

Invariant $\gamma\delta$ TCR natural killer-like effector T cells in the naked mole-rat

Received: 20 July 2023

Accepted: 3 May 2024

Published online: 18 May 2024

 Check for updates

Guillem Sanchez Sanchez ^{1,2,3,4}, Stephan Emmrich ⁵, Maria Georga ⁶,
Ariadni Papadaki ⁶, Sofia Kossida ⁶, Andrei Seluanov ^{5,7},
Vera Gorbunova^{5,7} & David Vermijlen ^{1,2,3,4} 

The naked mole-rat (*Heterocephalus glaber*) is a long-lived rodent species showing resistance to the development of cancer. Although naked mole-rats have been reported to lack natural killer (NK) cells, $\gamma\delta$ T cell-based immunity has been suggested in this species, which could represent an important arm of the immune system for antitumor responses. Here, we investigate the biology of these unconventional T cells in peripheral tissues (blood, spleen) and thymus of the naked mole-rat at different ages by TCR repertoire profiling and single-cell gene expression analysis. Using our own TCR annotation in the naked mole-rat genome, we report that the $\gamma\delta$ TCR repertoire is dominated by a public invariant V γ 4-2/V δ 1-4 TCR, containing the complementary-determining-region-3 (CDR3) γ CTYWDSNYAKKLF / CDR3 δ CALWELRTGGITAQLVF that are likely generated by short-homology-repeat-driven DNA rearrangements. This invariant TCR is specifically found in $\gamma\delta$ T cells expressing genes associated with NK cytotoxicity and is generated in both the thoracic and cervical thymus of the naked mole-rat until adult life. Our results indicate that invariant V γ 4-2/V δ 1-4 NK-like effector T cells in the naked mole-rat can contribute to tumor immunosurveillance by $\gamma\delta$ TCR-mediated recognition of a common molecular signal.

The naked mole-rat (*Heterocephalus glaber*) is an eusocial mammal, highly specialized for life underground through sensory and ecophysiological adaptations, but it has also been used in biomedical studies to investigate for example pain, hypoxia, cardiac function, and cancer^{1–3}. Naked mole-rats are highly resistant to cancer: thousands of individual animals have been monitored over decades in biomedical research laboratories, as well as in zoo colonies, and only six cases of tumors have been found¹. Cell-intrinsic mechanisms, for example driven by the production of high molecular mass hyaluronan, can contribute to this resistance against cancer cell development^{1,2}. However, it has been suggested more recently that cell-extrinsic factors such as immune cells can play major roles in the naked mole-rat cancer

resistance^{3–5}. It is therefore intriguing that the naked mole-rat lack canonical natural killer (NK) cells, a lymphocyte population with potent anti-tumor activities⁶. In contrast, the naked mole-rat has been shown to possess large ectopic cervical thymi in addition to the canonical thoracic thymus, organs where T cells are generated⁷. Furthermore, naked mole-rat thymi do not show involution upon aging and there are indications that the naked mole-rat T cell compartment shows a bias towards $\gamma\delta$ T cells⁷.

$\gamma\delta$ T cells are the “third” type of lymphocytes, besides $\alpha\beta$ T cells and B cells, that can rearrange genes at the DNA level in order to generate variable antigen receptors^{8,9}. These three cell lineages have been conserved seemingly since the emergence of jawed vertebrates,

¹Department of Pharmacotherapy and Pharmaceutics, Université Libre de Bruxelles (ULB), Brussels, Belgium. ²Institute for Medical Immunology, Université Libre de Bruxelles (ULB), Gosselies, Belgium. ³ULB Center for Research in Immunology (U-CRI), Université Libre de Bruxelles (ULB), Gosselies, Belgium. ⁴WELBIO Department, WEL Research Institute, Wavre, Belgium. ⁵Department of Biology, University of Rochester, Rochester, NY, USA. ⁶IMGT®, the international Immunogenetics information system®, Institut de Génétique Humaine (IGH), Centre National de la Recherche Scientifique (CNRS), Université de Montpellier (UM), Montpellier, France. ⁷Department of Medicine, University of Rochester Medical Center and Medicine, University of Rochester, Rochester, NY, USA.

 e-mail: David.Vermijlen@ulb.be

with the notable exception of squamate reptiles that lost $\gamma\delta$ T cells¹⁰, while a similar tripartite subdivision exists even in jawless vertebrates, such as lamprey and hagfish^{11,12}. The rearrangements of the TRG (gamma chain) and TRD (delta chain) loci take place in the thymus during the development of $\gamma\delta$ T cells from a common $\alpha\beta/\gamma\delta$ T cell precursor¹³. During TRG or TRD rearrangement, a V (variable), D (diversity; only for TRD), and J (joining) gene are joined to form a final T cell receptor (TCR) chain. The variability created during the V(D)J recombination is significantly enhanced by the junctional diversity, which includes the introduction of random nucleotides (N additions), a process mediated by the enzyme Terminal deoxynucleotidyl Transferase (TdT). The most variable domain, that includes these N additions and is usually accountable for antigen recognition, is found in the complementarity determining region 3 (CDR3) of the TRG and TRD chains. The pairing of a single TRG with a TRD chain at the protein level results in the final TCR expressed on the surface of the $\gamma\delta$ T cell. Based on their TCR repertoire, mouse and human $\gamma\delta$ T cells can be divided into innate subsets that possess all the same/similar invariant CDR3 and adaptive subsets that possess a variable TCR repertoire⁸. The innate $\gamma\delta$ T cells are typically produced in fetal waves before birth^{14,15}. Notwithstanding some examples of MHC-restricted $\gamma\delta$ TCRs¹⁶, the $\gamma\delta$ TCR is usually neither limited by MHC restriction nor primarily focused on processed antigens⁹. Still, a general rule of ligand recognition by the $\gamma\delta$ TCR is hampered by the lack of ligand knowledge for the majority of $\gamma\delta$ TCRs^{17–19}. However, based on the known $\gamma\delta$ TCR ligands, it appears that they have the potential to recognize diverse, largely host-derived, molecular entities^{17,18,20}. Of note, in recent years an emerging role for butyrophilin (BTN) and butyrophilin-like (BTNL) molecules became evident in the TCR-dependent regulation of $\gamma\delta$ T cells, involving TRGV regions outside the CDR3^{9,21}.

