1 **Title: Single-cell profiling identifies a spectrum of human unconventional**
2 **intraepithelial T lineage cells**

3

4 **Authors:** Lore Billiet[1†], Laurenz De Cock[2,3†], Guillem Sanchez Sanchez[4,5,6], Rupert L.
5 Mayer[2,7,8], Glenn Goetgeluk[1,3], Stijn De Munter[1,3], Melissa Pille[1], Joline Ingels[1,3], Hanne
6 Jansen[1], Karin Weening[1], Eva Pascal[1,3], Killian Raes[1], Sarah Bonte[3,9,11], Tessa Kerre[1,3,9], Niels
7 Vandamme[10,11], Ruth Seurinck[11], Jana Roels[10,11], Marieke Lavaert[1], Filip Van Nieuwerburgh[3,12],
8 Georges Leclercq[1,3], Tom Taghon[1,3], Francis Impens[2,7,8], Björn Menten[2], David Vermijlen[4,5,6]
9 and Bart Vandekerckhhove[1,3*]

10 **Affiliations:**

11 [1]Department of Diagnostic Sciences, Ghent University, Ghent, Belgium.

12 [2]Department of Biomolecular Medicine, Ghent University, Ghent, Belgium.

13 [3]Cancer Research Institute Ghent (CRIG), Ghent, Belgium.

14 [4]Department of Pharmacotherapy and Pharmaceutics, Université Libre de Bruxelles (ULB),
15 Brussels, Belgium.

16 [5]Institute for Medical Immunology, Université Libre de Bruxelles (ULB), Brussels, Belgium.

17 [6]ULB Center for Research in Immunology (U-CRI), Université Libre de Bruxelles (ULB),
18 Brussels, Belgium.

19 [7]VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium.

20 [8]VIB Proteomics Core, VIB, Ghent, Belgium.

21 [9]Department of Internal Medicine and Pediatrics, Ghent University, Ghent, Belgium.

22 [10]VIB Single Cell Core, VIB, Ghent, Belgium.

23 [11]Data Mining and Modeling for Biomedicine, VIB-UGent Center for Inflammation Research,
24 Ghent, Belgium

25 [12]Department of Pharmaceutics, Ghent University, Ghent, Belgium

26 [†]These authors contributed equally to this work

27 *Corresponding author. email: bart.vandekerckhove@ugent.be

28 **Summary:** Billiet et al. identify the postthymic progeny of the intraepithelial lymphocyte
29 precursors in human based on shared characteristics of the T cell receptor repertoire and the
30 transcriptome. This lineage represents a well-defined but heterogeneous, unconventional TCRαβ[+]
31 lineage mostly confined within the CD8 single positive T cells.

32 **Abstract**: In the human thymus, a CD10[+] PD-1[+] TCRαβ[+] differentiation pathway diverges from
33 the conventional single positive T cell lineages at the early double positive stage. These cells are
34 phenotypically and functionally similar to murine unconventional intraepithelial lymphocyte
35 (uIEL) precursors. Here, the progeny of the human uIEL lineage was identified in antigen-
36 inexperienced blood. The uIELs in thymus and peripheral blood share a transcriptomic profile,
37 characterized by hallmark transcription factors (i.e. *ZNF683* and *IKZF2*), and polyclonal TCR
38 repertoire with autoreactive features, exhibiting a bias towards early TCR alpha chain

39 rearrangements. Single-cell RNA sequencing confirmed a common developmental trajectory
40 between the thymic and peripheral uIELs, and clearly delineated this unconventional lineage in
41 peripheral blood. This population is phenotypically defined as CD3$^+$ TCRαβ$^+$ CD4$^-$ CCR7$^-$ CD26$^-$
42 . It contains CD10$^+$ recent thymic emigrants, Helios$^+$ KIR$^+$ CD8$^+$ Tregs and CD8αα$^+$ T cells. Thus,
43 the uIEL lineage represents a well-defined but heterogeneous, unconventional TCRαβ$^+$ lineage
44 mostly confined in human within the CD8 single positive T cells.

INTRODUCTION

46 After successful rearrangement of the T cell receptor (TCR) alpha and beta locus in the thymus,
47 each precursor T cell expresses a single, unique TCR. These cells are subsequently selected for
48 major histocompatibility complex (MHC) binding affinity and only the cells with a moderate
49 binding affinity further differentiate to mature CD4 or CD8 single positive (SP) conventional T
50 cells (CTCs), a process called positive selection. Conventional T cells leave the thymus as
51 dormant, stem cell-like cells without effector function. These cells will acquire effector function
52 only after encountering their cognate foreign antigen. Concurrently, several minor lineages of so-
53 called unconventional T cell populations are generated by different selection mechanisms.
54 Immature thymocytes may receive a strong TCR signal, resulting in agonist selection. These T
55 cells leave the thymus in an activated state, thereby facilitating rapid expansion in the tissues.
56 Mucosal-associated invariant T (MAIT) and natural killer T (NKT) cells are two unconventional
57 populations that express a semi-invariant TCRαβ reactive to microbial compounds. Besides these
58 two oligoclonal UTC populations, an agonist-selected, polyclonal intraepithelial lymphocyte (IEL)
59 lineage has been described [1,2].

60 CD8αα IELs form a prominent thymus-derived T cell population that guards the gut epithelium,
61 in addition to conventional memory T cells [3]. Their development has been predominantly studied
62 in mice, where it was shown that thymic IEL precursors (IELps) diverge from the conventional T
63 cell lineage upon high-affinity TCR interaction at the CD4$^+$ CD8$^+$ double positive (DP) stage.
64 These precursor T cells are induced to differentiate into CD4$^-$ CD8$^-$ double negative (DN) T cells
65 that express gut homing receptors [4–6]. IELp TCRs were shown to exhibit broad reactivity against
66 multiple MHC haplotypes and across MHC classes and therefore may be more generally
67 deployable compared to CTCs that have a narrow antigen specificity [7]. It has been reported in mice
68 that IELps progress through a PD-1$^+$ stage before upregulating T-bet, a process that is regulated
69 by the transcription factors (TFs) *ID2*, *ID3* and *IKZF2* (Helios)[8]. These IELs may play a role in
70 homeostasis by potentiating innate immune responses and/or by directly killing autologous
71 infected cells. However, their mode of action currently remains enigmatic. It is currently not
72 known whether CD8αα IELs present in other tissues or cancer micro-environments are all derived
73 from the same agonist-selected T cell lineage [9,10].

74 CD8αα$^+$ CD8β$^-$ IELs are not present in the human gut, suggesting that there may not be a human
75 equivalent for the murine CD8αα IEL lineage [11]. However, it is possible that the human IELs have
76 a different phenotype compared to their murine counterparts. Our research previously described
77 the likely IELps in human postnatal thymus (PNT) that is similar to the murine thymic precursors
78 of the CD8αα$^+$ IEL lineage. These human IELps present in the thymus with a CD10$^+$ PD-1$^+$
79 phenotype and, unlike in mice, express both CD8αα and CD8αβ dimers [12]. Similar to murine
80 IELps, these cells can be generated *in vitro* by agonist stimulation of DP thymocytes, have a
81 characteristic TCR repertoire different from CTCs, express TCRs with characteristics of
82 autoreactivity and are activated in the thymus by high affinity ligands as evidenced by the
83 expression of PD-1 and Helios [12–14]. The existence of this lineage in the human thymus as well as

84 the early divergence from the conventional T cell lineage was recently confirmed via single-cell
85 mapping [15,16]. It is currently unknown whether these human IELps leave the thymus and, if so,
86 which phenotype and function these cells have.

87 The aim of this study is to identify the progeny of the IELp lineage in the human periphery. Within
88 peripheral blood, the focus was on cord blood (CB) as CB CTCs have not yet been activated by
89 foreign antigens and therefore, are easily discriminated from activated unconventional T cells.
90 Therefore, the PNT IELp population and their likely progeny in CB were comprehensively
91 analyzed by means of transcriptome, TCR and proteome analyses. A single-cell RNA sequencing
92 (scRNA-seq) analysis was performed to further unravel the progeny in CB, combined with CITE-
93 sequencing to establish a phenotypic definition encompassing the entire heterogeneous population.
94 In summary, a detailed picture of the unconventional IEL (uIEL) lineage cells is provided in CB,
95 which may lead to a better understanding of these cells in autoimmunity as well as immune
96 reactivity to pathogens and will facilitate the study of the uIEL lineage in human.

97 **RESULTS**

98 **The PNT and CB PD-1$^+$ population share a similar transcriptomic and proteomic profile**

99 The phenotype of the IELps in human PNT was defined as CD3$^+$ TCRαβ$^+$ CD4$^-$ CD8α$^+$ CD10$^+$
100 PD-1$^+$ (fig. S1A). The progeny of the PNT IELp population in human CB was tentatively defined
101 as CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CD8α$^+$ PD-1$^+$ (fig. 1A)[12]. CD10 membrane expression, which is
102 prominent on PNT PD-1$^+$ IELps (fig. S1A), was less prominent on the CB PD-1$^+$ population (fig.
103 1B). However, *MME* mRNA (encoding CD10) was significantly upregulated in the CB PD-1$^+$
104 population compared to the conventional CD3$^+$ PD-1$^-$ population (fig. 1C). To examine the
105 relatedness of PNT and CB PD-1$^+$ populations, both were comprehensively analyzed by means of
106 transcriptome and proteome analyses and compared to the CD3$^+$ PD-1$^-$ population. Principle
107 component analysis (PCA) of the sorted CB populations indicated that the PD-1$^+$ populations from
108 the different donors clustered together and, similar to the respective PNT populations, shared more
109 features with the unconventional TCRγδ population than with conventional PD-1$^-$ populations (fig.
110 1D, S1C). Volcano plots comparing the transcriptomes of the PD1$^+$ and PD1$^-$ populations, showed
111 significant upregulation in both CB and PNT PD-1$^+$ populations of the hallmark TFs *ZNF683*
112 (Hobit), *IKZF2* (Helios), *RUNX3*, *ID3* and *TBX21* (T-bet), and downregulation of *RORA* and
113 *FOXP1* that mediates quiescence, and of *SATB1* that is required for positive and negative selection
114 [8,17,18]. Notably, both CB and PNT PD-1$^+$ populations highly expressed the unconventional TCRαβ
115 marker *TRGC2* (T Cell Receptor Gamma Constant 2), a gene progressively silenced in the
116 conventional T cell lineage during passage through the CD4$^+$ CD8$^+$ DP stage in the thymus (fig.
117 1C, S1B)[19]. Mass spectrometry-based proteomics confirmed upregulation of Helios in the PNT
118 PD-1$^+$ population (fig. S1B) and, although not reaching the significance threshold, in CB (fig. 1C).
119 Flow cytometric analysis validated the upregulation of Helios in both PNT and CB populations
120 (fig. 1E, S1D). In search of additional distinctive markers, membrane proteins identified in the
121 transcriptomic and proteomic profile of the PNT and CB PD-1$^+$ populations, were confirmed by
122 flow cytometry: CD3^{low} CD8β^{low} CCR7$^-$ and EVI2B$^+$ (fig. 1E, S1D). As expected, a correlation
123 between the significantly differentially expressed genes and abundance of the corresponding
124 proteins was observed (S1E). When zooming in on the genes or proteins that were differentially
125 expressed between the PD-1$^+$ and PD-1$^-$ populations either in PNT or in CB, a highly significant
126 positive correlation was revealed between the respective PNT and CB populations at both the RNA
127 and protein level (fig. 1F, S1F, table S1). Supporting this correlation between the PNT and CB
128 PD-1$^+$ population, Gene Set Enrichment Analysis (GSEA) confirmed that the significantly

129   upregulated genes in the PNT PD-1$^+$ population were also significantly enriched in the CB PD-1$^+$

130   population (fig. 1G). Based on the similarities in their transcriptomic and proteomic profile, it is

131   hypothesized here that the CB CD3$^{+/low}$ TCR$\gamma\delta^-$ CD4$^-$ CD8α^+ PD-1$^+$ population is the progeny of

132   the PNT PD-1$^+$ IELps.

