Hochberg adjusted  $P$  values are shown (nonsignificant results omitted). \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . **d**, UMAP of unstimulated, TCR-stimulated and cytokine-stimulated MAIT cells colored by expression of *CD40LG*. **i**, Violin plot showing expression of *IL2RA* by unstimulated, TCR-stimulated and cytokine-stimulated MAIT cells.



**Extended Data Fig. 7 | Transcriptional regulation of TCR- and cytokine-stimulated MAIT cells exhibits shared and distinct properties. a,b,** Violin plots showing the activity (AUCell scores) of selected TCR-specific (**a**) and cytokine-specific (**b**) transcription factor regulons in unstimulated, TCR-stimulated and cytokine-stimulated MAIT cells.  $n = 3$  donors. Differential

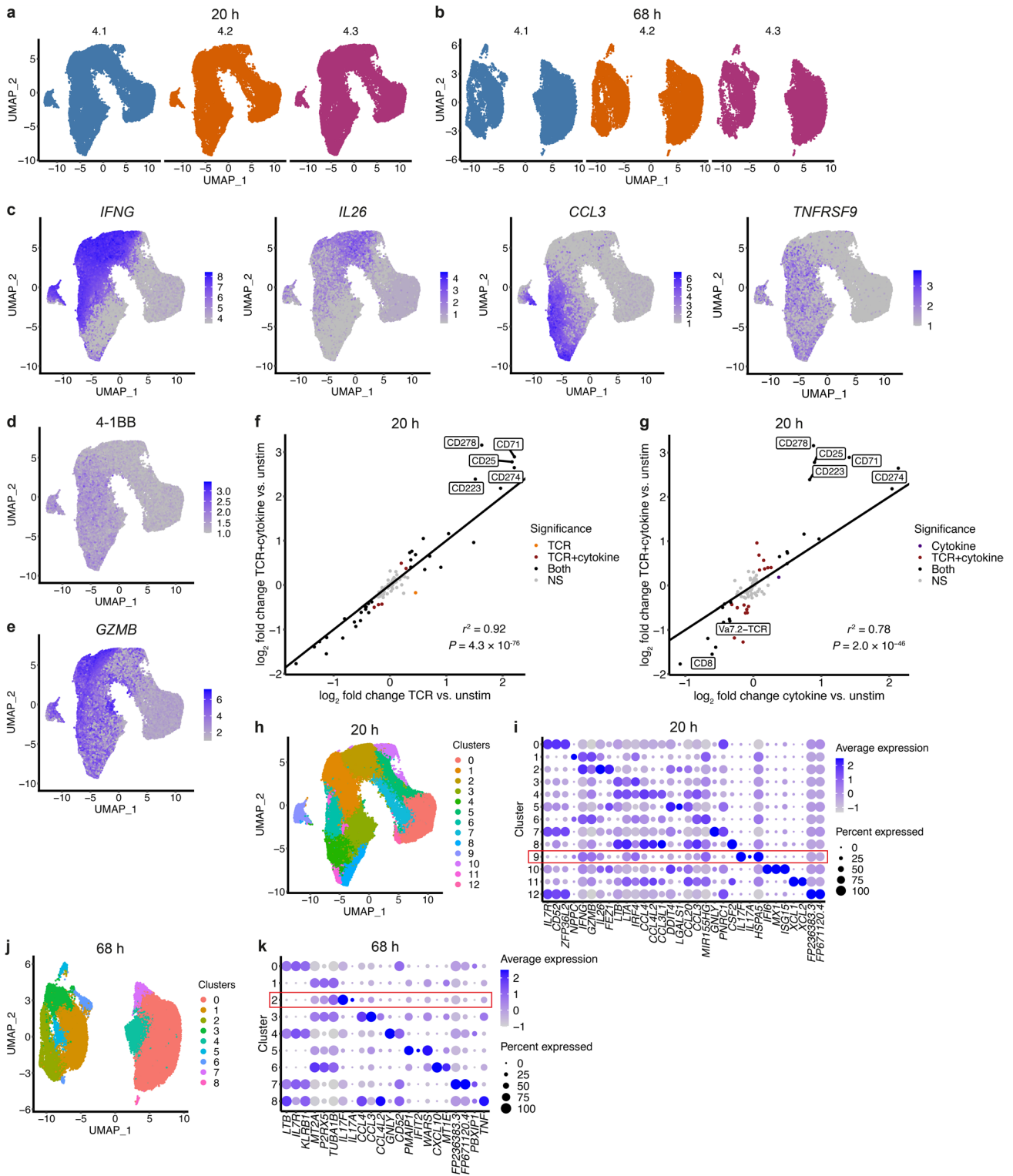
activity analysis was performed for all pairwise comparisons using MAST. Bonferroni adjusted  $P$  values are shown (nonsignificant results omitted). \*\*\*\* $P < 0.0001$ . **c,** Over-representation analysis on predicted T-bet (TBX21) target genes. Significant Reactome pathways and associated Benjamini-Hochberg adjusted  $P$  values are shown.