$\gamma\delta$ T cells can perform cancer immunosurveillance: for example, the mouse invariant V γ 5V δ 1 TCR senses cell normality in the skin epidermis^{22,23} and the (semi)invariant human V γ 9V δ 2 TCR senses the increased production of phosphorylated metabolites (known as phosphoantigens) in cancer cells²⁴. As a consequence, $\gamma\delta$ T cells are now being explored as a viable and promising approach for cancer immunotherapy in human, with the V γ 9V δ 2 T cell subset as a main target^{25–27}.

Since $\gamma\delta$ T cells are important in cancer immunosurveillance and thus could contribute to the natural cancer resistance observed in the naked mole-rat, we investigated the presence of these unconventional T cells in the naked mole-rat and characterized them in detail, comparing with conventional $\alpha\beta$ T cells, in peripheral tissues (blood, spleen) and in both thymus types. This required the annotation of the four TR loci (TRA, TRB, TRG and TRD) in *Heterocephalus glaber*. We found a striking prevalence, at a wide age range, of a $\gamma\delta$ T cell population that possess an invariant $\gamma\delta$ TCR and an NK-like cytotoxic effector phenotype.

Results

Annotation of naked mole-rat (*Heterocephalus glaber*) TR loci

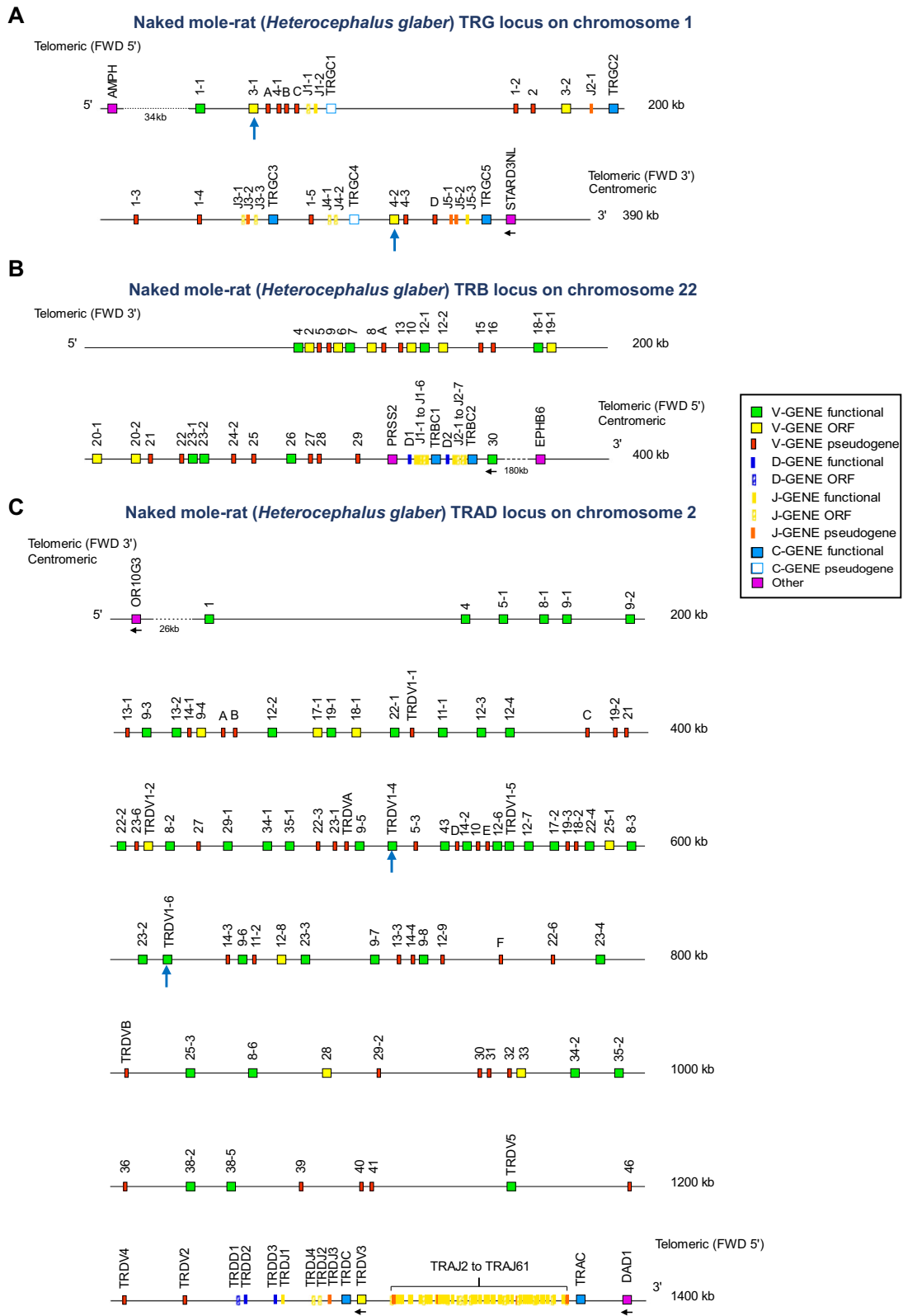
To determine the type of expressed naked mole-rat TCRs by high-throughput sequencing, we first annotated the TR loci (TRG, TRAD, and TRB) of *Heterocephalus glaber* (Fig. 1). While annotations were conducted for the TRG and TRB loci in both the maternal and paternal chromosomal assemblies, the annotation for the TRAD locus was exclusively carried out in the maternal assembly. This decision was prompted by the presence of gaps in the paternal assembly of the TRAD locus, indicative of lower quality sequencing in that genome region. Of note, there were no differences observed in the presence or order of TR genes between the annotated maternal and paternal haplotypes for TRB and TRG loci. The evaluation of the accuracy of the assemblies was performed through the detailed revision of the estimated number of variable, diversity, joining and constant genes at the locus, along with their