5
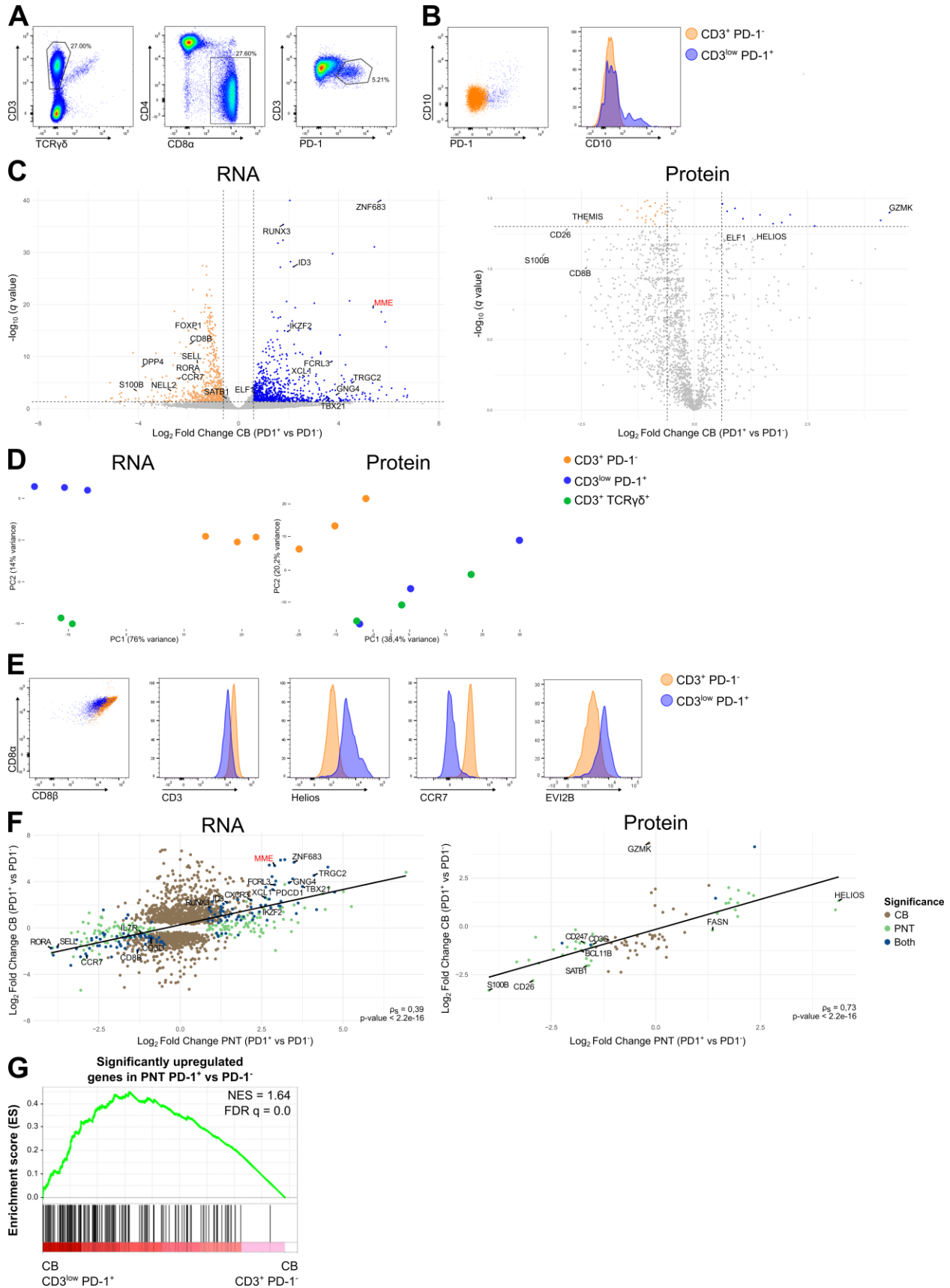
Fig. 1. **RNA and protein expression profile by the CB CD3$^{+/low}$ PD-1$^+$ population.** (**A**) Representative gating strategy for the CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CD8α$^+$ PD-1$^+$ population in human CB. (**B**) CD10 expression on the CD3$^+$ PD-1$^-$ (orange) and CD3$^{+/low}$ PD-1$^+$ (blue) populations in CB, representative of at least three CBs. (**C**) Volcano plots of differentially expressed genes (left) and proteins (right) between the CB CD3$^{+/low}$ PD-1$^+$ and CD3$^+$ PD-1$^-$ populations. Triangles indicate data points outside the y-axis range. Data points with a |Log$_2$ Fold Change| > 0.6 and adjusted P$ < 0.05$ are colored (upregulated in blue, downregulated in orange). (**D**) PCA of the transcriptome (left, donor corrected) and proteome (right) analysis of the sorted populations from three different CB donors. (**E**) Flow cytometric analysis of CD8α, CD8β, CD3, Helios, CCR7 and EVl2B on the CD3$^+$ PD-1$^-$ and CD3$^{+/low}$ PD-1$^+$ populations in CB, representative of at least three CBs. (**F**) Scatterplots and Spearman correlation coefficient comparing the Log$_2$ Fold Change of the significantly differentially expressed genes (left) and proteins (right) of CB (CD3$^{+/low}$ PD-1$^+$ versus CD3$^+$ PD-1$^-$) and PNT (CD10$^+$ PD-1$^+$ versus CD10$^-$ PD-1$^-$). (**G**) GSEA showing the significantly upregulated gene set from PNT CD10$^+$ PD-1$^+$ versus CD10$^-$ PD-1$^-$, on the CB CD3$^{+/low}$ PD-1$^+$ versus CD3$^+$ PD-1$^-$ population. Normalized enrichment score (NES) and false discovery rate q value (FDR q) are shown.

The TCR repertoires of the PNT PD-1$^+$ IELp and CB PD-1$^+$ populations share highly characteristic features

TCRα rearrangements are known to occur in a non-random manner, starting at the J-proximal V segments and the V-proximal J segments. As multiple sequential rearrangements may occur during the DP thymocyte stage, later rearrangements tend to be biased towards J-distal V segment and V-distal J segment usage. As shown in our previous publication and confirmed by others, the TCRα usage of PNT PD-1$^+$ IELp population is biased towards early rearrangements similar to early DP thymocytes, in contrast to late DP and conventional thymocytes [12,13,15]. Additionally, an association between autoreactive populations (T$_{reg}$ and IELp) and CDR3 amino acidic (AA) residue content, including the presence of cysteines within 2 positions of the CDR3 apex (cysteine index) and enrichment of hydrophobic AA doublets at positions 6 and 7 of the CDR3 (hydrophobic index), is reported [13,20]. Similar CDR3 properties are also reported to result in strong TCR-ligand interactions [21–23]. Finally, in contrast to NKT or MAIT cells, the repertoire of PNT IELps was determined to be polyclonal (fig. S2A-B). Thus, to obtain additional evidence for a precursor-progeny relationship between the PNT and CB PD-1$^+$ populations, the TCR repertoire of the CB PD-1$^+$ population was analyzed for these characteristics and compared to the PNT PD-1$^+$ TCR repertoire (fig. S2A-F for PNT).

Analysis of the CDR3α and CDR3β clonotypes revealed that the CB PD-1$^+$ population was polyclonal and the degree of polyclonality was similar to the conventional PD-1$^-$ T cell population (fig. 2A). This was quantified by calculating the D75 values, i.e. the percentage of unique clonotypes required to occupy 75% of the total TCR repertoire, for both populations. The D75 values were about 30% for both CB populations, indicating that the bulk of the repertoire consists of a wide variety of clonotypes. Moreover, no significant mean difference between the two CB populations could be identified, supporting the notion that the unconventional T cell population was equally polyclonal as the conventional T cell population in CB (fig. 2B). The CB PD-1$^+$ population exhibited biased usage of early J-proximal TCR Vα rearrangements and early V-proximal TCR Jα rearrangements, and this bias was similar to that found in the PNT PD-1$^+$ IELps (fig. 2C-E, S2D-E). The cysteine index of the TCRβ chain was significantly higher in the CB PD-

6

179   1^+ population compared to the PD-1⁻ population and the same trend could be observed for the
180   TCRα chain (fig. 2F). The hydrophobic index of the TCRβ chain was likewise significantly higher
181   in the CB PD-1$^+$ population compared to the PD-1⁻ population. However, such a trend could not
182   be established for the TCRα chain (fig. 2G). In line with this, the CDR3β repertoire exhibited
183   higher interaction strength values (strength and volume parameters) compared to the PD-1⁻
184   population counterpart (fig. 2H)[21,22]. In contrast, polarity, a property associated with conventional
185   T cells, was reduced in the CB PD-1$^+$ population (fig. 2H)[24]. Finally, TCR sequencing of the CB
186   populations revealed a significant higher percentage of TRAJ sequences using the T Cell Receptor
187   Delta Variable 1 (*TRDV1*) gene segment (instead of a TRAV gene segment) in the CB PD-1$^+$
188   population (fig. 2I). When subsequently analyzing the PNT populations, an increased *TRDV1*
189   usage was also observed in the PNT PD-1$^+$ IELps compared to their PD-1⁻ counterparts (fig. S2H).
190   Vδ1$^+$ cells expressing a hybrid TRDV1-TRAJ-TRAC TCR chain and co-expressing a TCRβ chain
191   rather than a TCRγ chain have been previously reported in human peripheral blood. This
192   population, termed δ/αβ T cells, recognizes antigens presented by both human leukocyte antigen
193   (HLA) and CD1d [25]. By using an anti-Vδ1 antibody, an enrichment of Vδ1 membrane expression
194   was shown in the CB PD-1$^+$ population (fig. 2J, S2I).

195   To conclude, a series of characteristic features of the TCR repertoire of the PNT PD-1$^+$ IELp can
196   be tracked within the CB PD-1$^+$ population. This strongly suggests that the CB PD-1$^+$ T cell
197   population is the progeny of the PNT PD-1$^+$ T cell population and that biased TCRα chain usage
198   and self-reactive features of both TCR chains are acquired during early thymic agonist selection
199   and preserved after thymic egress.

200

8
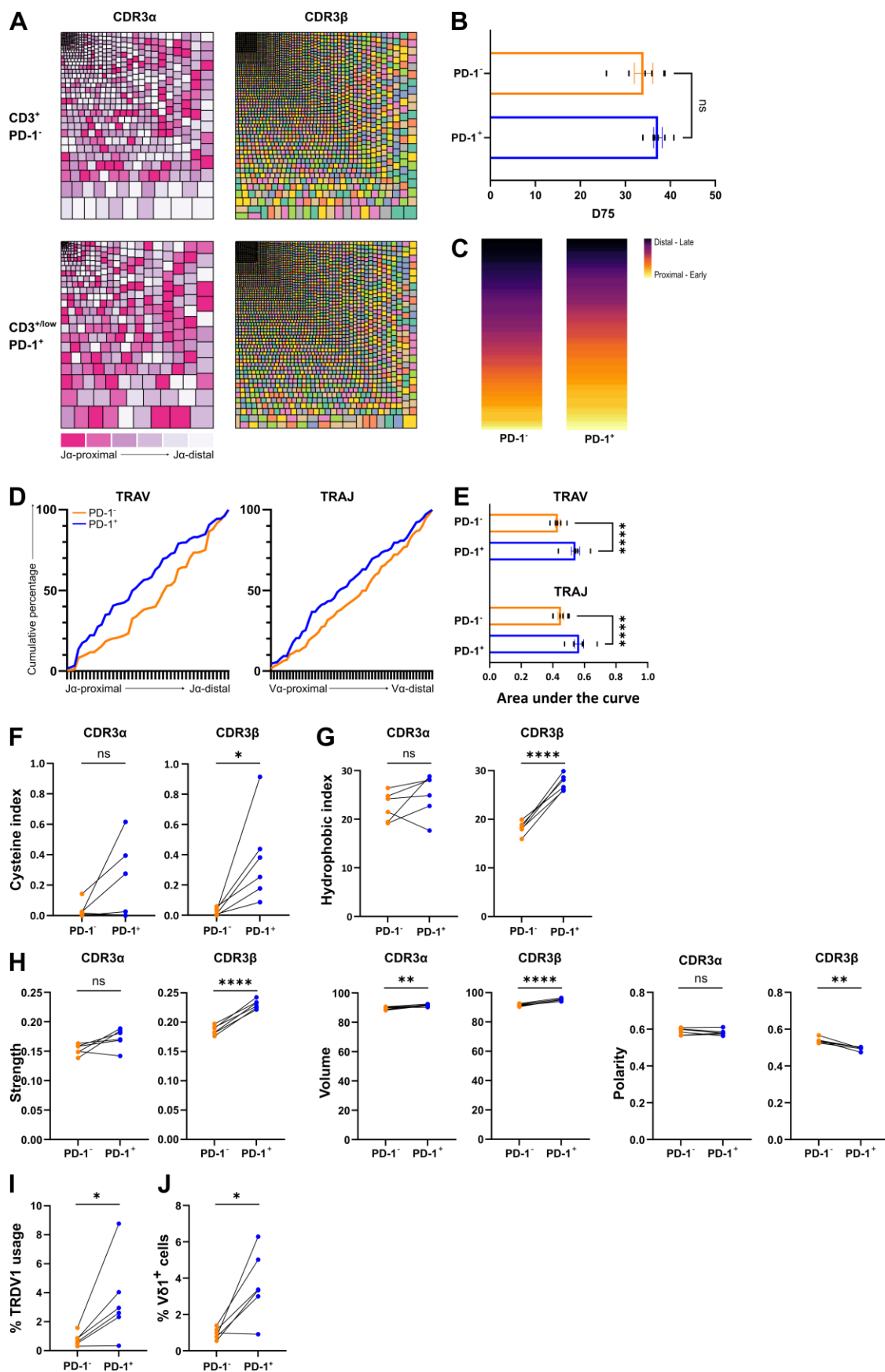
Fig. 2. **Distinctive TCR repertoire of the CB PD-1$^+$ population.** (**A**) Representative tree maps showing CDR3α (left) and CDR3β (right) clonotype usage in relation to repertoire size for the CD3$^+$ PD-1$^-$ (top) and CD3$^{+/low}$ PD-1$^+$ (bottom) populations. Each rectangle represents one CDR3 clonotype and its size corresponds to its relative frequency in the repertoire. Rectangle colors for CDR3α are categorized from J-proximal (pink) to J-distal (white) for the TRAV gene segments and for CDR3β are chosen randomly. (**B**) D75 (percentage of clonotypes required to occupy 75% of the total TCR repertoire) analysis comparing the PD-1$^-$ and PD-1$^+$ population (individual values and mean ± SEM). Mean difference was not significant. (**C**) Representative heatmap illustrating the difference in J-proximal versus J-distal TCR Vα usage between the PD-1$^-$ and PD-1$^+$ populations. (**D**) Cumulative percentage of TRAV (left) or TRAJ (right) gene segment usage by the PD-1$^-$ (orange) and PD-1$^+$ (blue) population in a representative donor. X-axis represents the location in the TRAV or TRAJ locus. (**E**) Area under the curve determined from the cumulative plots from each sample (individual values and mean ± SEM). Šídák's multiple comparisons test was used to assess the statistically significant difference. p-value < 0.0001 (****). (**F**) Cystein index (percentage of unique sequences with cysteine within 2 positions of the CDR3 apex) and (**G**) Hydrophobic index (percentage of unique sequences with self-reactive hydrophobic CDR3 position 6 and 7 doublets) of the CDR3α (left) and CDR3β (right). (**H**). Physicochemical properties (strength, volume and polarity) of the CDR3α (left) and CDR3β (right). (**I**) Percentage of unique sequences containing a *TRDV1* segment. (**J**) Flow cytometric analysis of the percentage of Vδ1$^+$ (A13 clone) cells in both CB populations. (**B, F-I**) Paired t-tests were used to assess statistical significance. Connected values correspond to paired populations of the same biological replicate (n= 6). p-value > 0.05 (ns), p-value < 0.05 (*), p-value < 0.0001 (****).

The CB unconventional T cell population extends beyond the CD3$^{+/low}$ PD-1$^+$ cells

Because murine IELps were shown to downregulate PD-1 expression during differentiation and PD-1$^-$ (T-bet$^+$) IELp populations were described [26,27], the unconventional CD3$^{+/low}$ PD-1$^+$ CB population may not include the complete human uIEL population, but only the more recent thymic emigrants of that population. Therefore, the discriminatory markers identified above were used in a flow cytometric analysis of the entire CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CD8α$^+$ fraction in CB (fig. 3A). In addition to the expected conventional CCR7$^+$ population, the resulting uniform manifold approximation and projection (UMAP) revealed a distinct cluster of CCR7$^-$ cells that contained all PD-1$^+$ cells (fig. 3A). Within this CCR7$^-$ cluster, the majority of the PD-1$^-$ cells were Helios$^+$ and EVI2B$^+$, two markers associated with the IELp population (fig. S1C). Here, it was hypothesized that these Helios$^+$ and EVI2B$^+$ cells may contain PD-1$^-$ uIELs. Consequently, the CB uIEL population was further studied within the CCR7$^-$ EVI2B$^+$ as well as within the CD3$^{+/low}$ PD-1$^+$ population.