**Extended Data Fig. 8 | Influence of TCR usage on MAIT cell activation potential.** **a,b**, Spearman's rank correlation between average *TRBV* pseudotimes on SCORPIUS and Slingshot (two trajectory analysis methods) TCR (**a**) and cytokine (**b**) trajectories. Pseudotime values scaled between 0 and 100. **c,d**, Spearman's rank correlation between average *TRBV* pseudotimes for pairs

of donors on SCORPIUS TCR (**c**) and cytokine (**d**) trajectories. **e,f**, Spearman's rank correlation between average clonotype pseudotimes on SCORPIUS and Slingshot TCR (**e**) and cytokine (**f**) trajectories. Pseudotime values scaled between 0 and 100. Plots show stimulated cells only, *TRBV* gene segments with a frequency >1% in any donor (**a-d**) and clonotypes containing  $\geq 20$  cells (**e, f**).

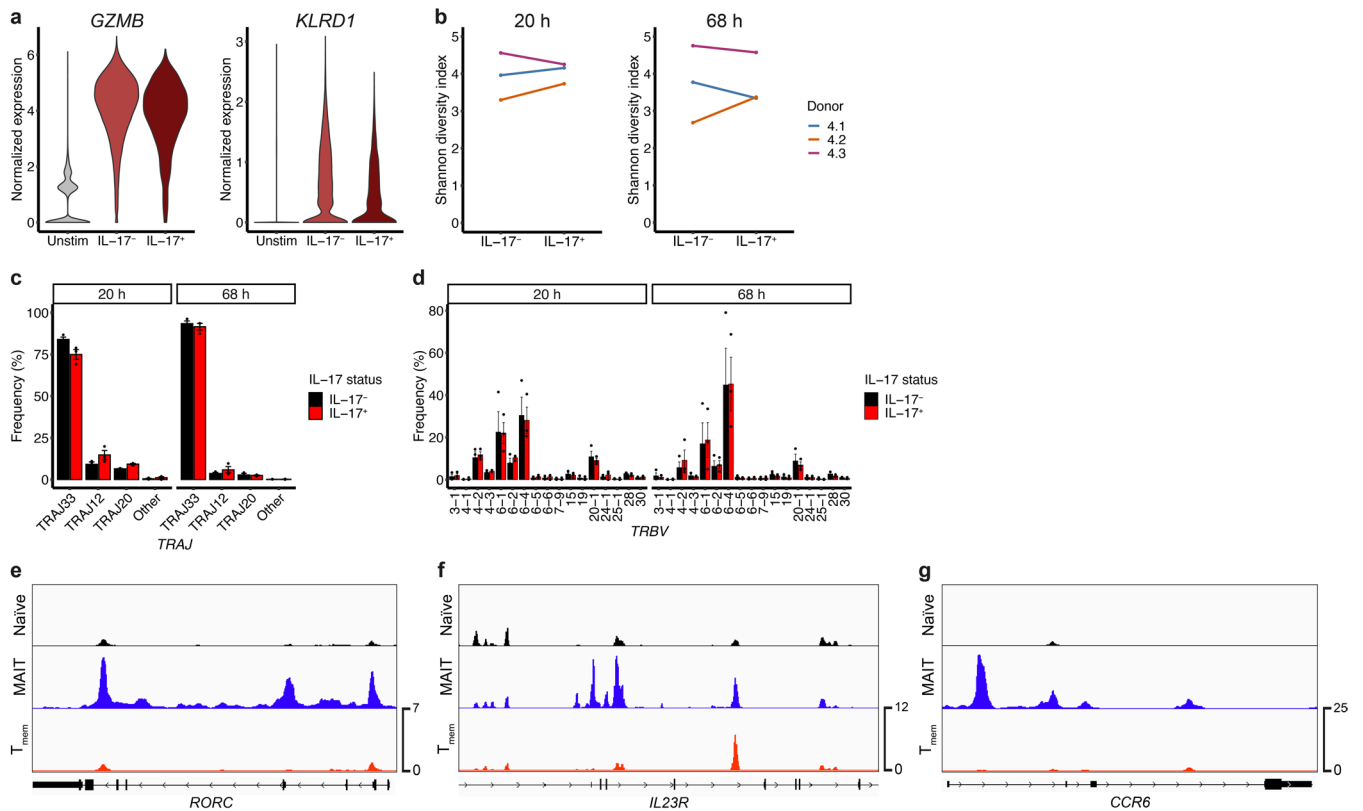


Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | TCR-, cytokine- and TCR+cytokine-stimulated MAIT cells.** **a,b**, UMAPs of 20 h- (**a**) and 68 h- (**b**) stimulated MAIT cells split and colored by donor. **c**, UMAPs of 20 h-stimulated MAIT cells colored by expression of *IFNG*, *IL26*, *CCL3* and *TNFRSF9*. **d**, UMAP of 20 h-stimulated MAIT cells colored by expression of 4-1BB protein. **e**, UMAP of 20 h-stimulated MAIT cells colored by expression of *GZMB*. **f**, Pearson's correlation between the log<sub>2</sub> fold change in protein expression between TCR-stimulated and unstimulated MAIT cells, and TCR+cytokine-stimulated and unstimulated MAIT cells. Labels highlight selected genes that were differentially regulated by the two stimuli. Point colors indicate whether the protein was significantly differentially expressed in response to TCR stimulation only (orange), TCR+cytokine stimulation only (maroon), both (black) or neither (NS; gray). **g**, Pearson's correlation between the log<sub>2</sub> fold change in protein expression between cytokine-stimulated and unstimulated MAIT cells,