organization, in comparison with the data of the most closely related evolutionary species.

The TRG locus of the naked mole-rat is located on chromosome 1, spanning -333 kilobases (kb) from 10 kb upstream of the most 5' gene, *TRGVI-1*, to 10 kb downstream of the most 3' gene, *TRGC5*. The delimitation of the locus was based on the investigation of the “IMGT bornes”, which are coding genes, other than TR, that remain conserved across different species and are located upstream of the first or downstream of the last gene of the locus, respectively²⁸. In both maternal and paternal assemblies, the 3' borne of TRG locus (*STARD3NL*) is detected 5 kb downstream of *TRGC5*. In the maternal assembly only, the 5' borne of TRG (*AMPH*) is present, 34 kb upstream the *TRGVI-1* (Fig. 1A). Within this locus 31 TRG genes were detected, consisting of 15 TRG variable genes (TRGV) (1 F, 3 ORF, 11 P), 11 TRG joining genes (TRGJ) (1 F, 6 ORF, 4 P), and 5 TRG constant genes (TRGC) (3 F, 2 P) (Fig. 1A). All the genes were present in both maternal and paternal assemblies and polymorphisms between the two assemblies were detected in 6 genes: 3 TRGV genes (*TRGVI-3*, *TRGVI-4* and *TRGVD*), 1 TRGJ gene (*TRGJ3-2*) and 2 TRGC genes (*TRGC2* and *TRGC5*). A gene is assigned ORF (open reading frame), rather than F (functional), when the coding region has an open reading frame without stop codons (like in a F) but shows (unlike the F designation) gene alterations in the recombination signals (compared to the canonical recombination signal sequences), deletions of more than 3 amino acids in the V-REGION and/or changes of conserved amino acids that might lead to incorrect protein folding. It is important to note that the annotation of a gene as ORF (versus F) is a dynamic process, particularly when applied to new species, where these changes may be unique and characteristic of a specific species. A gene is assigned P (pseudogene) when the coding region has stop codon(s) and/or frameshift mutation(s) and/or a mutation affecting the initiation codon. The naked mole-rat's TRG genes are organized into 5 V-J-C-CLUSTERS or “cassettes”. The number of TRGV (1–6) and TRGJ (1–3) varies in each cluster, while all clusters contain one TRGC (Fig. 1A). This organization into different clusters or cassettes is also observed in the mouse TRG locus but not in the human TRG locus.

The TRB locus of the naked mole-rat is located on chromosome 22 and extends over a distance of 283 kilobases (kb), ranging from 10 kb upstream of the most 5' gene in the locus (*TRBV4*) to 10 kb downstream of the most 3' gene (*TRBV30*) (Fig. 1B). The 3' borne, *EPHB6*, is located 180 kb downstream of *TRBV30*. The TRB locus of naked mole-rat consists of 29 TRBV categorized into 26 TRBV subgroups (8 F, 8 ORF and 13 P in the maternal assembly and 9 F, 8 ORF and 12 P in the paternal assembly), 2 TRBD belonging to 2 TRBD sets (2 F), 13 TRBJ (6 F and 7 ORF) belonging to 2 TRBJ sets, and 2 TRBC (2 F) (Fig. 1B). The genes were present in both maternal and paternal assemblies, while polymorphic variations were observed in 5 genes across the two assemblies: 2 TRBV genes (*TRBV9* and *TRBV27*), 1 TRBJ gene (*TRBJ1-6*) and 2 TRBC genes (*TRBC1* and *TRBC2*).