To comprehensively study the heterogeneity of the uIELs in CB, single-cell RNA sequencing (scRNA-seq) was performed. CB of two different donors was depleted of the CD4$^+$, CD14$^+$, CD19$^+$ and CD235$^+$ cells (fig. 3B). Of the first donor, the uIELs were sorted as CD8α$^+$ CD3$^{+/low}$ PD-1$^+$ (sort 1) or CD8α$^+$ CCR7$^-$ EVI2B$^+$ (sort 2) within the CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ window. Both fractions were labeled with different hashtags before they were further processed, enabling subsequent assignment of the single cells to their corresponding sorting strategy. For the second donor, the sorting strategy was expanded to include CD4$^-$ CD8α$^-$ DN cells (sort 3 and 4). In sort 3 and 4, the respective conventional populations were also sorted and added in equal portions before further

9

246 analysis (fig. 3C). Sort 4 was combined with cellular indexing of transcriptomes and epitopes
247 (CITE)-seq to capture expression of 277 membrane proteins. Using a droplet-based single-cell
248 platform 3′ gene expression libraries were constructed. Reciprocal PCA (RPCA) was used to
249 integrate the Seurat objects resulting from the separate sorts. This approach resulted in 24 727 cells
250 included in the scRNA-seq analysis after quality control and filtering. Leiden clustering applied to
251 this filtered and integrated Seurat object defined 13 distinct clusters (fig. 3D). Both $CD8\alpha^+$
252 $CD3^{+/low}$ PD-1$^+$ (sort 1) and $CD8\alpha^+$ CCR7$^-$ EVI2B$^+$ (sort 2) consisted mainly of clusters 1-5,
253 suggesting that these clusters represent the unconventional T cells. Focusing on these 5 clusters,
254 cluster 1 is relatively overrepresented in the $CD3^{+/low}$ PD-1$^+$ sorts 1 and 3 and cluster 4 is relatively
255 enriched in the CCR7$^-$ EVI2B$^+$ sorts 2 and 4, highlighting that indeed the two different sorting
256 strategies captured slightly different populations (Fig 3E).
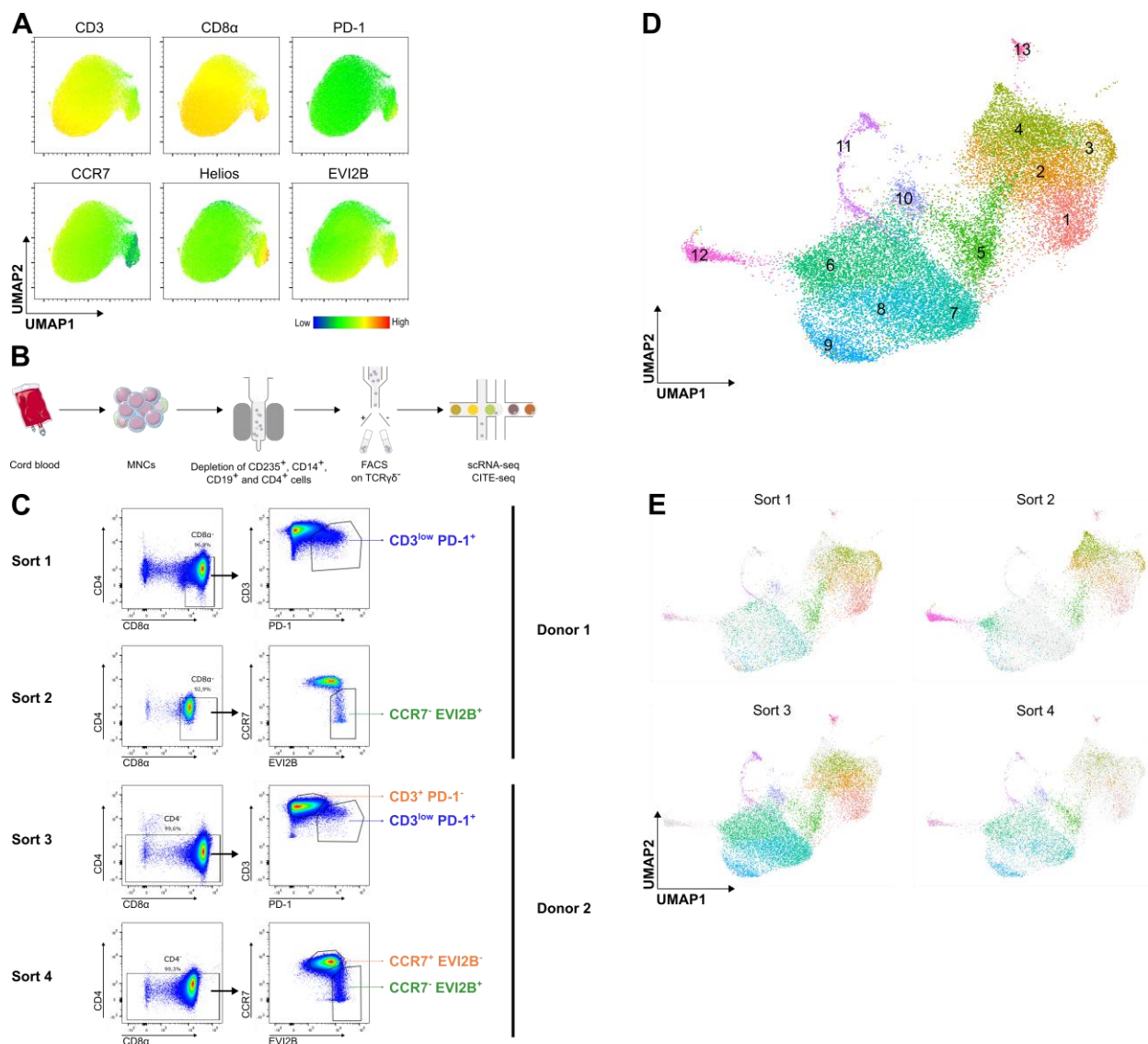
258 **Fig. 3**. **Single-cell RNA sequencing of human CB reveals a heterogeneous unconventional**
259 **population.** (**A**) UAMP analysis of flow cytometry data gated on all $CD3^{+/low}$ TCRγδ$^-$ CD4$^-$
260 $CD8\alpha^+$ cells from CB. (**B**) Schematic workflow of the CB processing. (**C**) Gating strategy

10

261   on the CD3$^{+/low}$ TCRγδ$^-$ cells for the four different sorts, to isolate the different populations
262   of interest. (**D**) UMAP of 24 727 CB single cells, colored by the 13 identified cell clusters.
263   (**E**) UMAP representation of the four different sorts, with the cells from the particular sorts
264   colored according to the cell clusters and the remaining cells in grey.

Defining unconventional T cell clusters using transcriptomics

266 This heterogeneity consisting of 13 different clusters has not been reported before for CB CD8$^+$ T
267 cells. Therefore, the clusters were manually annotated based on prominently upregulated genes
268 (fig. 4A-B; table S2-3). One non-T cell cluster was detected, which was annotated as NK cells
269 based on high expression of NK-associated genes (*GZMB*, *TYROBP*) and absence of membrane
270 CD3 and TCRαβ (fig. 4A, S4A). The NK cluster was assumed to be a contaminant due to lenient
271 gating for CD3. Two minor T cell clusters were annotated: an NKT/MAIT cluster based on high
272 *KLRB1* expression and a cycling T cell cluster based on upregulated effector (i.e. *GZMA*, *GZMK*)
273 and cycling genes (i.e. *PCNA*, *MKI67*, *CDC6*) (fig. 4A, S3A). The NKT/MAIT cells were the main
274 "contaminant" in the CCR7$^-$ EVI2B$^+$ sorts (fig. S3B). Based on their upregulation of hallmark
275 CD8αα$^+$ and IELp T cell genes [8,15], five unconventional T cell (UTC) clusters were annotated (fig.
276 4A-B). The unconventional marker genes *TRGC2*, *PDCD1* and *IKZF2* were expressed
277 homogeneously in these UTC clusters, with some expression in the cycling T and MAIT/NKT
278 cells (fig. 4C). The remaining clusters were considered conventional T cell (CTC) clusters based
279 on the expression of typical conventional naive T cell markers (*FOXP1*, *SELL*)(fig. 4A-B).
280 Furthermore, markers (i.e. *NELL2*, *S100B*) which were strongly overexpressed in the bulk
281 transcriptome of the conventional PD-1$^-$ populations in both PNT and CB (fig. 1C, S1C), were
282 homogeneously expressed in the CTC clusters and largely absent in the UTC clusters, except for
283 the GZMK$^+$ DN UTC cluster. Likewise, *DPP4* (encoding CD26) was expressed homogeneously
284 and exclusively by the CTC clusters and strongly in the NKT/MAIT cluster (fig. 4C).

285 Based on the differentially expressed genes, unique distinctive annotations were provided for the
286 individual UTC clusters (table S3). *MME* and *GNLY* (encoding granulysin) are solely expressed
287 by the UTC clusters. *IL32* is highly expressed by IL32$^+$ UTCs but is also expressed at lower levels
288 by different CTC and UTC clusters. The cluster that bridges the bulk of the UTCs and the CTCs
289 had a characteristically high expression of *GZMK* without overexpression of other cytolytic
290 effector genes and a low expression of *CD8A* and *CD8B*, and was annotated as the GZMK$^+$ DN
291 UTC cluster (fig. 4D-E).

292 The UTC clusters expressed the typical TFs of the IELp lineage, *ID3*, *ZNF683*, *IKZF2*, *RUNX3*
293 and *TBX21*, although very limited in the GZMK$^+$ DN UTCs. Importantly, expression of these TFs
294 was absent in the other T cell clusters (fig. 4F). *ID3, ZNF683* and *IKZF2* were expressed highest
295 in the MME$^+$ UTCs. As expected, the CD3$^{+/low}$ PD-1$^+$ population (sort 1 and 3) was enriched for
296 these MME$^+$ UTCs, the recent thymic emigrants (fig. S3B). The activator protein 1 (AP-1) TFs
297 (i.e. *FOS*, *JUN*) were constitutively expressed in both the UTC and NKT/MAIT clusters (fig. 4E).
298 With regard to effector function, low constitutive expression of cytokines (i.e. *TNF* and *IFNG*) and
299 genes involved in cytolysis (i.e. *GNLY*, granzymes) were observed in the GNLY$^+$ UTCs, IL32$^+$
300 UTCs and GZMK$^+$ DN UTCs (fig. 4D-E). These three clusters were considered effector UTC
301 clusters. These effector clusters showed expression of multiple NK receptors (i.e. *KLRC2*, *KLRD1*,
302 *NCR3*, *KIRs*). *EOMES*, which is reported to be absent in murine uIELs, was weakly expressed in
303 the effector clusters. Finally, UTCs expressed important components of the IL-2, IL-7 and IL-15
304 receptors, as well as IL-12 and IL-18 receptor components, which are required for inflammation-

305   induced cytokine responses. Expression of the genes discussed above was mostly absent in CTCs
306   and cycling T cells (fig. 4E).

307

13
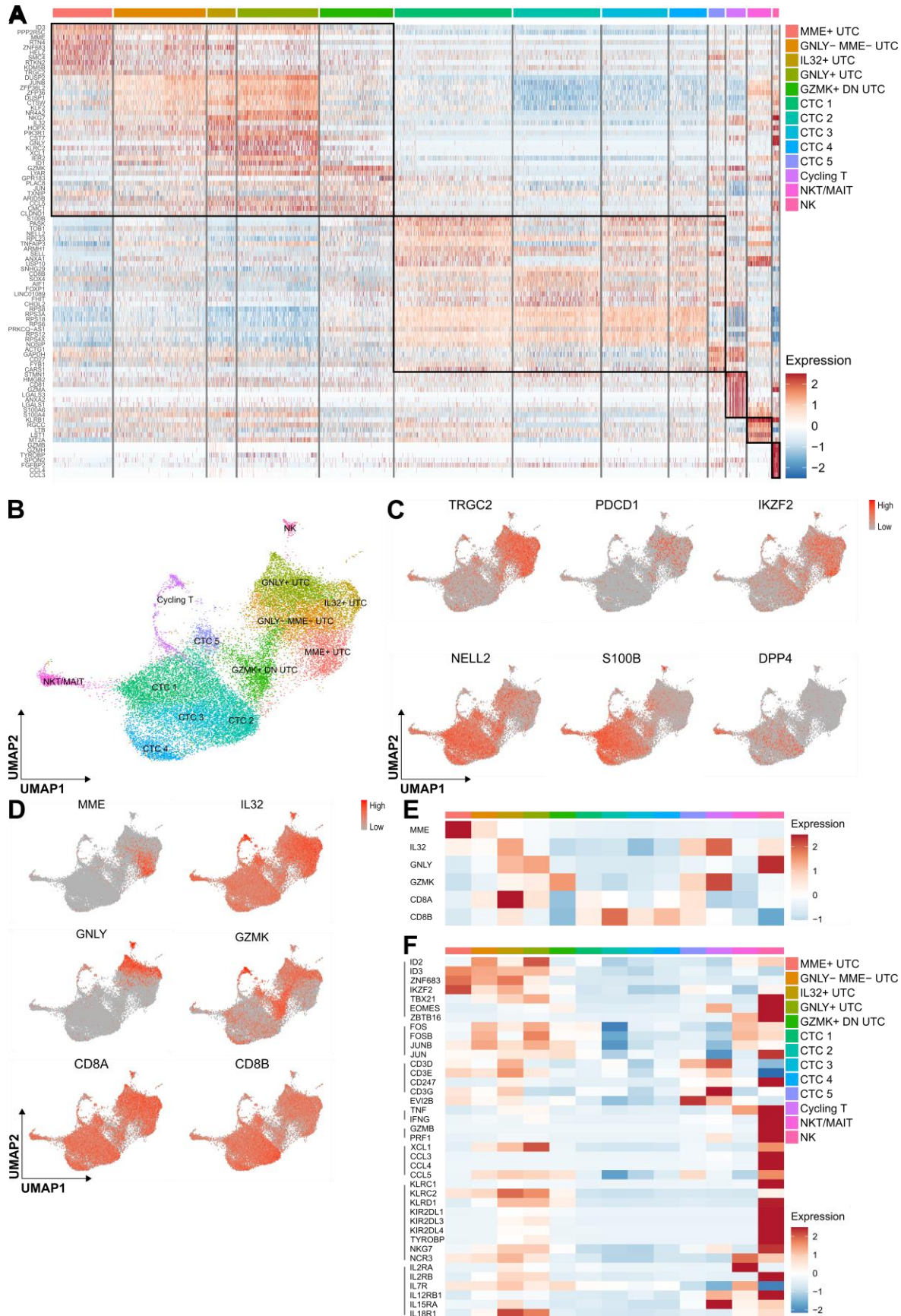
Fig. 4. **Annotation of the UTC clusters in CB.** (**A**) Heatmap showing the expression of the top 10 differentially expressed genes per defined cluster in CB. Recurrent genes are not repeated. The genes are listed in Table S3. (**B**) UMAP visualization of the 13 identified cell clusters in CB. (**C**) UMAP feature plots representing discriminating genes between the UTC and CTC clusters. (**D**) UMAP feature plots representing differentially expressed genes used to annotate the different UTC clusters. (**E**) Heatmap showing the mean expression of the UTC signature genes in the different clusters in CB. (**F**) Heatmap showing the mean expression of characteristic UTC genes in the different clusters in CB.