and TCR+cytokine-stimulated and unstimulated MAIT cells. Labels highlight selected genes that were differentially regulated by the two stimuli. Point colors indicate whether the protein was significantly differentially expressed in response to cytokine stimulation only (purple), TCR+cytokine stimulation only (maroon), both (black) or neither (NS; gray). **h**, UMAP of 20 h-stimulated MAIT cells colored by the 13 identified clusters. **i**, Dot plot showing the top three marker genes per cluster in **h**. Red box indicates the IL-17-expressing cluster. Dot color indicates the level of gene expression and dot size indicates the percentage of cells expressing the gene. **j**, UMAP of 68 h-stimulated MAIT cells colored by the nine identified clusters. **k**, Dot plot showing the top three marker genes per cluster in **j**. Red box indicates the IL-17-expressing cluster. Dot color indicates the level of gene expression and dot size indicates the percentage of cells expressing the gene.





**Extended Data Fig. 10 | IL-17-expressing MAIT cells and type 17 gene loci.**  
**a**, Violin plots showing *GZMB* (left) and *KLRD1* (CD94; right) expression by unstimulated MAIT cells, and IL-17<sup>-</sup> and IL-17<sup>+</sup> TCR+cytokine-stimulated MAIT cells at 20 h. **b**, Shannon diversity index for TCRαβ clonotypes in IL-17<sup>-</sup> and IL-17<sup>+</sup> MAIT cells from the TCR+cytokine stimulation condition at 20 h (left) and 68 h (right). Cell numbers for each donor were downsampled to ensure equal numbers of IL-17<sup>-</sup> and IL-17<sup>+</sup> cells at a given timepoint. **c,d**, Percentage of IL-17<sup>-</sup> and IL-17<sup>+</sup>

IL-17<sup>+</sup> TCR+cytokine-stimulated MAIT cells expressing different *TRAJ* (**c**) and *TRBV* (**d**) gene segments. *n* = 3 donors. *TRBV* genes expressed in >1% of IL-17<sup>-</sup> or IL-17<sup>+</sup> cells from any donor are included. Mean ± s.e.m. is shown. Two-sample Wilcoxon signed-rank test for all pairwise comparisons (nonsignificant results omitted). **e-g**, Representative ATAC-seq tracks showing *RORC* (**e**), *IL23R* (**f**) and *CCR6* (**g**) gene loci in naïve T (black), MAIT (blue) and T<sub>mem</sub> (red) cells.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Flow cytometry data were collected using BD FACSDiva Software (v8.0.1). Sequencing data were collected using HiSeq Control Software (v2.2.68; Illumina) or NovaSeq Control Software (v1.6.0-1.8.0; Illumina).

Data analysis

Flow cytometry data were analysed using FlowJo (v10.8.1; BD Biosciences).