The TRAD locus, thus containing the genes for both the TRA and TRD of $\alpha\beta$ and $\gamma\delta$ T cells respectively, of the naked mole-rat is located on chromosome 2 and spans 1368 kilobases (kb), ranging from 10 kb upstream of the most 5' gene (*TRAV1*), to 10 kb downstream of the most 3' gene (*TRAC*) (Fig. 1C). The 5' borne (*ORIOG3*) and 3' borne (*DADI*) have been identified 26 bp upstream of *TRAV1* and 20 kb downstream of *TRAC*, respectively (Fig. 1C). In the TRAD locus, 79 TRAV genes (39 F, 6 ORF and 34 P) were detected belonging to 39 TRAV subgroups. The naked mole-rat TRAD locus consists also of 11 TRDV genes (4 F, 2 ORF, and 5 P), 3 TRDD (2 F and 1 ORF), 4 TRDJ (1 F, 2 ORF, and 1 P), and 1 TRDC (F) (Fig. 1C). The TRDV genes are localized in 2 V-CLUSTERS: the 5' V-CLUSTER, in 5' of the TRDD, TRDJ, and TRDC, includes 10 TRDV genes and the 3' V-CLUSTER, downstream of the TRDC in inverted orientation, includes 1 TRDV. The TRDD, TRDJ, and TRDC genes are localized in 1 D-J-C-CLUSTER between the 5' V-CLUSTER and the 3' V-CLUSTER (Fig. 1C).



Comparison of naked mole-rat TRGV and TRDV genes with human and mouse

To gain a deeper understanding of the naked mole-rat’s TR loci and their variable genes, an unrooted phylogenetic analysis was performed, employing a holistic approach that incorporated TRGV, TRDV, TRAV, and TRBV nucleotide sequences from *Heterocephalus glaber* along with the well-studied *Homo sapiens* and *Mus musculus* (Fig. 2;

Supplementary Fig. 1). Within these trees, genes belonging to the same subgroup cluster together under the same branch, often with a corresponding human gene. The subgroup classification of the naked mole-rats’ TR loci is primarily based on human subgroups, with a few exceptions such as *TRBVI*, *TRDV4*, and *TRDV5*. These exceptions arise from their significantly low genetic identity percentages, which hinder correspondence with human subgroups. In terms of TRGV subgroups,

Fig. 1 | Locus representation of the *Heterocephalus glaber* TR loci of the assembly Naked mole-rat maternal (GCA_944319715.1). **A** Naked mole-rat (*Heterocephalus glaber*) TRG locus on chromosome 1; the orientation of the TRG locus is forward (FWD). **B** Naked mole-rat (*Heterocephalus glaber*) TRB locus on chromosome 22; the orientation of the TRB locus is reverse (REV). **C** Naked mole-rat (*Heterocephalus glaber*) TRA/TRD locus on chromosome 2; the orientation of the TRA/TRD locus is reverse (REV). **A–C** The figures display the position, nomenclature and functionality of the TR genes according to the IMGT nomenclature⁷³. The proposed nomenclature is considered provisional based on future data being available and involvement of IUIS at the appropriate time/level. The dotted line

indicates the distance in kb between the locus and its IMGT bornes²⁸ (5' borne and the most 5' gene in the locus and/or 3' borne and the most 3' gene in the locus) on the top and bottom lines, respectively. These distances are not represented at scale and are not included in the numbers displayed at the right ends of these two lines. The telomeric and centromeric indications refer to the position of the locus in the chromosome. There are two telomeric ends referenced within the locus (one at the 5' and one at the 3' of the locus), with the centromeric indication added to its nearest end. Finally, the black arrows indicate an inverse transcriptional orientation in the locus while the blue arrows indicate the position in the TRG and TRD locus of V genes that are highlighted in the current study.

both the 11 variable genes of the naked mole-rat and the 13 variable genes of humans are classified in four subgroups. The most prevalent subgroup in both species is TRGV1, which consists of 5 variable genes in the naked-mole rat and 10 variable genes in humans (containing for example *TRGV8*). In mouse, no direct correspondence between its TRGV subgroups and human has been identified so far. However, this phylogenetic analysis reveals that subgroup TRGV1 (containing *Trgv1*, *Trgv2*, and *Trgv3*) and subgroup TRGV2 (containing *Trgv4*) of mouse are clustered together in the same branch with subgroup TRGV4 (containing the non-functional *TRGVII*) of human (Fig. 2A).

Interestingly, the naked mole-rat possess TRGV (*TRGV2*) and TRDV (*TRDV2*) genes that show high identity with the human *TRGV9* and *TRDV2* genes respectively (Fig. 2), while the mouse TRGV or TRDV genes do not show significant identity with these human genes. These human *TRGV9* and *TRDV2* genes compose the variable domains of the human semi-invariant V γ 9V δ 2 TCR expressed by the phosphoantigen-reactive $\gamma\delta$ T cells, the most prevalent $\gamma\delta$ T cell subset in human peripheral blood and which are regarded as innate T cells⁸. Of note, both the naked mole-rat *TRGV9* and *TRDV2* genes are pseudogenes (Fig. 1A, C).

In the mouse, *Trdv4* (often referred to as “V δ 1”) is a main gene used in the invariant $\gamma\delta$ TCR of mouse $\gamma\delta$ T cells⁸. As for the human *TRDV2*, also for this “mouse invariant” TRDV gene a close orthologue exists in the naked mole-rat, *TRDV4* (Fig. 2B), but which is again annotated as a pseudogene (Fig. 1C). For the mouse TRGV genes used in invariant $\gamma\delta$ TCRs (*Trgv5* and *Trgv6*), no clear orthologue exists in the naked mole-rat (Fig. 2A).

Thus overall, we identified naked mole-rat TRGV and TRDV genes showing high identity with mouse/human genes that are used in mouse and human invariant $\gamma\delta$ TCRs. However, these naked mole-rat genes are pseudogenes, questioning the presence of conserved invariant naked mole-rat $\gamma\delta$ TCRs.