CB UTCs originate from ZNF683$^+$ CD8αα$^+$ thymocytes

It was established above that the TCR repertoires of the PNT PD-1$^+$ IELp and CB PD-1$^+$ populations are similar, suggesting that the thymic population is the precursor of the CB uIELs. To further explore this developmental pathway, the CB scRNA-seq dataset was integrated with a previously published PNT CD3$^+$ scRNA-seq dataset [15], followed by a trajectory analysis. As published, the CD3$^+$ DP PNT cells diverged into two main pathways: the unconventional pathway consisting of GNG4$^+$ CD8αα(I), ZNF683$^+$ CD8αα(II) and TCRγδ cells and the conventional pathway of CD4$^+$ and CD8$^+$ SP T cells (fig. 5A). Integration with the CB populations showed that the CB UTC clusters partially overlapped with the PNT UTCs, whereas the CB CTC partially overlapped with the PNT SP T cells (fig. 5A-B). When focusing on the UTC pathway, the CB MME$^+$ UTCs overlapped with the PNT CD8αα(I) as well as CD8αα(II) (fig. 5B). Based on the expression of hallmark differentially expressed genes (i.e. *GNG4*, *MME* and *ZNF683*), the CB MME$^+$ UTCs seemed to originate from the PNT ZNF683$^+$ CD8αα(II) rather than from the GNG4$^+$ CD8αα(I) (fig. 5C-D). Therefore, a TSCAN trajectory analysis was performed with the ZNF683$^+$ CD8αα(II) as the population of origin. This analysis revealed a common pathway passing through PNT ZNF683$^+$ CD8αα(II) and CB MME$^+$ UTCs, leading to a branching point at the GNLY$^-$ MME$^-$ UTC cluster. The GNLY$^-$ MME$^-$ UTC cluster gave rise to three distinct lineages: the GNLY$^+$ UTCs (lineage 1), the GZMK$^+$ DN UTC (lineage 2) and the IL32$^+$ UTCs (lineage 3)(fig. 8E-F). The common pathway included ZNF683$^+$ cells, which upregulated *ID3* and differentiated in *TBX21*(T-bet) positive cells (fig. 5G-H). During terminal differentiation, all three lineages expressed high levels of AP-1 TFs and gradually upregulated different effector markers. When analyzing the data in regulons, the transcriptional regulation of the UTCs and the CTCs was significantly different. As expected, the CTCs were mainly regulated by the conventional TFs *FOXP1* and *RORA*, while in the UTCs, *KLF4*, which negatively regulates TCR-mediated proliferation in CD8$^+$ T cells, and *RUNX3* were prominent (fig. 5I)[28,29].
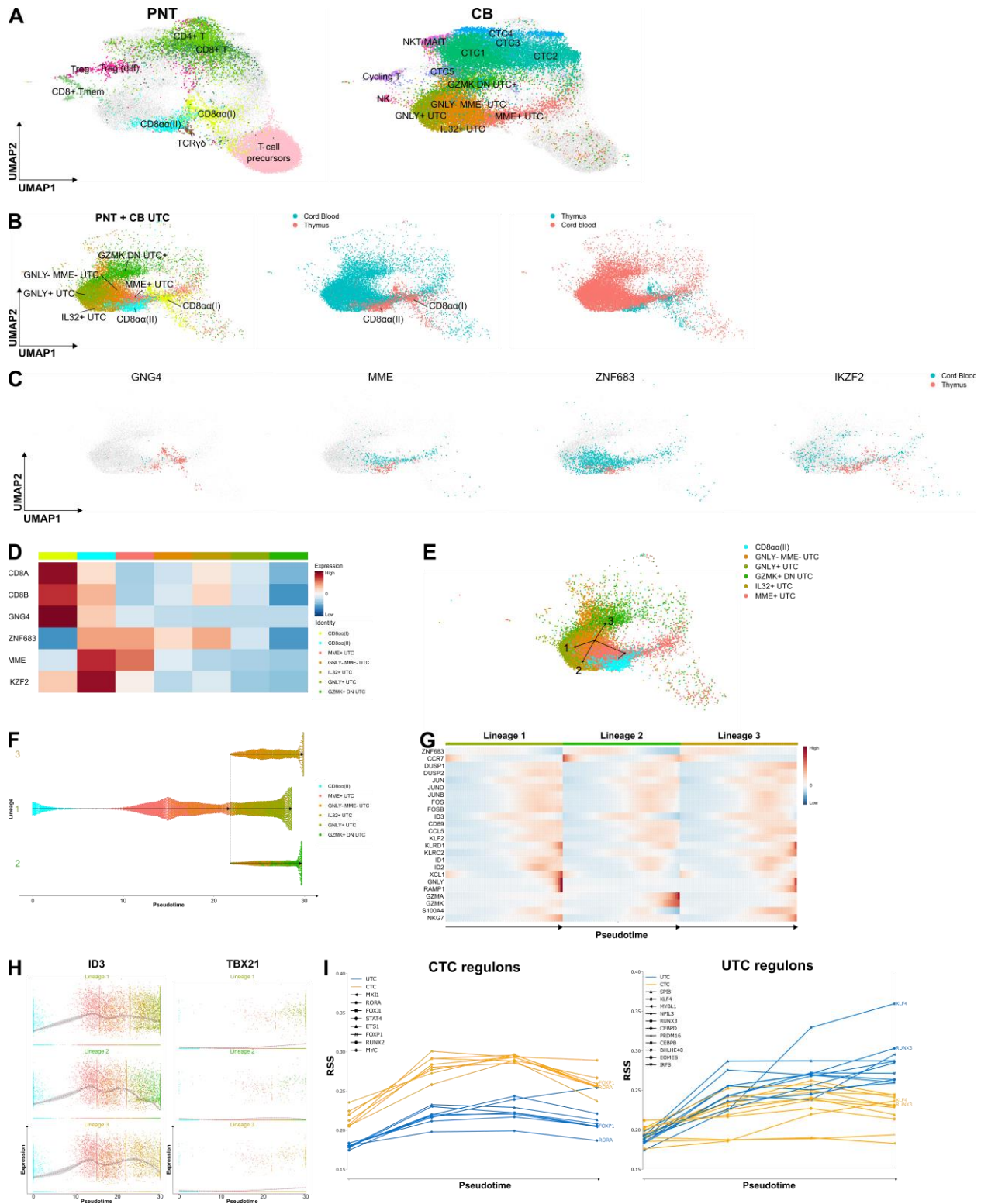
Fig. 5. **UTC pathway analysis reveals 3 effector lineages** (**A**) UMAP visualization of the annotated PNT clusters (left) and CB clusters (right) after integration. (**B**) UMAP visualization highlighting the PNT and CB UTCs and showing the overlay of the integrated UTCs derived from PNT or CB. (**C**) UMAP visualization *GNG4*, *MME*, *ZNF683* and

15

346   *IKZF2* expression by UTCs from PNT (red) or CB (blue). Only cells with a relatively high
347   expression are colored. (**D**) Heatmap showing the mean expression for hallmark
348   differentially expressed genes per UTC population. (**E**) UMAP visualization of the TSCAN
349   trajectory analysis of the UTC populations with the PNT ZNF683$^+$ CD8αα(II) as the origin
350   population. (**F**) Dendrogram of the predicted UTC lineages. (**G**) Heatmap showing the
351   varying gene expression in the pseudotime for the different lineages. (**H**) For each lineage,
352   *ID3* or *TBX21* expression is shown per single cell and summarized as the mean (grey line).
353   (**I**) The regulon specificity score (RSS) is plotted per cell type for the most prominent UTC
354   and CTC pathway. The cell populations on the x-axis are ordered according to pseudo-
355   time.

356 **TCRαβ$^+$ CCR7$^-$ CD26$^-$ cells represent the uIEL population in cord blood**

357 To identify phenotypical differences between the different clusters, CITE-seq was included in sort
358 4 of our scRNA-seq experimental setup (fig. 3C). Membrane protein data were acquired for 3 615
359 single cells across all clusters (fig. 6A). CD10 (encoded by *MME*) was expressed exclusively in
360 the MME$^+$ UTC cluster. A small fraction of the MME$^+$ UTC cluster expressed the thymic T cell
361 immaturity marker CD1a, confirming that this cluster included the recent thymic emigrants.
362 Likewise, PD-1 was expressed by the immature UTCs, as well as by the cycling T cells. In contrast,
363 CD26 (*DPP4*) was strongly expressed by the CTCs and NKT/MAIT cells. The latter also highly
364 expressed CD161 (*KLRB1*), similar to the NK cells. CD54 (*ICAM-1*) and CD244 (*2B4*) are known
365 to be induced in many immune cell types during inflammatory responses [30,31]. CD54 and CD244
366 were highly expressed by the effector type UTCs, but not by the earliest CD1a$^+$ MME$^+$ cells and
367 the GZMK$^+$ DN UTC cluster. Expression of NK receptors such as CD158b (*KIR2DL/DL3*),
368 CD244 and CD94 was mainly observed in the GNLY$^+$ UTC (fig. 6B, S4A, table S4).

369 Because of their scarcity in human CB, the unconventional semi-invariant NKT/MAIT populations
370 have only been studied to a limited extent. The NKT/MAIT cells are the main non-UTC
371 "contaminant" in CCR7$^-$ EVI2B$^+$ cells (sort 2, fig. 3F). However, they can easily be discriminated
372 from the UTCs using the markers CD161, CD26, CD117 (*KIT*), CD194 (*CCR4*), CD103 (*ITGAE*)
373 and CD196 (*CCR6*) (fig. 6C). Our scRNA-seq and CITE-seq data did not incorporate TCR
374 sequencing. Therefore no further distinction was made between NKT or MAIT cells.

375 Based on these CITE-seq results, the uIEL lineage in CB was redefined to include the effector
376 UTC clusters. Flow cytometric analysis of the CD3$^+$ TCRγδ$^-$ cells of a CD4-depleted CB showed
377 a distinct CCR7$^-$ CD26$^-$ population. The UMAP visualization of this CCR7$^-$ CD26$^-$ population
378 clearly showed a gradient of Helios expression with the CD10$^+$ cells having the highest expression.
379 Whereas CD10 stained the immature UTCs, CD54 preferentially stains the effector UTCs. Finally,
380 Eomes$^+$ Granzyme K$^+$ cells are included of which a minority is ultimately CD8α$^-$. CD161$^+$
381 NKT/MAIT cells are not included in the CCR7$^-$CD26$^-$ population (fig. 6D).

382 A population of virtual memory T (T$_{VM}$) cells has been documented in mice, expressing TCRs
383 with a strong binding affinity towards self-antigens. These cells express Eomes in the thymus and
384 acquire a memory phenotype in the periphery [32–34]. A population of CD8$^+$ KIR/NKG2A$^+$ T cells
385 expressing *EOMES* has been described in CB and is proposed as a putative human analog of the
386 T$_{VM}$ population in mice [35]. As KIR expression in CB T cells could only be observed in the UTC
387 clusters (fig. 6B), the CD3$^+$ TCRγδ$^-$ CD4$^-$ CCR7$^-$ CD26$^-$ population was further assessed for these
388 markers. Flow cytometric analysis indeed showed that the cells with the highest expression for
389 KIR or NKG2A expressed Eomes. scRNA-seq analysis clearly indicated that *EOMES*, *KIR*,

16

390   *KLRC1* (encoding NKG2A) and *IL2RB* (encoding CD122) were, besides in the NK cluster, almost
391   exclusively present in the UTC clusters and mainly enriched in the GNLY$^-$ UTC cluster (fig. S4B).
392   No separate cluster of T cells expressing these markers could be observed, suggesting that these
393   T_{VM} cells are part of the uIEL population in human CB.

394   Finally, the size of the CD3$^+$ TCRγδ$^-$ CD4$^-$ CCR7$^-$ CD26$^-$ population expressed as percentage of
395   the CD3$^+$ CD4$^-$ T cells was determined (fig. 6F, S4C). Although the percentages in the different
396   donors varied substantially, the uIEL population is generally significantly larger than the TCRγδ$^+$
397   or CD161^{high} NKT/MAIT population in CB. It constitutes the largest unconventional T cell
398   population in CB.