R version 4.1.1-4.2.0. R packages: AUCell (v1.16.0-1.20.1), biomaRt (v2.54.1), cellhashR (v1.0.3), clusterProfiler (v4.7.1), clustree (v0.4.3-0.5.0), ComplexHeatmap (v2.14.0), DropletQC (v0.0.0.9000), edgeR (v3.24.3), EMT (v1.3), ggplot2 (v3.3.4-3.4.2), ggpubr (v0.4.0), ggrepel (v0.9.1-0.9.3), ggseqlogo (v0.1), gridExtra (v2.3), Harmony (v0.1.1), MAST (v1.18.0-1.24.1), org.Hs.eg.db (v3.13.0-3.16.0), pheatmap (v1.0.12), plotrix (v3.8.2), ReactomePA (v1.36.0), readxl (v1.4.2), rstatix (v0.7.2), scales (v1.1.1-1.2.1), scatter (v1.20.1-1.26.1), sctransform (v0.3.2-0.3.5), SCORPIUS (v1.0.8), Seurat (v4.0.3-4.3.0), Slingshot (v2.0.0), STACAS (v2.0.1), tidyverse (v1.3.1-2.0.0), vegan (v2.6.4), venn (v1.11), VennDiagram (v1.7.3), viridis (v0.6.1-0.6.2).

Python version 3.7.10-3.10.8. Python packages: pandas (v1.2.5-1.5.2), pySCENIC (v0.11.2-0.12.1).

Command-line software: bedtools (v2.26.0), Bowtie 2 (v2.3.4.1), Cell Ranger (v3.0.1-7.0.1; 10x Genomics), deepTools (v3.1.0-3.5.1), FastQC (v0.11.5), featureCounts (v1.6.0), HOMER (v4.8), MACS2 (v2.1.1), Picard (v2.15.0), samtools (v1.6), Trimmomatic (v0.36).

Other software: GraphPad Prism (v9.5.1; GraphPad Software, LLC), IMGJunctionAnalysis (v2.3.0; [https://www.imgt.org/IMGT\\_jcta/analysis](https://www.imgt.org/IMGT_jcta/analysis)), GSEA (v4.3.2), Integrative Genomics Viewer (v2.16.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Sequencing data generated in this study have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO SuperSeries accession number GSE194189. The Bioconductor org.Hs.eg.db annotation package and ENSEMBL\_MART\_ENSEMBL BioMart database are publicly available.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

### Reporting on sex and gender

Blood-liver scRNA-seq experiments: the sex of liver patients is listed in Supplementary Table 1. Blood was additionally obtained from one male and four females. Stimulation scRNA-seq experiments: for Exp 3, blood was obtained from three females; for Exp 4, blood was obtained from two males and one female. Experiments were underpowered for formal sex-based analysis. Sex information is not available for the donors used in flow cytometry validation experiments.

### Reporting on race, ethnicity, or other socially relevant groupings

Race and ethnicity information is not available for the donors in this study.

### Population characteristics

Liver patient information, including age, sex, diagnosis and treatment is detailed in Supplementary Table 1. Age and gender were not accounted for in the analyses. Donor was included as a covariate for differential expression analyses between conditions e.g. cell types, tissues, stimuli.

### Recruitment

Human participants were recruited at the Churchill Hospital, Oxford, UK and the University Hospital Basel, Basel, Switzerland. Patients had no chronic liver disease, active excess alcohol consumption (> 14 g/day), infection, immunosuppression or family history of liver disease. The requirement for written informed consent may introduce selection bias. Patient recruitment was influenced by practical considerations, such as the time of day of the surgery and the amount of liver tissue available.

### Ethics oversight

Samples were obtained with written informed consent through the Oxford Gastrointestinal Illnesses Biobank (REC ref. 16/YH/0247) or under Ethikkommission Nordwest- und Zentralschweiz (EKNZ) numbers EKNZ-2014-362, EKNZ-2016-01188 and EKNZ-2019-02118.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Sample size

Statistical methods were not used to predetermine sample size. scRNA-seq and flow cytometry experiments included  $\geq 3$  donors per group (tissue or stimulation condition). Initial blood-liver (Exp 1) and stimulation (Exp 3) scRNA-seq experiments were followed by validation experiments (Exp 2 and 4) – the findings of these were highly concordant.

### Data exclusions

No samples were excluded from scRNA-seq, ATAC-seq or flow cytometry datasets.

scRNA-seq cell exclusions: cells with low unique molecular identifier counts, low gene counts and/or a high percentage of mitochondrial reads were assumed to be dead/dying cells and were removed. For Exp 1 and 2, cells labelled as empty droplets or damaged cells by DropletQC were removed (damaged cells in Exp 4 also removed). For Exp 2 and 4, only cells called as consensus singlets by hashtag demultiplexing (cellhashR) were retained. Cells with two TCR $\alpha$  and two TCR $\beta$  chains, or more than two TCR $\alpha$  and/or TCR $\beta$  chains, were assumed to be doublets and discarded.