Selective conservation of BTN(L) members in the naked mole-rat

In recent years, it has become increasingly clear that members of the BTN(L) family are crucial for $\gamma\delta$ T cell biology, especially for those $\gamma\delta$ T cell subsets with (semi)invariant $\gamma\delta$ TCRs⁸. Mouse *Skint1/2* and *Btnl1/Btnl6* are important for the development of skin V γ 5⁺ and intestinal V γ 7⁺ $\gamma\delta$ T cells respectively^{9,29–31}, while human *BTN3A1/BTN2A1* have been shown to be crucial for phosphoantigen-reactivity^{32–35}. As naked mole-rat TRGV/TRDV genes with high identity to “invariant” human/mouse genes are pseudogenes (Fig. 2), we wondered whether the naked mole-rat lacks conservation of mouse/human BNT(L) members. Thus, we reconstructed an unrooted phylogenetic tree analysis of naked mole-rat BTN(L) amino acid sequences with their counterparts in mouse and human (Fig. 3). This showed that that some BTN(L) members were conserved across all three species (*BTNL1A1*, *BTNL9*), while others were only shared by pairs of species (*BTNL3/BTNL8* between human and naked mole-rat, *BTNL2* between mouse and human and *BTNL10* between mouse and naked mole-rat). Furthermore, members of the *BTN3* family were specific to human while members related to *Skint1* were specific to mouse species (Fig. 3). Thus, the naked mole-rat may possess a selected conservation of mouse/human BTN(L) members in its genome.

The $\gamma\delta$ TCR repertoire of the naked mole-rat is highly oligoclonal

To obtain an unbiased insight into the re-arranged and expressed TCR repertoire of naked mole-rat, we conducted TCR sequencing (TCR-seq) experiments targeting the CDR3 of TRA, TRB, TRG, and TRD chains on RNA derived from blood samples from 3 neonates, all from the same litter, and blood and spleen samples from 7 adults ranging from 3 to 11 years old, originating from 4 different litters. Analysis of general TCR repertoire properties revealed lower levels of N additions and shorter CDR3 in neonatal TRB, TRG, and TRD repertoires compared to adult repertoires (Fig. 4A, Supplementary Fig. 2A, B).

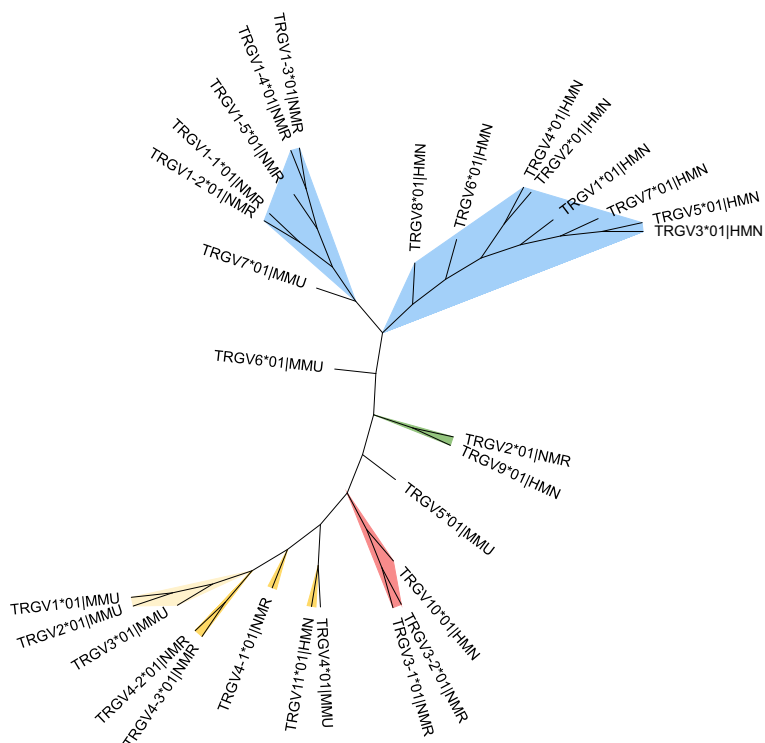
Interestingly, there was a huge bias, both in the neonatal and adult group, towards the usage of the *TRGV4-2* gene in the TRG repertoire (Fig. 4B, top panel, Supplementary Table 2). The neonatal group showed the strongest bias towards the *TRGV4-2* usage and thus the adult animals had a tendency to have more (though not statistically significant) expression of *TRGV1-1*, *TRGV3-1* and *TRGV3-2* compared to the neonatal group (Fig. 4B, top row). The TRD repertoire exhibited a strong preference for the utilization of *TRDV1-4* and *TRDV1-6* genes, with *TRDV1-4* being more prominently enriched in neonates and *TRDV1-6* in adults (Fig. 4B bottom row, Supplementary Table 2). Differences in junctional diversity, CDR3 nucleotide length and NDN size between neonates and adults had the tendency to be more prominent in *TRDV1-6*-containing than in *TRDV4-1*-containing sequences (Fig. 4C). While a similar strong overall bias in V gene usage could not be observed in the TRA and TRB repertoire, some differences were found between the neonatal and adult groups (Supplementary Fig. 2C, D).