399
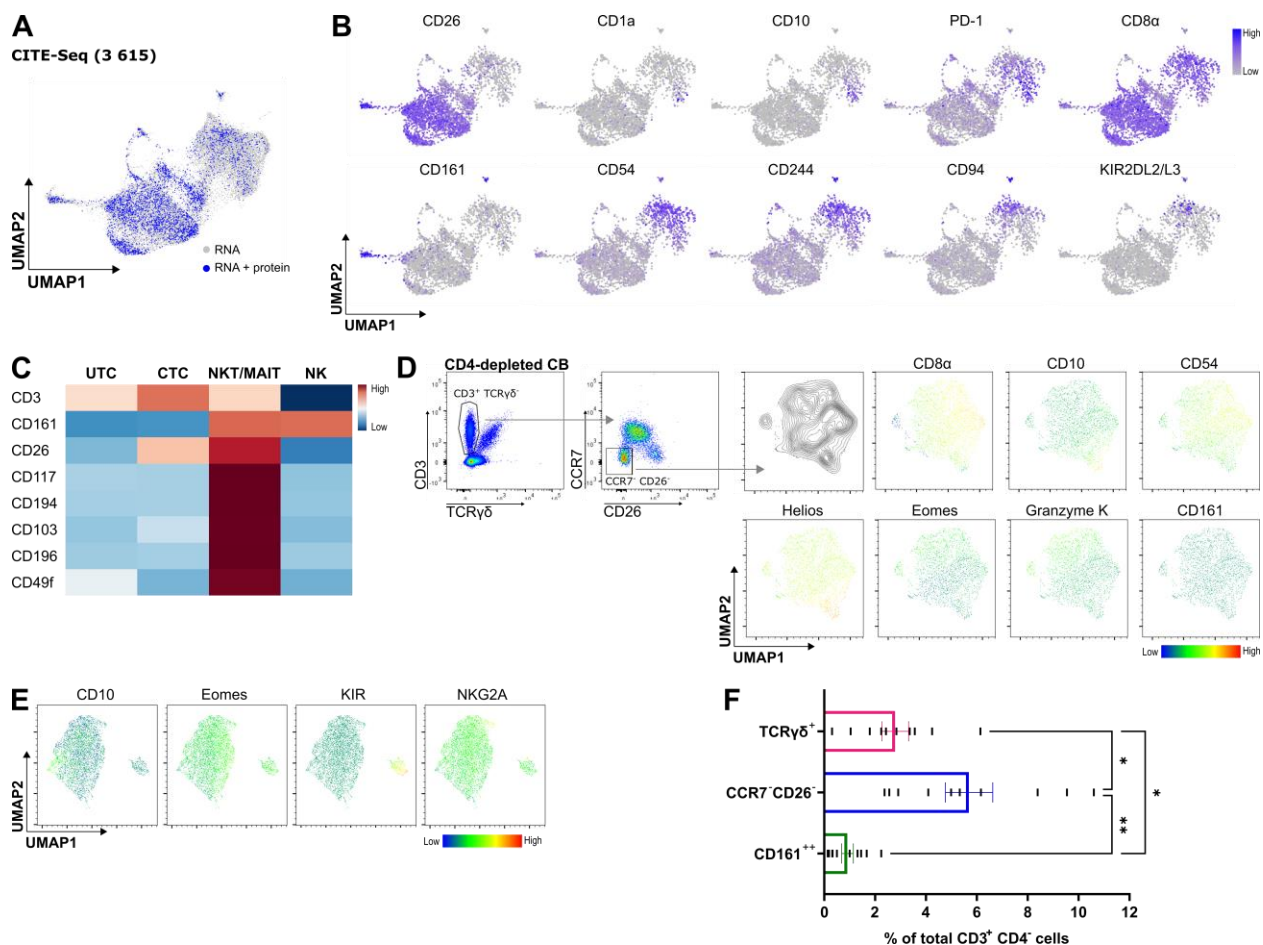
Fig. 6. **CCR7$^-$ CD26$^-$ constitutes the largest unconventional population in CB.** (**A**) UMAP
       visualization of the single cells in the scRNA-seq analysis from which RNA data (grey)
       or combined RNA and protein data (blue) was determined. (**B**) Protein-based UMAP
       visualizations showing the expression of the indicated cell surface protein markers, only
       visualizing the single cells from which protein data was collected. (**C**) Heatmap depicting
       the mean expression of cell surface protein markers used to differentiate the UTC and
       CTC clusters from the NKT/MAIT and NK clusters. (**D**) Representative flow cytometric
       analysis of a CD4-depleted CB. The CD3$^{+/low}$ TCRγδ$^-$ cells were further gated for CCR7$^-$
       CD26$^-$ cells, for which UMAP visualizations are shown. (**E**) Representative flow
       cytometric analysis of CD10, Eomes, KIR (KIR2DL1/DS1, KIR2DL2/DL3 and
   KIR3DL/DS1) and NKG2A expression on the CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CCR7$^-$ CD26$^-$

17

411 population in CB, for which UMAP visualizations are shown. (**F**) Percentage of the CD3$^+$
412 CD4$^-$ cells in CB which are TCRγδ$^+$, CCR7$^-$ CD26$^-$ or CD161^{high} (individual values and
413 mean ± SEM, n = 10). Holm-Šídák's multiple comparisons test was used to assess the
414 statistically significant difference. p-value < 0.05 (*), p-value < 0.01 (**).

Downregulation of CD8β is a unique characteristic of the CCR7$^-$ CD26$^-$ population

416 Many characteristic markers of the uIELs (i.e. PD-1, Helios) are related to activation by
417 autoantigens in the thymus. Therefore, the stability and specificity of these hallmarks was tested
418 during culture-expansion. Both CB populations were culture-expanded with interleukins only, as
419 an *in vitro* equivalent for steady state persistence of the cells in tissues. To include the entire
420 population, the CTCs were isolated from CD4-depleted CBs as CD3$^+$ TCRγδ$^-$ CD8α$^+$ CCR7$^+$ and
421 the UTCs as CD3$^+$ TCRγδ$^-$ CCR7$^-$ CD26$^-$. Similar to the PNT PD-1$^+$ population, the CB UTCs
422 extensively proliferated in the presence of interleukin-15 (IL-15)(fig. S5A-B)[36]. CD26 expression
423 was upregulated by the UTCs in culture, while Helios proved to be a stable distinguishing feature
424 between the CTCs and UTCs, even after proliferation (fig. 7A, fig. S5C). CD158b
425 (KIR2DL2/DL3) was expressed on a minority of the IELs and this expression remained stable
426 during culture. Moreover, no expression could be observed on IL-7 culture-expanded CTCs (fig.
427 7A, S5D-E). This suggested that all KIR$^+$ T cells detected *in vivo* belong to the UTC lineage. CD8β
428 expression decreased on UTCs during culture. whereas expression on CTCs was stable (fig. 7B).

429 To investigate the long-term stability of the CB phenotype, single cell clones of both lineages were
430 culture-expanded and the phenotype of the resulting clones was analyzed. CD8αα$^+$ and DN clones
431 were frequently observed in UTC-derived clones, whereas CTC-derived clones remained
432 predominantly CD8αβ$^+$ (fig. 7C-D). This confirms that downregulation of CD8(β), and therefore
433 the expression of CD8αα homodimers, is an exclusive characteristic of UTCs. Although Helios
434 was induced in CTC-derived clones, expression remained significantly higher in UTC-derived
435 clones. Similarly, CD26 expression was induced on UTC-derived clones, but remained
436 significantly higher in CTC-derived clones (fig. 7E). In conclusion, the markers CD26, Helios,
437 CD8α, CD8β and KIR reliably discriminated between UTCs and CTCs, even after culture
438 expansion and activation.

439 Functional testing of the CD3$^+$ TCRγδ$^-$ CCR7$^-$ CD26$^-$ population revealed *ex vivo* CD3-induced
440 killing activity. Upon activation, a spectrum of chemokines including IL-8, MIP-1α (CCL3), MIP-
441 1β (CCL4) and fractalkine (CX3CL1) and the cytokines IL-2, FLT-3L, PDGF-AA, GM-CSF, IL-
442 10, IFN-γ and TNFα was produced (fig. S5F-H).

18
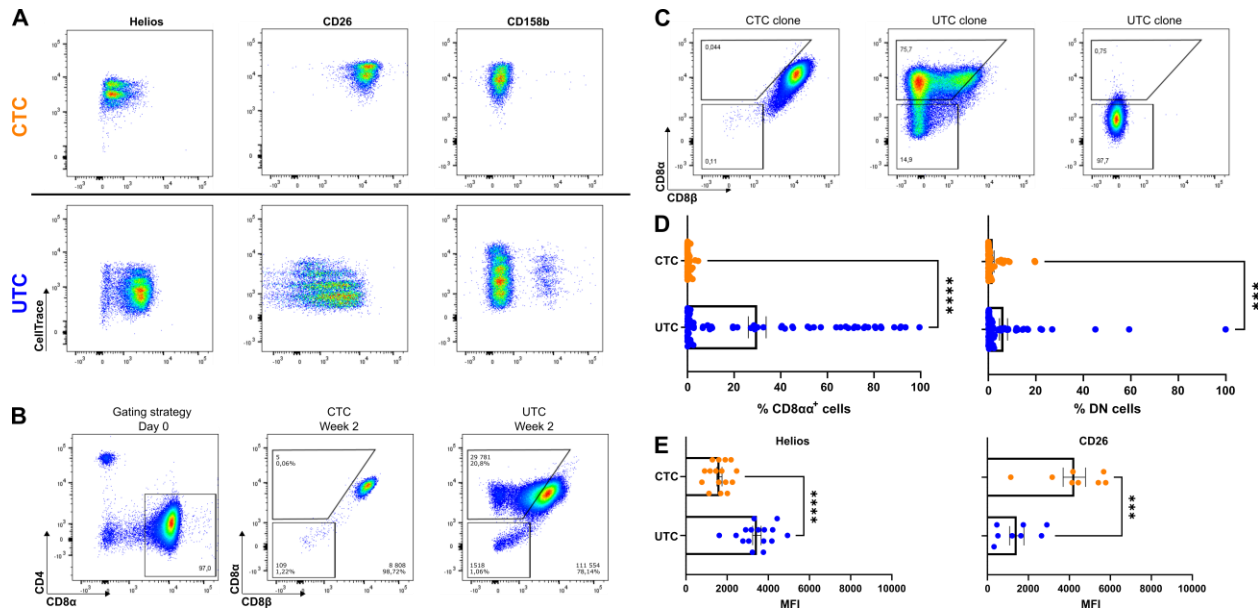
443

444 **Fig. 7**. **Stable phenotypic uIEL markers following proliferation.** (**A**) Representative dot plots
445 of flow cytometric markers expressed by the CTCs after 5 days of incubation with IL-7 (10
446 ng/mL, upper row) and by the UTCs after incubation with IL-15 (10 ng/mL, bottom row).
447 Proliferation assessed by CellTrace Violet dye dilution. (**B**) Both the CTCs and UTCs were
448 sorted as strictly CD4$^-$ CD8α$^+$ (left). Representative assessment of CD8α and CD8β
449 expression by the CTC and UTC populations after two weeks of proliferation with IL-15.
450 (**C**) Representative examples of CD8α and CD8β expression by CTC and UTC clones. (**D**)
451 Percentage of CD8αα$^+$ or DN cells (gated as shown in figure 7C) in CTC or UTC clones
452 after at least 2 weeks of expansion. Individual values and mean ± SEM of 60 clones per
453 population are shown. Mann-Whitney was used to assess statistical significance. (**E**) Mean
454 fluorescence intensity (MFI) of Helios and CD26 expression by CTC or UTC clones.
455 Individual values and mean ± SEM of 8 or 16 clones per population are shown. Unpaired
456 t-tests were used to assess statistical significance. p-value < 0.001 (***), p-value < 0.0001
457 (****).

DISCUSSION

459 In the present study, the peripheral progeny of the human thymic IELps was identified in CB as a
460 separate population evidently distinct from conventional CD8$^+$ T cells and from NKT/MAIT cells.
461 This population was remarkably heterogeneous. Five clusters were identified of which the MME$^+$
462 UTCs and the GNLY$^-$ MME$^-$ UTCs are the closest related progeny of the thymic CD10$^+$ PD-1$^+$
463 population and are themselves the precursor clusters of the three differentiated clusters. It is
464 therefore possible that in adulthood, when the output of the thymus diminishes, these two precursor
465 populations gradually differentiate into effector cells and are no longer detectable in the blood and
466 tissues. Of the effector clusters, the GNLY$^+$ UTC and IL32$^+$ UTC both contain cells expressing
467 NK receptors including KIRs. KIR$^+$ cells were not detected in the CTC clusters, *ex vivo* or after
468 culture, indicating that KIR expression is confined to the uIEL lineage cells. A fifth UTC cluster,
469 adjacent to the CTCs, was notable for high *GZMK* expression and low expression of *CD8A* and
470 *CB8B*. A distinctive set of markers to further study this cluster was not found. Although some
471 characteristics of the CTCs were present, it was argued that the GZMK$^+$ DN UTC cluster also
472 belongs to the uIEL lineage, because GZMK$^+$ DN UTCs consistently co-occurred with the UTC

19

473 population either in the CD3$^{+/low}$ PD-1$^+$ or CCR7$^-$ EVI2B$^+$ sorts and constitutively expressed low
474 levels of CD8α and CD8β, in contrast to CTCs. Moreover, CD8$^+$ UTCs have the discriminating
475 ability to downregulate or lose CD8β (and CD8α) expression after *in vitro* stimulation whereas
476 this was not observed for CD8$^+$ CTCs.

477 The uIEL lineage in CB was likewise heterogeneous with regard to the expression of hallmark
478 protein markers such as CD8β, PD-1, Helios and NK receptors. Despite the variable phenotypes
479 within the uIEL population, evidence is presented that this population is well defined in CB with
480 the markers CD26 and CCR7: the CCR7$^-$ fraction of TCRαβ$^+$ CD8$^+$ cells consisted exclusively of
481 CD26$^-$ uIELs and CD26$^+$ NKT/MAIT cells. However, CCR7 and to a lesser extent CD26 and
482 Helios expression was unstable during T cell activation, which made these markers less valuable
483 for analysis of samples containing antigen-experienced T cells such as adult blood. Hence, the
484 presence of KIR$^+$ T cells, CD8αα$^+$ cells and DN T cells (excluding TCRγδ and NKT/MAIT cells),
485 other markers that defined subtypes of this lineage and were not affected by T cell activation, was
486 investigated and validated as a characteristic of the uIELs.

487 It was evidenced in CB that all KIR$^+$ CD8$^+$ T cells belong to the uIEL lineage, but only constitute
488 a small fraction of the total uIEL population. Moreover, as CD8αα$^+$ and DN cells could only be
489 generated from UTCs and not from CTCs, KIR$^-$ CD8αα$^+$ and possibly DN T cells (after exclusion
490 of TCRγδ$^+$, NKT and MAIT cells) most likely also belong to the uIEL lineage. In addition to their
491 transcriptional profile, it is therefore suggested that previously described populations, such as
492 immunosuppressive KIR$^+$ CD8$^+$ T cells, KIR/NKG2A$^+$ CD8$^+$ T$_{VM}$ cells and CD8αα$^+$ cells
493 described in tissues, probably belong to the unconventional, broadly reactive uIEL lineage rather
494 than to the foreign peptide-reactive conventional CD8$^+$ T cell lineage [32–35,37].

495 KIR$^+$ CD8$^+$ T cells were recently isolated from human adult blood based on their promiscuous
496 binding to tetramers of HLA-A3 and HLA-A11. These cells express KLRC2, KIR2DL3 and
497 NCR3, all of which are expressed by cells within the GNLY$^+$ UTC and IL32$^+$ UTC clusters. These
498 cells are characterized by prominent expression of *IKZF2* and *ZNF683*, two characteristics of the
499 uIEL lineage [38]. KIR$^+$ CD8$^+$ cells, expressing *IKZF2*, are found to be enriched at inflammatory
500 sites induced by viral infection as well as autoimmune inflammation in patients with celiac disease,
501 multiple sclerosis and lupus. Importantly, these KIR$^+$ CD8$^+$ cells are shown to suppress
502 autoreactivity by direct killing of pathogenic CD4 T cells [39]. Functionally and phenotypically
503 similar cells were described earlier in mice as CD8$^+$ T$_{reg}$ cells [40–42]. Here, the uIELs were found to
504 be able to kill target cells *ex vivo*. Furthermore, uIELs could produce chemokines (i.e. CCL3,
505 CCL4 and CCL5), demonstrated as a mechanism to attract T cells [43]. It is therefore possible that
506 uIELs described here, may display *in vivo* immune suppressive properties by killing of autologous,
507 metabolically active immune blasts. Indeed, it was previously reported that IELps are generated in
508 the thymus by agonist selection at the early active DP blasts stage [12]. As the IELps lineage is
509 agonist selected on hematopoietic cells and not on thymic epithelial cells, it would be possible that
510 IELps specifically react to autoantigens expressed by potent, metabolically active immune blasts
511 [44,45]. Therefore, whereas natural T$_{reg}$ cells are reactive to antigens presented in the thymic medulla,
512 the spectrum of antigens to which the uIELs react is probably presented by early DP blasts in the
513 thymic cortex.