### Replication

Initial blood-liver (Exp 1) and stimulation (Exp 3) scRNA-seq experiments were followed by validation experiments (Exp 2 and 4) – the findings of these were highly concordant. Genes of interest from Exp 1 and 3 were validated at the protein level by CITE-seq (Exp 2 and 4) and flow cytometry. As described in the methods, pySCENIC transcription factor regulon analysis was performed 100 times and results aggregated to define high-confidence regulons.

ATAC-seq was performed once with three donors.

Activation markers and cytokines identified by scRNA-seq were validated by flow cytometry in 14 donors (three independent experiments) and 11 donors (two independent experiments), respectively. Stimulation of sorted CD56- and CD56+ MAIT cells was performed once with three donors. All attempts at replication were successful.

#### Randomization

For blood-liver scRNA-seq experiments, randomization to groups was not possible since liver patients were compared with healthy donors. For stimulation experiments, cells from each donor were present in all conditions.

Blood-liver scRNA-seq experiments (Exp 1 and 2) were each performed in two batches - donors were randomly distributed across batches. Stimulation scRNA-seq experiments (Exp 3 and 4) were each performed as a single batch. Donor was included as a covariate in differential expression analyses.

#### Blinding

Experiments and analyses were not performed blinded as the same investigator performed sample collection, sample processing, data generation and data analysis.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

### Antibodies used

Specificity; fluorophore; clone; manufacturer; catalogue number; RRID

Anti-human CCL4; APC; REA511; Miltenyi Biotec; 130-129-199; AB\_2922004  
 Anti-human CCR7; PE-CF594; 150503; BD Biosciences; 562381; AB\_11153301  
 Anti-human CD3e; APC-Cy7; HIT3A; BioLegend; 300318; AB\_314054  
 Anti-human CD3e; APC-Cy7; UCHT1; BioLegend; 300426; AB\_830755  
 Anti-human CD3e; BV650; UCHT1; BioLegend; 300468; AB\_2629574  
 Anti-human CD3e; BV650; UCHT1; BioLegend; 300464; AB\_2566036  
 Anti-human CD3e; FITC; UCHT1; BioLegend; 300406; AB\_314060  
 Anti-human CD3e; PerCP-Cy5.5; UCHT1; BioLegend; 300430; AB\_893299  
 Anti-human CD4; BV650; OKT4; BioLegend; 317436; AB\_2563050  
 Anti-human CD4; BV711; OKT4; BioLegend; 317440; AB\_2562912  
 Anti-human CD4; BV785; OKT4; BioLegend; 317442; AB\_2563242  
 Anti-human CD8α; AF700; SK1; BioLegend; 344724; AB\_2562790  
 Anti-human CD8α; BV650; SK1; BioLegend; 344730; AB\_2564510  
 Anti-human CD14; APC-Cy7; M5E2; BioLegend; 301820; AB\_493695  
 Anti-human CD14; FITC; M5E2; BioLegend; 301804; AB\_314186  
 Anti-human CD19; APC-Cy7; HIB19; BioLegend; 302218; AB\_314248  
 Anti-human CD19; FITC; HIB19; BioLegend; 302206; AB\_314236  
 Anti-human CD25; PerCP-Cy5.5; BC96; BioLegend; 302626; AB\_2125478  
 Anti-human CD26; APC; BA5b; BioLegend; 30271; AB\_10916120  
 Anti-human CD45; AF700; HI30; BioLegend; 304024; AB\_493761  
 Anti-human CD45RO; BV786; UCHL1; BD Biosciences; 564290; AB\_2738733  
 Anti-human CD56; BV421; HCD56; BioLegend; 318328; AB\_11218798  
 Anti-human CD69; PerCP-Cy5.5; FN50; BioLegend; 310926; AB\_2074956  
 Anti-human CD94; FITC; DX22; BioLegend; 305504; AB\_314534  
 Anti-human CD137 (4-1BB); BV421; 4B4-1; BD Biosciences; 564091; AB\_2722503  
 Anti-human CD154 (CD40L); BV785; 24-31; BioLegend; 310842; AB\_2572187  
 Anti-human CD161; APC; 191B8; Miltenyi Biotec; 130-113-590; AB\_2733346  
 Anti-human CD161; PE; 191B8; Miltenyi Biotec; 130-113-592; AB\_2733625  
 Anti-human CD161; PE-Vio770; 191B8; Miltenyi Biotec; 130-113-594; AB\_2751134  
 Anti-human Granzyme B; AF647; GB11; BioLegend; 515406; AB\_2566333  
 Anti-human Granzyme B; Pacific Blue; GB11; BioLegend; 515408; AB\_2562196  
 Anti-human IFNγ; BV605; 4S.B3; BioLegend; 502536; AB\_2563881  
 Anti-human IFNγ; BV711; 4S.B3; BioLegend; 502540; AB\_2563506