Importantly, examination of repertoire clonality revealed massive enrichment in the TRG repertoire and an oligoclonal pattern in the TRD repertoire, with the most abundant clones dominating a significant portion of the total repertoire (Fig. 4D, E; Supplementary Fig. 3). The TRB repertoire exhibited high polyclonality, while the TRA repertoire only displayed oligoclonality in the neonatal group, albeit at lower levels compared to the TRG and TRD repertoires (Fig. 4D, E; Supplementary Fig. 3). Note that an increase in clonality, indicated by a higher frequency of clones with lower abundance in the total repertoire, was observed in the adult group from all repertoires (Fig. 4E), possibly related to a limited expression of TdT in naked mole-rats at birth. Consistent with reduced junctional diversity during the neonatal period in the naked mole-rat, the levels of TRB, TRG, and TRD repertoire sharing (as defined by the F value) within neonatal samples were higher compared to their adult counterparts (Fig. 4F). The higher TRD sharing was due to the enriched expression of *TRDV1-4*-containing CDR3 sequences (Fig. 4F bottom row). When performing neonate vs. adult paired comparisons, the neonatal group consistently showed higher sharing levels in all repertoires (Supplementary Fig. 4A).

Invariant *TRGV4-2* and *TRDV1-4* CDR3 sequences dominate the naked mole-rat TRG and TRD repertoire

Because of the high bias in TRGV and TRDV usage in the naked mole-rat, we wondered whether this was associated with particular patterns at the level of TCR arrangement. Therefore, we conducted an analysis

A



B

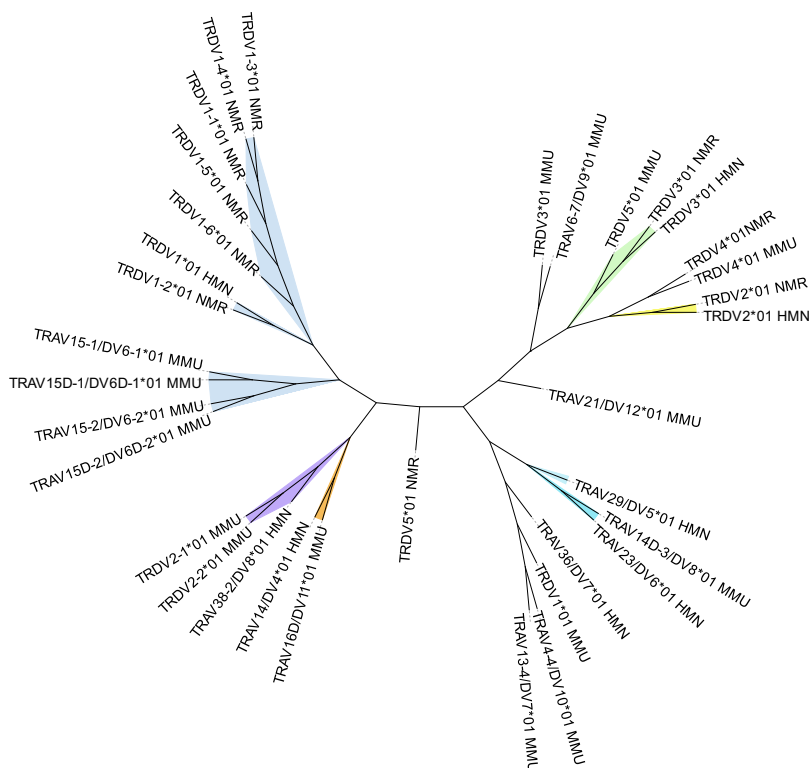


Fig. 2 | Unrooted phylogenetic trees of TRGV (A) and TRDV (B) genes in *Homo sapiens*, *Mus musculus*, and *Heterocephalus glaber*. The trees were constructed using the Maximum Likelihood method in MEGA software⁷⁹. The trees are based on nucleotide sequence alignments, generated with representatives of V-REGION sequences from each subgroup using MAFFT⁸⁰. Visualization of the trees was carried out using iTOL v6⁸¹, with distinct colors representing the different human

subgroups and their orthologs. For instance, under (A) the TRGV1 subgroup includes the human *TRGV8* gene and the naked mole-rat *TRGV4-2* gene; under (B), the TRDV1 subgroup includes the human *TRDVI* gene and the naked mole-rat *TRDVI-4* gene. Species abbreviations are as follows: HMN *Homo sapiens*, MMU *Mus musculus*, NMR *Heterocephalus glaber*. Source data are provided as a Source Data file.