514 CD8αα$^+$ cells have been described in various organs of the human body, including liver and lungs,
515 and at tumor sites [9,10,46,47]. CD8αα$^+$ herpes simplex virus-specific T cells are found at the dermal-
516 epidermal interface [48].

517 DN T cells represent a yet poorly characterized subset of TCRαβ[+] T cells, partially due to their
518 relatively low frequencies in human peripheral blood. They have mainly been studied in the
519 context of autoimmunity, where increases of a heterogeneous group of DN T cells have been
520 reported in patients with systemic lupus erythematosus and other autoimmune diseases. It is shown
521 in human that peripheral DN T cells are maintained primarily by differentiation from CD8[+] T cells
522 [49]. It is also reported in mice that only CD8[+] T cells expressing PD-1 and Helios convert to DN T
523 cells after encountering autoantigens [37]. ScRNA-seq of splenic DN T cells in mice revealed five
524 DN clusters of which two clusters were characterized by high expression of *IKZF2*, *IL2RB*, NK
525 receptor genes such as *KLRD1* and chemokines such as *XCL1* and *CCL5*, genes also overexpressed
526 in the uIEL lineage cells described here [50]. Despite their low frequencies, DN T cells are potent
527 producers of cytokines and therefore essential for immune responses. The therapeutic potential of
528 DN T cells as Chimeric antigen receptor (CAR)-T cell is recently explored. It is evidenced that
529 DN CAR-T cells are as effective as conventional CAR-T cells, without inducing toxicity,
530 demonstrating the potential of using allogeneic DN T cells [51]. A relationship between these DN T
531 cells and the uIELs has not yet been conclusively addressed.

532 Finally, T_{VM} cells are described in mice as a subtype of CCR7[+] conventional T cells with somewhat
533 higher affinity for self-ligands in the thymus. The murine T_{VM} phenotype is imposed by self-
534 reactivity of the TCR, similar to IELps. These cells start to express Eomes upon leaving the thymus
535 and become activated in the periphery, proliferate on IL-15 and home to the tissues [32–34]. The
536 human counterpart is reported to be KIR/NKG2A[+] Eomes[+] and was detected in adult blood as well
537 as CB samples [35]. Here, it is reported that indeed a population is present in the UTC clusters in CB
538 that weakly expresses Eomes and either KIRs or NKG2A.

539 In conclusion, the full spectrum of the uIEL lineage present in human CB is here described. The
540 more immature uIELs, derived from thymic IELps, differentiate further and generate a
541 heterogeneous mixture of effector cells. This population includes excellent killers, which may
542 contribute to immune defence activity as well as exert immune suppressive activity by killing
543 autologous immune blasts. The concept of these cells originating in the thymus by agonist selection
544 as a consequence of the special characteristics of their TCR, may throw a different light on the
545 study of these cells in immune reactivity to pathogens as well as autoimmunity.

MATERIALS AND METHODS

Sample Processing

548 Human CB was obtained from the Cord Blood Bank UZ Gent. Samples were used following the
549 guidelines of the Medical Ethical Committee of Ghent University Hospital (CG20171208A, 8
550 December 2017) after informed consent had been obtained in accordance with the Declaration of
551 Helsinki. Mononuclear cells were isolated using density gradient centrifugation (LymphoPrep;
552 Axis-Shield, 1114547) and were enriched by magnetically activated cell sorting (MACS) through
553 negative selection using anti-CD4-biotin, anti-CD14-biotin, anti-CD19-biotin, anti-CD235-biotin
554 (homemade) and anti-biotin Microbeads (Miltenyi Biotec, 130-090-485). Human postnatal thymus
555 was processed as previously described [36].

Flow Cytometry and Antibodies

557 Staining of surface markers was performed in DPBS (Lonza, 17-512F) with 1% fetal calf serum
558 (FCS; Biowest, S1810) using the antibody to cell ratio recommended by the supplier. Intracellular
559 and intranuclear stainings were performed following the supplier's protocol using BD

560 Cytofix&Cytoperm (BD Biosciences, 554714) and the eBioscience™ Foxp3 / Transcription
561 Factor Staining Buffer Set (eBioscience, 00-5523-00) respectively. Flow cytometric analysis was
562 performed on the LSR II and cell sorting on the FacsARIA Fusion (both BD Biosciences). Flow
563 cytometry data were analyzed using FACS DIVA software (BD Biosciences) and FlowJo software
564 (TreeStar Inc). Viable cells were gated based on propidium iodide (PI) negativity or Fixable
565 Viability Dye (eFluor 506; Thermo Fisher Scientific, 65-0866-18) negativity for surface and
566 intracellular stainings respectively. The following list of anti-human monoclonal antibodies was
567 used. Allophycocyanin (APC)/AF647-conjugated: CD4 (Miltenyi, 130-113-250), CD158b
568 (KIR2DL2/DL3, Miltenyi, 130-092-617), Helios (Biolegend, 137221), Granzyme K (Biolegend,
569 370503), NKG2A (CD159a, Miltenyi, 130-114-089), PD-1 (CD279, Biolegend, 367420), TCRγδ
570 (Miltenyi, 130-113-500); APC Cy7/APC Fire750-conjugated: CD8α (Biolegend, 344746), CCR7
571 (CD197, Biolegend, 353246); Brilliant Violet 421-conjugated: CD3 (Biolegend, 317344), CD54
572 (Biolegend, 353131); Brilliant Violet 605-conjugated: CD161 (Biolegend, 339915); Brilliant
573 Violet 650-conjugated: CD3 (Biolegend, 317323); Brilliant Violet 711-conjugated: CCR7
574 (CD197, Biolegend, 353227); Brilliant Violet 785-conjugated: CD10 (Biolegend, 312237);
575 Fluorescein isothiocyanate (FITC)-conjugated: CD8α (homemade), CD161 (Miltenyi, 130-114-
576 118), IFN-γ (BD Biosciences, 554551), TCRγδ (BD Biosciences, 347903); Phycoerythrin (PE)-
577 conjugated: CD158a/h (KIR2DL1/DS1, Miltenyi, 130-116-975), CD158b1/b2 (KIR2DL2/DL3,
578 Beckman Coulter, IM2278U), CD158e1/e2 (KIR3DL/DS1, Beckman Coulter, IM3292), EVI2B
579 (CD361, Thermo Fisher Scientific, A15806), Granzyme B (eBioscience, 12-8899-42), Granzyme
580 K (Biolegend, 370511), PD-1 (CD279, Biolegend, 367404), Perforin (eBioscience, 12-9994-42);
581 PE Cy7-conjugated: CD8β (eBioscience, 25-5273-42), CD10 (Biolegend, 312214), Eomes
582 (eBioscience , 25-4877-41), TCRγδ (Biolegend, 331222); Peridinin chlorophyll protein complex
583 (PerCP) Cy5.5-conjugated: CD4 (Biolegend, 344608), CD26 (Biolegend, 302715). CB MNCs
584 were stained with Vδ1 (clone A13 supernatant, which can bind to Vδ1 when incorporated in hybrid
585 Vδ1-Jα-Cα TCR chains, a kind gift from Prof. Dr. Lorenzo Moretta's laboratory), anti–mouse Ig
586 light chain κ (Biolegend, 409506), 5% normal mouse serum (Invitrogen, 10410), followed by the
587 appropriate antibodies above to isolate the populations.

588 CD4/CD14/CD19/CD235-depleted CB MNCs were sorted into CD3$^+$ TCRγδ$^-$ CD4$^-$ CD8α$^+$ PD-1$^-$
589 (PD-1$^-$ population) and CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CD8α$^+$ PD-1$^+$ (PD-1$^+$ population) for the
590 transcriptome, TCR and proteome analyses. The PNT populations were sorted as previously
591 described [36]. For the single cell assay, CD4/CD14/CD19/CD235-depleted CB MNCs were sorted
592 into CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CD8α$^+$ PD-1$^+$ (sort 1), CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CD8α$^+$ CCR7$^-$ EVI2B$^+$
593 (sort 2), CD3$^+$ TCRγδ$^-$ CD4$^-$ PD-1$^-$ and CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ PD-1$^+$ (sort 3), CD3$^+$ TCRγδ$^-$ CD4$^-$
594 CCR7$^+$ EVI2B$^-$ and CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CCR7$^-$ EVI2B$^+$ (sort 4). For the proliferation and
595 stimulation experiments, CD4/CD14/CD19/CD235-depleted CB MNCs were sorted into CD3$^+$
596 TCRγδ$^-$ CD4$^-$ CD8α$^+$ CCR7$^+$ (CTC) and CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CCR7$^-$ CD26$^-$ (UTC).

RNA Sequencing

598 The populations of interest were each time sorted from three CB and PNT donors. The CD3$^+$
599 TCRγδ$^+$ population was also sorted, only 2 TCRγδ$^+$ samples were analyzed for CB. The PNT and
600 CB populations were sorted in IMDM (Thermo Fisher Scientific, 12440053) supplemented with
601 10% FCS, 2 mM L-glutamine (Thermo Fisher Scientific, 25030-081), 100 IU/mL penicillin and
602 100 IU/mL streptomycin (Thermo Fisher Scientific, 15140-122) (complete IMDM, cIMDM) and
603 washed 3 times in phosphate-buffered saline (PBS). RNA extraction was performed using the
604 miRNeasy Mini Kit (Qiagen, 217004). For poly(A) RNA-seq, the QuantSeq 3′ mRNA FWD kit

605 (Lexogen) was used, followed by single-ended sequencing on the NextSeq500 Sequencing System
606 (Illumina) with a read length of 75bp. RNA-seq reads were aligned to hg38-noalt using STAR
607 v2.6.0c and quantified on Ensembl v93.

TCR sequencing

609 For the TCR analysis, the populations of interest were sorted from six CB donors and two new
610 PNT donors. Previously published PNT samples were also reanalyzed (12). RNA extraction was
611 performed using the miRNeasy Micro Kit (Qiagen, 217084), followed by template-switch
612 anchored RT-PCR. High-throughput sequencing of TRA and TRB loci was performed as
613 previously described [52]. Raw sequencing reads from fastq files were aligned to reference V, D and
614 J genes from GenBank database specifically for 'TRA' or 'TRB' to build CDR3 sequences using
615 the MiXCR software version 3.0.12 [53]. Following, the CDR3 sequences were analysed using
616 VDJtools software version 1.2.1 [54]. Out of frame sequences were excluded from the analysis, as
617 well as non-functional TRA and TRB segments using *IMGT* (the international ImMunoGeneTics
618 information system®) annotation. TRDV gene segment-containing sequences were filtered as
619 well, except for the analysis where the amount of TRDV1 containing sequences was assessed.
620 Calcbasicstats default function was used to calculate the number of CDR3 N additions. Cumulative
621 gene segment plots were generated using the output from CalcSegmentUsage function. Tree maps
622 were generated using the Treemap Package on RStudio, grouping TRAV and TRAJ segments
623 according to their locus position. D75 repertoire diversity metrics were calculated by measuring
624 the percentage of clonotypes required to occupy 75% of the total TCR repertoire. Determination
625 of the CDR3α and CDR3β apex region and cysteine usage was performed following previously
626 described indications [14]. Hydrophobic CDR3α and CDR3β doublet containing sequences were
627 determined by calculating the percentage of sequences using any of the 175 amino acid doublets
628 previously identified as promoting self-reactivity [13]. Physicochemical characteristics (strength,
629 volume and polarity) of the CDR3β were analysed using VDJtools software version 1.2.

LC-MS/MS proteomic analysis

631 *Sample preparation*: The populations of interest were each time sorted from three CB and PNT
632 donors. Cell pellets ($\pm 1.10^6$ cells per pellet) were resuspended in lysis buffer (8 M urea; 20 mM
633 HEPES, pH 8.0). Samples were sonicated by three pulses of 15 s, interspaced by 1 min pauses on
634 ice at an intensity output of 15 W, and centrifuged for 15 min at 20 000g at room temperature to
635 remove insoluble components. Proteins were reduced with 5 mM dithiothreitol (DTT) (Sigma-
636 Aldrich) for 30 min at 55 °C and then alkylated by the addition of 10 mM iodoacetamide (Sigma-
637 Aldrich) for 15 min at room temperature in the dark. Samples were further diluted with 20 mM
638 HEPES, pH 8.0, to a final urea concentration of 4 M and proteins were digested with LysC (Wako)
639 (1/100, w/w) for 4 hours at 37°C. Samples were again diluted to 2 M urea and digested with trypsin
640 (Promega) (1/100, w/w) overnight at 37℃. The resulting peptide mixture was acidified by addition
641 of 1% trifluoroacetic acid (TFA). Peptides were then purified on a SampliQ SPE C18 cartridge
642 (Agilent), vacuum-dried, and kept at −20 °C until measured by LC-MS/MS.