Anti-human Perforin; BV711; dG9; BioLegend; 308130; AB\_2687190  
 Anti-human TCR  $\gamma\delta$ ; BV480; B1; BD Biosciences; 566076; AB\_2739491  
 Anti-human TCR  $\gamma\delta$ ; FITC; B1; BD Biosciences; 559878; AB\_397353  
 Anti-human TCR V $\alpha$ 7.2; PE-Cy7; 3C10; BioLegend; 351712; AB\_2561994  
 Anti-human TCR V $\alpha$ 24-J $\alpha$ 18; FITC; 6B11; BioLegend; 342906; AB\_1731856  
 Anti-human TCR V $\delta$ 2; FITC; B6; BioLegend; 331406; AB\_1089230  
 Anti-human TNF; BV605; MAb11; BioLegend; 502909; AB\_315261

## Validation

All antibodies are commercially available and validation statements can be found on the manufacturers' websites using the catalogue number or in the Antibody Registry database (<https://antibodyregistry.org>) using the RRID. Antibodies were titrated to achieve optimal separation between negative and positive populations.

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

#### Sample preparation

Liver tissue collection and processing (Exp 1 and 2):

Liver tissue (n=7) and matched blood (n=6) were obtained from patients undergoing liver resection at the Churchill Hospital, Oxford, UK and the University Hospital Basel, Basel, Switzerland (Supplementary Table 1). Patients had no chronic liver disease, active excess alcohol consumption (> 14 g/day), infection, immunosuppression or family history of liver disease.

Disease-free liver tissue was collected from the resection margin, cut into small pieces with a scalpel, and ground through a 70  $\mu$ m cell strainer. Cells were washed with R10 (RPMI-1640 [Sigma-Aldrich], 10% FBS [Sigma-Aldrich], 1% penicillin-streptomycin [Thermo Fisher Scientific]; 931g, 10 min, 4  $^{\circ}$ C) and mononuclear cells isolated by density gradient centrifugation on a discontinuous 35%/70% Percoll (GE Healthcare) gradient (931g, 20 min, 21  $^{\circ}$ C, no brake). Mononuclear cells were collected from the interface and washed with R10 (596g, 10 min, 4  $^{\circ}$ C). Residual red blood cells were lysed with ACK for 3-5 min. Cells were washed twice (596g, 10 min, 4  $^{\circ}$ C) and cryopreserved (90% FBS, 10% DMSO [Sigma-Aldrich]) in liquid nitrogen.

Ethics statement:

Samples were obtained with written informed consent through the Oxford Gastrointestinal Illnesses Biobank (REC ref. 16/YH/0247) or under Ethikkommission Nordwest- und Zentralschweiz (EKNZ) numbers EKNZ-2014-362, EKNZ-2016-01188 and EKNZ-2019-02118.

Peripheral blood mononuclear cell (PBMC) isolation:

PBMCs were isolated from fresh whole blood by density gradient centrifugation (Lymphoprep, Axis-Shield) at 931g for 30 min with no brake. Cells were cryopreserved in liquid nitrogen and thawed in complete medium (R10, 1X nonessential amino acids [Thermo Fisher Scientific], 1 mM sodium pyruvate [Thermo Fisher Scientific], 10 mM HEPES [pH 7.0-7.5; Thermo Fisher Scientific], 50  $\mu$ M  $\beta$ -mercaptoethanol [Thermo Fisher Scientific]) on the day of use.