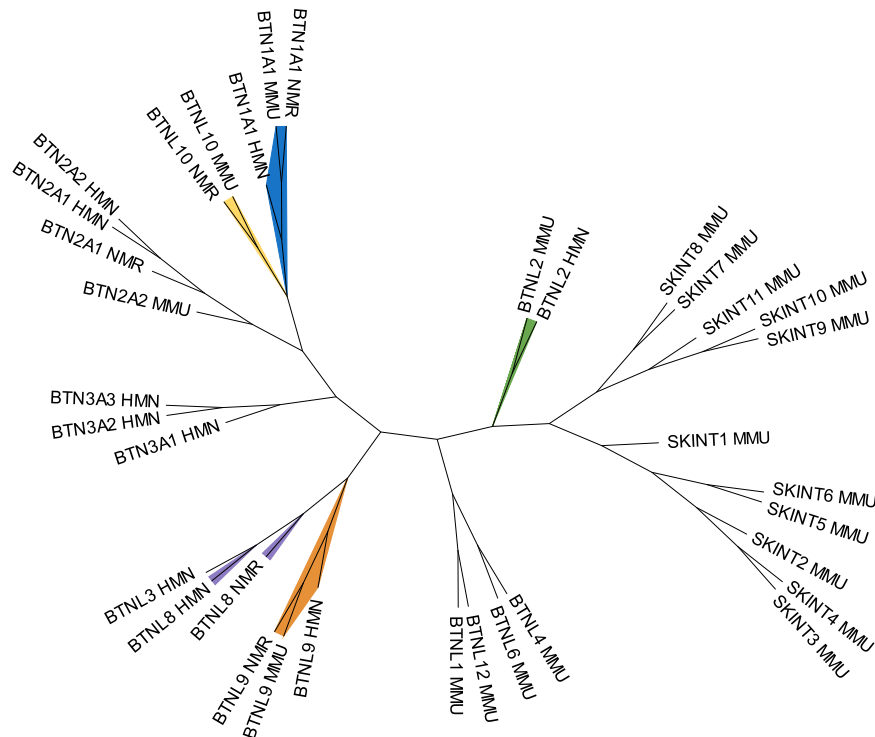


Fig. 3 | Unrooted phylogenetic trees of BTN(L) molecules from *Homo sapiens*, *Mus musculus*, and *Heterocephalus glaber* species. Maximum Likelihood tree of BTN(L) proteins was constructed utilizing MEGA software⁷⁹, based on amino acid MAFFT⁸⁰ sequence alignments of the BTN(L) and SKINT proteins obtained from NCBI. Protein sequences were used as there was enough information to reconstruct the phylogenetic tree. A single representative sequence from each protein family

was selected for inclusion in the analysis. The tree was visualized with iTOL v6⁸¹, and manually assigned colors indicate orthologous proteins, determined using the NCBI ortholog database. For example, the BTNL9 proteins from HMN, MMU, and NMR share a branch (colored orange) as orthologs. Species abbreviations are as follows: HMN *Homo sapiens*, MMU *Mus musculus*, NMR *Heterocephalus glaber*. Source data are provided as a Source Data file.

of CDR3 sequences. Remarkably, the TRG repertoire was overwhelmingly dominated by a highly prevalent public CDR3 sequence (that is: shared between different individuals), which was present in all the samples: CTYWDSNYAKKLF (Fig. 5A left panel; Supplementary Fig. 4B; Supplementary Fig. 5A).

This sequence, composed of a *TRGV4-2* gene and a *TRGJ5-3* gene, was germline-encoded without junctional diversity (Fig. 5B). Intriguingly, when aligning this sequence with invariant CDR3 γ sequences from mice and humans⁸, we found that the CTYWDSNYAKKLF sequence shared the same length and conserved amino acid intrinsic physicochemical properties as the human fetal-enriched invariant *TRGV8* CDR3 sequence³⁶ and the mouse invariant fetal-restricted *TRGV5* and *TRGV6* CDR3 sequence^{37,38} (Fig. 5C).

Similarly, the TRDV1-4 repertoire exhibited high levels of public CDR3 sequences, accounting for more than 10% of the repertoire, while the TRDV1-6 repertoire displayed more modest values (Supplementary Fig. 4B). When comparing publicity between neonates and adults, the repertoire space occupied by these highly public sequences tended to be higher in neonates for TRDV1-4 and TRDV1-6 sequences, with a smaller difference observed in the case of TRG sequences (Supplementary Fig. 4B). The TRDV1-4 repertoire was dominated by a single CDR3 sequence: CALWELRTGGITAQLVF (Fig. 5A right panel; Supplementary Fig. 5B). This sequence consisted of *TRDV1-4*, *TRDD2*, *TRDD3*, and *TRDJ2* genes (Fig. 5D).

Both the highly prevalent invariant TRGV4-2/TRGJ5-3 sequence CTYWDSNYAKKLF and invariant TRDV1-4/TRDD2/TRDD3/TRDJ2 sequence CALWELRTGGITAQLVF contained short homology repeats (also known as microhomology regions)⁸ that likely contribute to their formation (Fig. 5B, D). Finally, we observed a low-frequency public TRDV1-6-containing CDR3 sequence in distinct individuals

(Fig. 5A right panel), containing the same short homology repeats between TRDD2-TRDD3 and between TRDD3-TRDJ2 as the TRDV1-4 invariant sequence (Fig. 5D), but in contrast to the TRDV1-4 invariant sequence it does not possess a short homology repeat between the *TRDV1-6* and *TRDD2*, possibly explaining its lower frequency (Supplementary Fig. 6A, B).

In contrast to the TRG and TRD repertoire, no highly public sequences were detected in the TRA and TRB repertoire (Supplementary Fig. 4B; Supplementary Fig. 6C). However, a shared *TRAV1-TRAJ33* sequence was present at low levels among the most prevalent TRA sequences (Supplementary Fig. 6C). This TRA-TRAJ combination is highly prevalent among mucosal-associated invariant T (MAIT) cells, an innate-like T cell subset that is highly conserved across species and recognizes small non-peptidic molecules presented by MRI molecules, a nonclassical MHC³⁹. Alignment of the detected public TRAV1-TRAJ33 sequence from the naked mole-rat with the mouse invariant *Trav1-Traj33* MAIT sequence revealed a very high level of similarity (Supplementary Fig. 6D), suggesting that this sequence might be the TRA fingerprint of the conserved MAIT lineage, present at low frequencies (Supplementary Fig. 6E), in the naked mole-rat. Invariant natural killer T (iNKT) cells are another prototypical innate-like subset that have a semi-invariant TCR that is able to recognize glycolipid antigens presented by the MHC-like molecule CD1d⁴⁰. Notably, no similar invariant rearrangements were detected in the naked mole-rat TCR ecosystem, which is in line with the absence of CD1d from their genome.