643 *LC–MS/MS Analysis*: Immediately before injection, purified peptides were redissolved in 15 µL
644 loading solvent (0.1% trifluoroacetic acid/water/acetonitrile (0.1:98:2, v/v/v)) and the peptide
645 concentration was determined by measuring on a Lunatic spectrophotometer (Unchained Labs). 2
646 µg of peptide material of each sample was injected for LC-MS/MS analysis on an Ultimate 3000
647 RSLC nano-LC (Thermo Fisher Scientific, Bremen, Germany) in-line connected to a Q Exactive
648 HF mass spectrometer (Thermo Fisher Scientific) equipped with a nanospray flex ion source

649 (Thermo Fisher Scientific). Trapping was performed at 10 µL/min for 4 min in loading solvent A
650 on a 20-mm trapping column (made in-house, 100-µm internal diameter, 5-µm beads, C18
651 Reprosil-HD, Dr Maisch, Germany). Peptide separation after trapping was performed on a 200-
652 cm-long micropillar array column (PharmaFluidics) with C18-endcapped functionality. The
653 Ultimate 3000's column oven was set to 50°C. For proper ionization, a fused silica PicoTip emitter
654 (10-µm inner diameter) (New Objective) was connected to the µPAC outlet union and a grounded
655 connection was provided to this union. Peptides were eluted by a nonlinear gradient from 1 to 55%
656 MS solvent B (0.1% FA in water/acetonitrile (2:8, v/v)) over 145 min, starting at a flow rate of
657 750 nL/min switching to 300 nL/min after 15 min, followed by a 15-min washing phase plateauing
658 at 99% MS solvent B. Re-equilibration with 99% MS solvent A (0.1% FA in water) was performed
659 at 300 nL/min for 45 min followed by 5 min at 750 nL/min adding up to a total run length of 210
660 min. The mass spectrometer was operated in a data-dependent, positive ionization mode,
661 automatically switching between MS and MS/MS acquisition for the 16 most abundant peaks in
662 each MS spectrum. The source voltage was 2.2 kV, and the capillary temperature was 275 °C. One
663 MS1 scan (m/z 375–1,500, AGC target 3.10^6 ions, maximum ion injection time 60 ms), acquired
664 at a resolution of 60,000 (at 200 m/z), was followed by up to 16 tandem MS scans (resolution
665 15,000 at 200 m/z) of the most intense ions fulfilling predefined selection criteria (AGC target
666 1.10^5 ions, maximum ion injection time 80 ms, isolation window 1.5 Da, fixed first mass 145 m/z,
667 spectrum data type: centroid, intensity threshold 1.3×10^4, exclusion of unassigned, 1, 7, 8, >8
668 positively charged precursors, peptide match preferred, exclude isotopes on, dynamic exclusion
669 time 12 s). The higher-energy collisional dissociation was set to 28% normalized collision energy,
670 and the polydimethylcyclosiloxane background ion at 445.12003 Da was used for internal
671 calibration (lock mass).

672 *Data analysis:* Data analysis was performed with MaxQuant (version 1.6.2.6) using the
673 Andromeda search engine with default search settings including a false discovery rate (FDR) set
674 at 1% on both the peptide and protein level. Spectra were searched against the human Swiss-Prot
675 database (from November 2018 with 20 424 entries) separately for PNT and CB T cells. The mass
676 tolerance for precursor and fragment ions was set to 4.5 and 20 ppm, respectively, during the main
677 search. Enzyme specificity was set to the C-terminal of arginine and lysine, also allowing cleavage
678 next to prolines with a maximum of two missed cleavages. Variable modifications were set to
679 oxidation of methionine residues and acetylation of protein N-termini. Matching between runs was
680 enabled with a matching time window of 1.5 min and an alignment time window of 20 min. Only
681 proteins with at least one unique or razor peptide were retained, leading to the identification of
682 4539 and 3584 proteins for PNT and CB, respectively. Proteins were quantified by the MaxLFQ
683 algorithm integrated into the MaxQuant software. A minimum ratio count of two unique or razor
684 peptides was required for quantification. Further data analysis was performed with the Perseus
685 software (version 1.6.2.1) separately for the PNT and CB data set after uploading the protein
686 groups file from MaxQuant. Reverse database hits, potential contaminants and proteins that are
687 only identified by peptides carrying at least one modified amino acid were removed. Replicate
688 samples were grouped and proteins with less than three valid values in at least one group were
689 removed, and missing values were imputed from a normal distribution around the detection limit
690 resulting in 3,001 and 2,024 quantified proteins for PNT and CB, respectively.

691 **Gene set enrichment analysis**

692 GSEA was performed using the GSEA software version 4.1.0., a joint project of UC San Diego
693 (San Diego, CA, USA) and Broad Institute (Cambridge, MA, USA)[55,56]. The GSEAPreranked tool

24

694 was run using standard parameters and 1000 permutations. The gene set contained the significantly
695 differentially expressed genes when comparing the CD10$^+$ PD-1$^+$ IELp population to the CD10$^-$
696 PD-1$^-$ cells from human PNT. The gene list was ranked by comparing the PD-1$^+$ population to the
697 PD-1$^-$ cells from human CB, ranked from the upregulated genes (left) to the downregulated genes
698 (right). The normalized enrichment score (NES) reflects the degree to which the gene set is
699 overrepresented in the upregulated genes (positive value) or downregulated genes (negative value).
700 The false discovery rate q value (FDR q) is the estimated probability that a gene set with a given
701 NES represents a false-positive finding.

Single-cell RNA sequencing analysis

703 *Single cell library preparation and sequencing:* The populations of interest were sorted from
704 CD4/CD14/CD19/CD235-depleted CB from two different donors. From the first donor, CD3$^{+/low}$
705 TCRγδ$^-$ CD4$^-$ CD8α$^+$ as CD3$^{+/low}$ PD-1$^+$ (sort 1) or CCR7$^-$ EVI2B$^+$ (sort 2) were sorted separately.
706 Both fractions were labeled with different TotalSeq anti-human Hashtag antibodies (Biolegend)
707 before being pooled in equal portions and processed together. From the second donor, CD3$^+$
708 TCRγδ$^-$ CD4$^-$ PD-1$^-$ and CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ PD-1$^+$ (sort 3), CD3$^+$ TCRγδ$^-$ CD4$^-$ CCR7$^+$ EVI2B$^-$
709 and CD3$^{+/low}$ TCRγδ$^-$ CD4$^-$ CCR7$^-$ EVI2B$^+$ (sort 4) were sorted. Considering the smaller
710 percentage of unconventional T cells, the conventional cells were sorted separately and later added
711 in equal portions. Sort 4 was combined with CITE-seq labeling. Cells were incubated for 30 min
712 on ice with 50 µL of staining mix in PBS containing 0.04% BSA, Fc receptor block (PN 422301,
713 TruStain FcX, BioLegend) and a human cell surface protein antibody panel containing 277 oligo-
714 conjugated antibodies (TotalSeq-A, BioLegend) including 6 TotalSeq-A isotype controls (table
715 S4). TotalSeq antibodies were diluted in concentrations as recommended by the manufacturer.
716 Sorted single-cell suspensions were resuspended at an estimated final concentration of 1200
717 cells/µl and loaded on a Chromium GemCode Single Cell Instrument (10x Genomics) to generate
718 single-cell gel beads-in-emulsion (GEM) at the VIB Single Cell Core.The scRNA-Seq libraries
719 were prepared using the GemCode Single Cell 3' Gel Bead and Library kit, version NextGEM 3.1
720 (10x Genomics) according to the manufacturer's instructions with the addition of amplification
721 primers   (3nM),   5'CCTTGGCACCCGAGAATT*C*C   and
722 5'GTGACTGGAGTTCAGACGTGTGC*T*C during cDNA amplification to enrich the
723 TotalSeq-A cell surface and hashtag protein oligos. Library construction was performed according
724 to the manufacturer's instructions. Sequencing libraries were loaded on an Illumina NovaSeq flow
725 cell at VIB Nucleomics core with sequencing settings according to the recommendations of 10x
726 Genomics, pooled in a 80:25 ratio for the combined 3' gene expression and cell surface protein
727 libraries, respectively.

728 *Preprocessing of the scRNA-seq and CITE-seq data:* The Cell Ranger pipeline (10x Genomics,
729 version 3.1.0) was used to perform sample demultiplexing and to generate FASTQ files for read
730 1, read 2 and the i7 sample index for the gene expression and cell surface protein libraries. Read 2
731 of the gene expression libraries was mapped to the reference genome (GRCh38.99) using STAR.
732 The resulting count matrices were subsequently loaded into R for further processing using Seurat
733 version 4.0.5 [57]. Empty and/or damaged cells were removed from the datasets by filtering on the
734 following three parameters: (i) number of genes per cell (nFeature > 600), (ii) number of UMI
735 counts per cell (nCount > 1150) and (iii) percentage mitochondrial genes per cell (percent.mt <
736 15). The remaining cells were normalized with the built-in normalization function from the Seurat
737 package using the log normalization method and a scale factor of 10000. Highly variable
738 features/genes were identified with FindVariableFeatures with the selection method set to vst and

739 the nfeatures parameter to 4000. Differential gene expression analysis between cell clusters and
740 conditions was performed using Wilcoxon Rank Sum test through the Seurat function
741 "FindMarkers". P-value adjustment was performed using Bonferroni correction.

742 CITE-seq antibody reads were quantified using the feature-barcoding functionality within the
743 Seurat package. Antibodies with low expression were filtered out based on inspection of
744 the feature plots for each antibody. After processing, the CITE-Seq and scRNA-seq data of Sort 4
745 were merged into the same Seurat object.

746 *Dataset integration and batch correction:* To verify that our different sorting strategies/definitions
747 of the uIEL in CB included the same or different cell types, a data integration of our three CB
748 samples was performed using the Seurat package. Therefore, the integration anchor strategy was
749 followed. Integration anchors were identified following the reciprocal PCA (rPCA) method instead
750 of the CCA method, due to the speed of the former method and the recommendation of the
751 developers that rPCA is more conservative and thus better equipped to handle cell populations that
752 have no matching type between samples. For this integration, integration features were first
753 selected with the "SelectIntegrationFeatures" function and subsequently the samples were scaled
754 ("ScaleData") and a PCA analysis ("RunPCA") was executed for each sample separately.
755 Afterwards the integration anchors were identified with the "FindIntegrationAnchors" function.
756 After integration with the "IntegrateData" function the data was scaled once again and a new PCA
757 analysis was performed for visualization and data exploration.

758 Subsequently, to identify the progeny of the thymic uIEL lineage in CB a second data integration
759 was performed with CB and PNT samples. Therefore, four PNT samples (TTA9, TTA10, TTA12
760 and TTA14) were selected from Park et al. [15]. The FASTQ-files of these four samples were
761 acquired and processed them as described under 'Preprocessing of the scRNA-seq and CITE-seq
762 data'. By processing the samples of both organs in the same way, the variation that needed to be
763 corrected in the subsequent data integration and batch correction step, could be limited. For the
764 latter, the R package Harmony was used [58]. Prior to the data integration with Harmony all samples
765 were merged into a combined Seurat object afterwards the following steps were performed: (i)
766 identification of highly variable genes/features ("FindVariableFeatures"), (ii) data scaling
767 ("ScaleData") and (iii) PCA analysis ("RunPCA"). The data integration was executed with the
768 function "RunHarmony", taking into account three sources of potential batch effects, namely
769 donor, sort and chemistry. The later was added since the PNT samples were processed with the v2
770 version of the 10X genomics scRNA-seq kit, while for the CB samples the v3 version was used.
771 Afterwards, the integrated object was processed for visualization and data exploration.

772 *Trajectory Analysis and SCENIC:* Based on the data integration of the PNT and CB samples, a
773 trajectory analysis could be performed to identify the progeny of the uIELs. For this trajectory
774 analysis, the TSCAN algorithm (version 1.28.0) was used [59]. To simplify the problem for the
775 algorithm, the populations of interest were selected prior to the actual analysis. For the
776 unconventional populations, these selected cell types were: CD8αα(II), MME$^+$ UTC, GNLY$^-$
777 MME$^-$ UTC, GNLY$^+$ UTC, GZMK$^+$ DN UTC, and IL32$^+$ UTC. While for the conventional
778 populations these were: CD8$^+$T, CTC1, CTC2, CTC3, CTC4, and CTC5. The first step in the
779 TSCAN analysis was the calculation of the centroid of each cluster with the "reducedDim"
780 function. These centroids were then connected with a minimum spanning tree (MST) in the
781 subsequent step using the "createClusterMST" function. The cells were subsequently mapped and
782 ordered along this MST to determine their pseudo-time point. This was done with the functions
783 "mapCellsToEdges" and "orderCells". To identify the genes that might play a role in this

784 differentiation process along the pseudo-time, a differential expression analysis with the tradeSeq
785 package (version 1.4.0) was performed [60]. Therefore, a generalized additive model (GAM) was
786 fitted to the data along the pseudotime and subsequently the associationTest was performed to
787 identify the differentially expressed genes.

788 To better identify which transcription factors are important or active in the different cell
789 populations, the data was analyzed with the SCENIC algorithm [61]. Note that for the actual analysis
790 the python implementation of this package, pySCENIC, was used. In a first step gene regulatory
791 networks and co-expression modules are generated by the GRNBoost 2 algorithm based on the
792 correlation between transcription factors and other genes. Subsequently regulons are predicted
793 based on known binding motifs provided by the Aerts lab. And in the final step the cellular
794 enrichment for the different regulons was calculated using the Aucell algorithm. The output of the
795 different steps was loaded into python to calculate the regulon specificity scores (RSS) for the
796 different cell populations through the pySCENIC package.

797 The regulon specificity score (RSS) was calculated from the regulon activity score (RAS) and lies
798 between 0 and 1. With a higher value for RSS indicating a higher specificity of that regulon in the
799 cell type compared to others. The UTC and CTC specific regulons were selected according to the
800 following strategy: a regulon that appeared in the top 5 of one of the UTC or CTC populations and
801 in the top 10 of one or more of the remaining UTC or CTC populations was selected for inclusion
802 in the plot. The RAS was calculated based on the rank of the expression value in the cell of all
803 genes involved in the regulon [62].

T cell expansion

805 The CellTrace proliferation assays were performed as previously described [36]. CTC and UTC
806 clones were generated by FACS sorting single cells and expanding them on irradiated allogenic
807 feeder cells, consisting of a mixture of 40 Gy irradiated peripheral blood mononuclear cells and
808 50 Gy irradiated JY cells. Cells were cultured in cIMDM, supplemented with 1 μg/mL
809 phytohemagglutinin (PHA, Sigma–Aldrich). IL-2 (5 ng/mL; Miltenyi, 130-097-748) was added
810 on day five and day ten. Cells were restimulated every 7 to 14 days. After 14-28 days, grown
811 clones were harvested and assessed via flow cytometry.

[51]Chromium Release Assay

813 Target cells (W6/32 or OKT3 hybridoma) were labelled with [51]Chromium (Perkin Elmer) for 90
814 min at 37 °C, washed and added at 10^3 cells per well to various ratios of effector T cells (CTC or
815 UTC population after overnight incubation with IL-15) in 96 well V-bottomed plates (NUNC,
816 Thermo Fisher Scientific). After 4 hours of co-incubation, the supernatant was harvested and
817 measured in a 1450 LSC & Luminescence Counter (Perkin Elmer). Specific lysis was calculated
818 as follows: (experimental release–spontaneous release)/(maximal release–spontaneous release) ×
819 100%.