Stimulation of isolated CD8<sup>+</sup>/CD3<sup>+</sup> T cells for scRNA-seq and scTCR-seq (Exp 3 and 4) or activation marker/cytokine validation:

Pierce streptavidin-coated high-capacity flat-bottom 96-well plates (Thermo Fisher Scientific) were coated with 50  $\mu$ l biotinylated MR1/5-OP-RU monomer (NIH Tetramer Core Facility) at 10  $\mu$ g/ml in PBS (Sigma-Aldrich) overnight at 4  $^{\circ}$ C. Cryopreserved PBMCs were thawed in complete medium. CD8<sup>+</sup> T cells were isolated using CD8 MicroBeads (Exp 3; Miltenyi Biotec) and CD3<sup>+</sup> T cells using the REAlease CD3 MicroBead Kit (Exp 4 and validation experiments; Miltenyi Biotec) following the manufacturer's instructions. Isolated CD8<sup>+</sup>/CD3<sup>+</sup> T cells were washed in complete medium and resuspended at  $1 \times 10^7$  cells/ml. One million (20 h stimulation) or 500,000 (68 h stimulation) cells were added per well to the appropriate 96-well plates (MR1/5-OP-RU-coated plate for TCR and TCR+cytokine stimulation, round-bottom plate for unstimulated and cytokine stimulation). IL-12 (50 ng/ml; R&D Systems) and IL-18 (50 ng/ml; R&D Systems) were added for cytokine stimulation;  $\alpha$ CD28 (1  $\mu$ g/ml; clone: CD28.2; BioLegend) for TCR stimulation; IL-12, IL-18, and  $\alpha$ CD28 for TCR+cytokine stimulation; and complete medium for unstimulated cells (final volume 200  $\mu$ l/well). Cells were incubated for 20 h or 68 h at 37  $^{\circ}$ C, 5% CO<sub>2</sub>. For intracellular cytokine staining, brefeldin A (BioLegend) and monensin (BioLegend) were added for the final 4 h.

Tetramer staining (Exp 1 and 2):

Biotinylated human MR1/5-OP-RU and MR1/6-FP monomers were provided by the NIH Tetramer Core Facility. Tetramers were generated using streptavidin-PE (high concentration) or streptavidin-BV421 (both BioLegend) following the NIH Tetramer Core Facility protocol. Tetramer staining was performed for 40 min at 21  $^{\circ}$ C in FACS buffer (PBS, 0.5% BSA [Sigma-Aldrich], 1 mM EDTA [Sigma-Aldrich]).

Surface staining and cell sorting for scRNA-seq and scTCR-seq (Exp 1-4):

TotalSeq-C hashtag antibodies (BioLegend) were used in Exp 2 and 4. Hashtag antibody dilutions were prepared according to the manufacturer's instructions. Namely, antibody vials were centrifuged at 10,000g, 30 s, 4  $^{\circ}$ C, before antibody dilution in

FACS buffer. Diluted hashtags were centrifuged at 14,000g, 10 min, 4 °C. Cells were incubated in Human TruStain FcX (BioLegend) for 10 min at 4 °C before the addition of diluted hashtag antibodies (0.2 µg/well) for 10 min at 4 °C. Surface fluorochrome-conjugated antibodies were added without washing off the hashtag antibodies. Surface staining was performed in Brilliant Stain Buffer Plus (BD Biosciences) for 30 min at 4 °C. Cells were washed twice in PBS with 0.5% BSA, resuspended in presort buffer (PBS, 1% BSA, 25 mM HEPES) containing 3-5 nM SYTOX Green Nucleic Acid Stain (Thermo Fisher Scientific) and incubated for 20 min at 4 °C. Cells were sorted on a BD FACSAria III with an 85 µm nozzle. Sorted cells were collected in RPMI-1640, 10% FBS, 25 mM HEPES, or HBSS (Thermo Fisher Scientific), 50% FBS, 25 mM HEPES. Sort purity was > 99%. For Exp 2 and 4, sorted cells were stained with the TotalSeq-C Human Universal Cocktail V1.0 (BioLegend) according to the manufacturer's instructions. Staining reagents are listed in Supplementary Table 12.

#### Stimulation of CD56- and CD56+ MAIT cells:

CD3+ T cells were isolated using the REAlease CD3 MicroBead Kit following the manufacturer's instructions. Surface antibody and live/dead (SYTOX Green Nuclear Acid Stain) staining were performed as above, then CD56- and CD56+ MAIT cells (Vα7.2+CD161hi) sorted on a BD FACSAria III with an 85 µm nozzle. Sorted cells were collected in HBSS, 50% FBS, 25 mM HEPES, then centrifuged at 400g, 5 min, 21 °C and incubated overnight at 37 °C, 5% CO<sub>2</sub>. Rested cells were washed in complete medium, plated in a 96-well round-bottom plate and stimulated with IL-12 (50 ng/ml) and IL-18 (50 ng/ml) at 37 °C, 5% CO<sub>2</sub> for 20 h, with the addition of brefeldin A and monensin for the final 4 h.

#### Surface marker and intracellular cytokine staining for flow cytometry:

Surface staining was performed in Brilliant Stain Buffer Plus for 30 min at 4 °C. Stained cells were washed twice in FACS buffer. For intracellular cytokine staining, cells were fixed in Cytofix/Cytoperm (BD Biosciences) for 20 min at 4 °C, then washed twice in 1X Perm/Wash (BD Biosciences). Intracellular staining was performed in 1X Perm/Wash for 30 min at 4 °C. Cells were acquired on a BD LSR II flow cytometer with BD FACSDiva Software (v8.0.1). Staining reagents are listed in Supplementary Table 12.

#### Instrument

Cells sorting was performed on a BD FACSAria III with an 85 µm nozzle. Cells were analysed on a BD LSR II.

#### Software

FACSDiva (v8.0.1; BD Biosciences) for data collection, FlowJo (v10.8.1; BD Biosciences) for data analysis.

#### Cell population abundance

Postsort purity, determined for representative samples, was > 99%.

#### Gating strategy

See Supplementary Figures 1-3.

Exp 1 and 2 - FSC-A vs. SSC-A (lymphocytes), CD45 vs. SSC-A (hematopoietic), FSC-H vs. FSC-A (singlets), CD3ε vs. Dump (live CD3+ T cells; Dump = CD14, CD19, TCR γδ, TCR Vα24-Jα18, TCR Vδ2, SYTOX Green Nucleic Acid Stain). MAIT cells are MR1/5-OP-RU+ and Tmem cells are MR1/5-OP-RU-CCR7-.

Exp 3 - FSC-A vs. SSC-A (lymphocytes), CD3ε vs. Dump (live CD3+ T cells; Dump = CD14, CD19, TCR γδ, TCR Vα24-Jα18, TCR Vδ2, SYTOX Green Nucleic Acid Stain), FSC-H vs. FSC-A (singlets), CD4 vs. CD8α (CD8+ T cells). MAIT cells are CD26+CD161hiVα7.2+.

Exp 4 - FSC-A vs. SSC-A (lymphocytes), CD3ε vs. Dump (live CD3+ T cells; Dump = CD14, CD19, TCR γδ, TCR Vα24-Jα18, TCR Vδ2, SYTOX Green Nucleic Acid Stain), FSC-H vs. FSC-A (singlets). MAIT cells are CD26+CD161hiVα7.2+.

#### CD56- and CD56+ MAIT cell sorting and poststimulation analysis:

Sorting - FSC-A vs. SSC-A (lymphocytes), CD3ε vs. Dump (live CD3+ T cells; Dump = CD14, CD19, TCR γδ, TCR Vα24-Jα18, TCR Vδ2, SYTOX Green Nucleic Acid Stain), FSC-H vs. FSC-A (singlets). CD56- and CD56+ MAIT cells are Vα7.2+CD161hiCD56- and Vα7.2+CD161hiCD56+, respectively.

Poststimulation analysis - FSC-H vs. FSC-A (singlets), FSC-A vs. SSC-A (lymphocytes), CD3ε vs. live/dead (live CD3+ T cells). CD56- and CD56+ MAIT cells are Vα7.2+CD161hiCD56- and Vα7.2+CD161hiCD56+, respectively.

Stimulation and activation marker staining - FSC-H vs. FSC-A (singlets), FSC-A vs. SSC-A (lymphocytes), CD3ε vs. Dump (live CD3+ T cells; Dump = CD14, CD19, TCR γδ, TCR Vα24-Jα18, TCR Vδ2, SYTOX Green Nucleic Acid Stain). MAIT cells are CD26+CD161hiVα7.2+.

Stimulation and cytokine staining - FSC-H vs. FSC-A (singlets), FSC-A vs. SSC-A (lymphocytes), CD3ε vs. live/dead (live CD3+ T cells). MAIT cells are Vα7.2+CD161hi.

ATAC-seq - FSC-A vs. SSC-A (lymphocytes), FSC-A vs. Dump (non-monocyte, non-B, live; Dump = CD14, CD19, live/dead), FSC-H vs. FSC-A (singlets), CD3ε vs. TCR γδ (non-γδ T cells), CD4 vs. CD8α (CD8+ T cells). Naive T cells are CD45RO-CCR7+, MAIT cells are CCR7-MR1/5-OP-RU+ and Tmem cells are CCR7-MR1/5-OP-RU-.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.