Cytokine Production

821 To explore the secreted cytokine profile of the CB populations, Luminex High Performance
822 Assays (R&D Systems) were performed. The supernatant of both freshly sorted UTCs and UTC-
823 derived clones after 24 hours of stimulation with phorbol myristate acetate (PMA, 1 ng/mL;
824 Sigma-Aldrich, 16561-29-8) + ionomycin (0.5 µg/mL; Sigma-Aldrich, 56092-82-1) was assessed
825 with the Luminex Performance Human XL Cytokine Magnetic Panel 44-plex Fixed Panel (Bio-
826 Techne, LKTM014). Following, a custom mixed multiplex was used to determine the IFN-γ,

granzyme B, GM-CSF, MIP-1α, MIP-1β and IL-2 concentrations in the supernatant of both freshly sorted CTCs and UTCs after 24 hours of stimulation with PMA + ionomycin. All assays were performed conform the manufacturer's protocol and measured with the Bio-Plex 200 system (Biorad) and analysed with the Bio-Plex Manager software version 6.2. Levels below or above the detection level were set as the lower or upper detection level, respectively.

Statistical Analysis

Statistical analyses were performed in Prism version 9.3.1. (GraphPad Software, San Diego, CA, USA), using statistical tests as indicated in figure legends. Results were considered statistically significant when the p-value was less than 0.05.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [63] partner repository with the dataset identifier PXD033392. The sequencing data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [64] and will be made accessible upon publication through GEO Series accession number GSE201811.

Supplementary Material:

Fig. S1. RNA and protein expression profile by the PNT CD10$^+$ PD-1$^+$ population.

Fig. S2. Distinctive TCR repertoire of the PNT CD10$^+$ PD-1$^+$ population.

Fig. S3. Defining the T cell clusters in CB.

Fig. S4. Phenotyping the unconventional populations in CB.

Fig. S5. UTC phenotype and functionality after proliferation with IL-15.

Table S1. Significantly differentially expressed between the PD-1$^+$ and PD-1$^-$ population, in both PNT and CB

Table S2. Cell number in each identified cluster

Table S3. Top 10 differentially expressed genes for each identified cluster

Table S4. CITE-seq cell surface protein antibody panel

Table S5. Luminex Performance Human XL Cytokine Magnetic Panel 44-plex Fixed Panel

References

1. 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).

2. Pellicci, D. G., Koay, H. F. & Berzins, S. P. Thymic development of unconventional T cells: how NKT cells, MAIT cells and γδ T cells emerge. *Nat. Rev. Immunol.* **20**, 756–770 (2020).

3. McDonald, B. D., Jabri, B. & Bendelac, A. Diverse developmental pathways of intestinal intraepithelial lymphocytes. *Nat. Rev. Immunol.* **18**, 1–12 (2018).

4. Yamagata, T., Mathis, D. & Benoist, C. Self-reactivity in thymic double-positive cells commits cells to a CD8αα lineage with characteristics of innate immune cells. *Nat. Immunol. 2004 56* **5**, 597–605 (2004).

5. Mayans, S. *et al.* αβT cell receptors expressed by CD4-CD8αβ- intraepithelial T cells drive their fate into a unique lineage with unusual MHC reactivities. *Immunity* **41**, 207–218 (2014).

6. McDonald, B. D., Bunker, J. J., Ishizuka, I. E., Jabri, B. & Bendelac, A. Elevated T cell receptor signaling identifies a thymic precursor to the TCRαβ+CD4-CD8β- intraepithelial lymphocyte lineage. *Immunity* **41**, 219–229 (2014).

7. McDonald, B. D., Bunker, J. J., Erickson, S. A., Oh-Hora, M. & Bendelac, A. Crossreactive αβ T Cell Receptors Are the Predominant Targets of Thymocyte Negative Selection. *Immunity* **43**, 859–869 (2015).

8. Hummel, J. F. *et al.* Single-cell RNA-sequencing identifies the developmental trajectory of C-Myc-dependent NK1.1− T-bet+ intraepithelial lymphocyte precursors. *Mucosal Immunol.* **13**, 257–270 (2020).

9. Dadi, S. *et al.* Cancer Immunosurveillance by Tissue-Resident Innate Lymphoid Cells and Innate-like T Cells. *Cell* **164**, 365–377 (2016).

10. Wu, R. *et al.* Identification and Regulation of TCRαβ+CD8αα+ Intraepithelial Lymphocytes in Murine Oral Mucosa. *Front. Immunol.* **11**, 1702 (2020).

11. Mayassi, T. & Jabri, B. Human intraepithelial lymphocytes. *Mucosal Immunology* vol. 11 1281–1289 (2018).

907   12.   Verstichel, G. *et al.* The checkpoint for agonist selection precedes conventional selection
908   in human thymus. *Sci. Immunol.* **2**, (2017).

909   13.   Daley, S. R. *et al.* Cysteine and hydrophobic residues in CDR3 serve as distinct T-cell
910   self-reactivity indices. *J. Allergy Clin. Immunol.* **144**, 333–336 (2019).

911   14.   Wirasinha, R. C. *et al.* αβ T-cell receptors with a central CDR3 cysteine are enriched in
912   CD8αα intraepithelial lymphocytes and their thymic precursors. *Immunol. Cell Biol.* **96**,
913   553–561 (2018).

914   15.   Park, J. E. *et al.* A cell atlas of human thymic development defines T cell repertoire
915   formation. *Science.* **367**, eaay3224 (2020).

916   16.   Le, J. *et al.* Single-Cell RNA-Seq Mapping of Human Thymopoiesis Reveals Lineage
917   Specification Trajectories and a Commitment Spectrum in T Cell Development. *Immunity*
918   **52**, 1105-1118.e9 (2020).

919   17.   Kondo, M. *et al.* SATB1 Plays a Critical Role in Establishment of Immune Tolerance. *J.*
920   *Immunol.* **196**, 563–572 (2016).

921   18.   Wei, H. *et al.* Cutting Edge: Foxp1 Controls Naive CD8 + T Cell Quiescence by
922   Simultaneously Repressing Key Pathways in Cellular Metabolism and Cell Cycle
923   Progression. *J. Immunol.* **196**, 3537–3541 (2016).

924   19.   Kisielow, J., Tortola, L., Weber, J., Karjalainen, K. & Kopf, M. Evidence for the
925   divergence of innate and adaptive T-cell precursors before commitment to the αβ and γδ
926   lineages. *Blood* **118**, 6591–6600 (2011).

927   20.   Stadinski, B. D. *et al.* Hydrophobic CDR3 residues promote the development of self-
928   reactive T cells. *Nat. Immunol.* **17**, 946–955 (2016).

929   21.   Lagattuta, K. A. *et al.* Repertoire analyses reveal T cell antigen receptor sequence features
930   that influence T cell fate. *Nat. Immunol. 2022 233* **23**, 446–457 (2022).

931   22.   Košmrlj, A., Jha, A. K., Huseby, E. S., Kardar, M. & Chakraborty, A. K. How the thymus
932   designs antigen-specific and self-tolerant T cell receptor sequences. *Proc. Natl. Acad. Sci.*
933   *U. S. A.* **105**, 16671–16676 (2008).

934   23.   Logunova, N. N. *et al.* MHC-II alleles shape the CDR3 repertoires of conventional and
935   regulatory naïve CD4+ T cells. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 13659–13669 (2020).

936   24.   Monk, J. M. *et al.* Antigen receptor repertoire profiling from RNA-seq data. *Nat.*
937   *Biotechnol. 2017 3510* **35**, 908–911 (2017).

938   25.   Pellicci, D. G. *et al.* The molecular bases of δ/αβ T cell–mediated antigen recognition. *J.*
939   *Exp. Med.* **211**, 2599–2615 (2014).

940   26.   Ruscher, R. *et al.* Intestinal CD8αα IELs derived from two distinct thymic precursors have
941   staggered ontogeny. *J. Exp. Med.* **217**, (2020).

942   27.   Kurd, N. S. *et al.* Factors that influence the thymic selection of CD8αα intraepithelial
943   lymphocytes. *Mucosal Immunol.* 1–12 (2020) doi:10.1038/s41385-020-0295-5.

944   28.   Mamonkin, M. *et al.* Differential roles of KLF4 in the development and differentiation of
945   CD8+ T cells. *Immunol. Lett.* **156**, 94–101 (2013).

946 29. Hao, Z. *et al.* K48-linked KLF4 ubiquitination by E3 ligase Mule controls T-cell
947   proliferation and cell cycle progression. *Nat. Commun. 2017 81* **8**, 1–14 (2017).

948 30. Bui, T. M., Wiesolek, H. L. & Sumagin, R. ICAM-1: A master regulator of cellular
949   responses in inflammation, injury resolution, and tumorigenesis. *J. Leukoc. Biol.* **108**,
950   787–799 (2020).

951 31. Speiser, D. E. *et al.* The Activatory Receptor 2B4 Is Expressed In Vivo by Human CD8 +
952   Effector αβ T Cells. *J. Immunol.* **167**, 6165–6170 (2001).

953 32. Miller, C. H. *et al.* Eomes identifies thymic precursors of self-specific memory-phenotype
954   CD8+ T cells. *Nat. Immunol. 2020 215* **21**, 567–577 (2020).

955 33. Drobek, A. *et al.* Strong homeostatic TCR signals induce formation of self-tolerant virtual
956   memory CD8 T cells. *EMBO J.* **37**, e98518 (2018).

957 34. White, J. T. *et al.* Virtual memory T cells develop and mediate bystander protective
958   immunity in an IL-15-dependent manner. *Nat. Commun. 2016 71* **7**, 1–13 (2016).

959 35. Jacomet, F. *et al.* Evidence for eomesodermin-expressing innate-like CD8+
960   KIR/NKG2A+ T cells in human adults and cord blood samples. *Eur. J. Immunol.* **45**,
961   1926–1933 (2015).

962 36. Billiet, L. *et al.* Human Thymic CD10+ PD-1+ Intraepithelial Lymphocyte Precursors
963   Acquire Interleukin-15 Responsiveness at the CD1a– CD95+ CD28– CCR7–
964   Developmental Stage. *Int. J. Mol. Sci.* **21**, (2020).

965 37. Rodríguez-Rodríguez, N. *et al.* Programmed Cell Death 1 and Helios Distinguish TCR-αβ
966   + Double-Negative (CD4 − CD8 −) T Cells That Derive from Self-Reactive CD8 T Cells.
967   *J. Immunol.* **194**, 4207–4214 (2015).

968 38. Schattgen, S. A. *et al.* Integrating T cell receptor sequences and transcriptional profiles by
969   clonotype neighbor graph analysis (CoNGA). *Nat. Biotechnol.* **40**, 54–63 (2022).

970 39. Li, J. *et al.* KIR + CD8 + T cells suppress pathogenic T cells and are active in autoimmune
971   diseases and COVID-19. *Science (80-.).* **376**, eabi9591 (2022).

972 40. Cantor, H. & Kim, H. J. A new chapter in the CD8 T reg story. *J. Exp. Med.* **218**, (2020).

973 41. Kim, H. J. *et al.* Stable inhibitory activity of regulatory T cells requires the transcription
974   factor Helios. *Science (80-.).* **350**, 334–339 (2015).

975 42. Mishra, S. *et al.* TGF-β and Eomes control the homeostasis of CD8+regulatory T cells. *J.*
976   *Exp. Med.* **218**, (2021).

977 43. Patterson, S. J. *et al.* T regulatory cell chemokine production mediates pathogenic T cell
978   attraction and suppression. *J. Clin. Invest.* **126**, 1039–1051 (2016).

979 44. Lee, S. T., Georgiev, H., Breed, E. R., Ruscher, R. & Hogquist, K. A. MHC Class I on
980   murine hematopoietic APC selects Type A IEL precursors in the thymus. *Eur. J. Immunol.*
981   **51**, 1080–1088 (2021).

982 45. Sullivan, B. A., Kraj, P., Weber, D. A., Ignatowicz, L. & Jensen, P. E. Positive Selection
983   of a Qa-1-Restricted T Cell Receptor with Specificity for Insulin. *Immunity* **17**, 95–105
984   (2002).

46. Sheng, H. *et al.* Distinct PLZF + CD8αα + Unconventional T Cells Enriched in Liver Use a Cytotoxic Mechanism to Limit Autoimmunity. *J. Immunol.* **203**, 2150–2162 (2019).

47. Zhang, Y. Y. *et al.* CD8αα + T cells exert a pro-inflammatory role in patients with psoriasis. *Ski. Heal. Dis.* **1**, e64 (2021).

48. Zhu, J. *et al.* Immune surveillance by CD8αα+ skin-resident T cells in human herpes virus infection. *Nature* **497**, 494–497 (2013).

49. Brandt, D. & Hedrich, C. M. TCRαβ + CD3 + CD4 - CD8 - (double negative) T cells in autoimmunity. *Autoimmun. Rev.* **17**, 422–430 (2018).

50. Yang, L. *et al.* Transcriptome landscape of double negative T cells by single-cell RNA sequencing. *J. Autoimmun.* **121**, 102653 (2021).

51. Vasic, D. *et al.* Allogeneic double-negative CAR-T cells inhibit tumor growth without off-tumor toxicities. *Sci. Immunol.* **7**, 3642 (2022).

52. Van Caeneghem, Y. *et al.* Antigen receptor-redirected T cells derived from hematopoietic precursor cells lack expression of the endogenous TCR/CD3 receptor and exhibit specific antitumor capacities. *Oncoimmunology* **6**, e1283460 (2017).

53. Bolotin, D. A. *et al.* MiXCR: Software for comprehensive adaptive immunity profiling. *Nature Methods* vol. 12 380–381 (2015).

54. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).

55. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

56. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

57. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

58. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

59. Ji, Z. & Ji, H. TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* **44**, e117 (2016).

60. Van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1–13 (2020).

61. Aibar, S. *et al.* SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

62. Ma, A. *et al.* IRIS3: Integrated cell-type-specific regulon inference server from single-cell RNA-Seq. *Nucleic Acids Res.* **48**, W275–W286 (2020).

63. Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, 543–552 (2022).

64. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene

1024   expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210
1025   (2002).

1026

Abbreviations

1028   CB, cord blood
1029   CTC, conventional T cell
1030   DN, double negative
1031   GSEA, gene set enrichment analysis
1032   IEL, intraepithelial lymphocyte
1033   IELp, intraepithelial lymphocyte precursor
1034   KIR, killer Ig-like receptor
1035   MAIT, mucosal-associated invariant T cell
1036   NKT, natural killer T cell
1037   PNT, postnatal thymus
1038   SP, single positive
1039   TCR, T cell receptor
1040   TF, transcription factor
1041 T_{VM}, virtual memory T cell
1042   uIEL, unconventional intraepithelial lymphocyte
1043   UMAP, uniform manifold approximation and projection
1044   UTC, unconventional T cell

1045