In sum, the naked mole-rat TRG and TRD repertoire is dominated by highly public invariant CDR3 sequences, likely directed by short-homology repeat recombination, while in the TRA repertoire a low-level MAIT-like sequence could be detected.

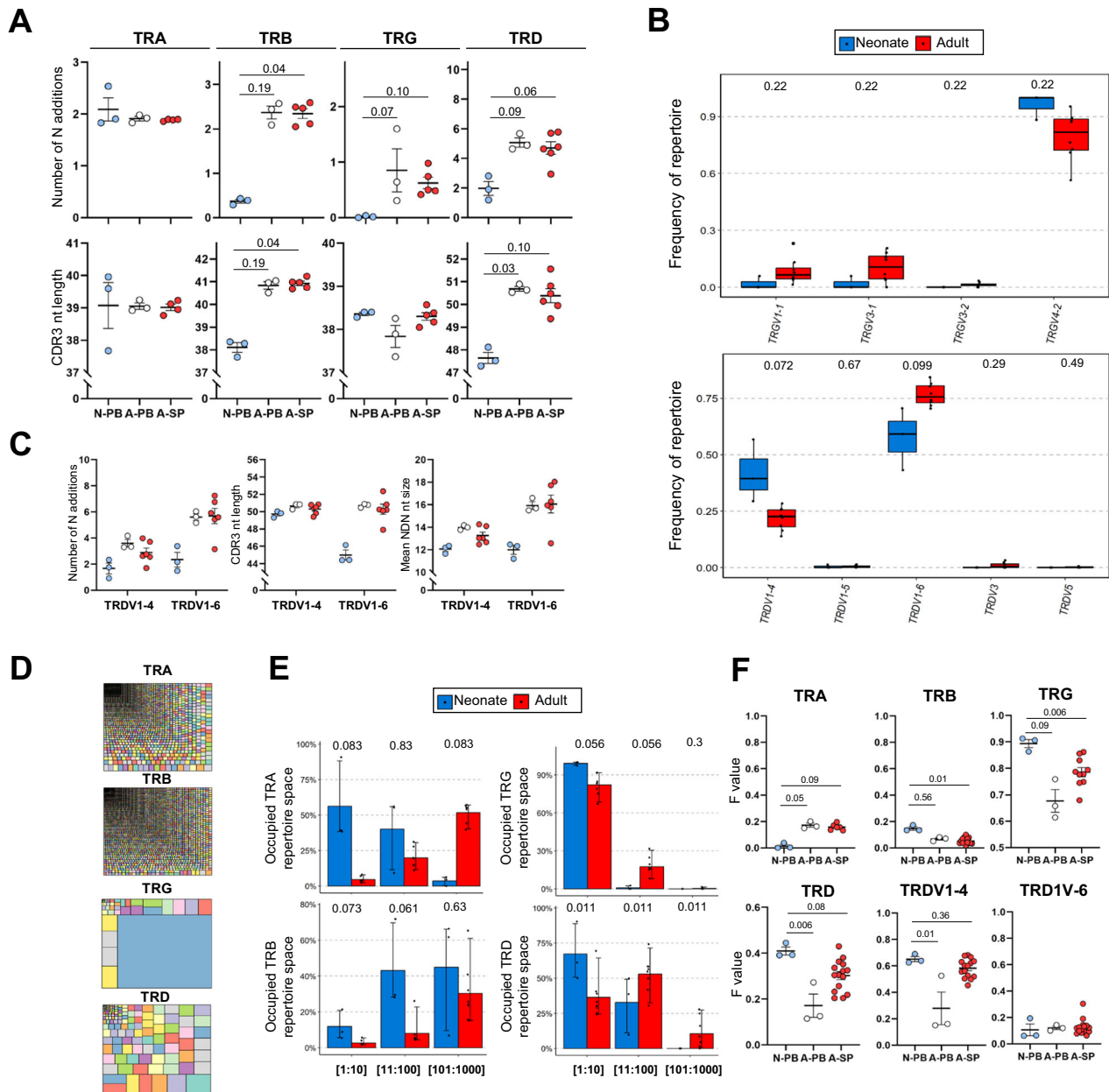


Fig. 4 | The TCR repertoire expressed by the naked mole-rat is influenced by age and possesses an oligoclonal TCR $\gamma\delta$ repertoire. **A Dot plots showing number of N additions (top) and CDR3 nucleotide length (bottom) in expressed CDR3 α , β , γ , and δ sequences. Blue dots: neonate blood (N-PB, $n = 3$), white dots: adult blood (A-PB, $n = 3$), red dots: adult spleen (A-SP, $n = 4-6$). Each dot represents a naked mole-rat. Lines indicate means. **B** Boxplots of TRGV (top) and TRDV (bottom) usage of neonate blood samples ($n = 3$) and adult blood and spleen ($n = 7$). All detected V genes (expressed at RNA level) are displayed. The adult blood and spleen data sets were combined to increase statistical power. **C** Dot plots showing number of N additions (left), CDR3 nucleotide length (middle) and NDN nucleotide size (right) from TRDV1-4 and 6 repertoires. Blue dots: neonate blood, white dots: adult blood, red dots: adult spleen. Lines indicates means. **D** Example of a tree map visualization of a TRA, TRB, TRG, and TRD repertoires. **E** Clonality assessment of the distinct repertoires of neonate blood samples ($n = 3$) and adult blood and spleen ($n = 7$).**

Clonotypes are grouped according to their abundance. 1:10 = first 10 clones, 11:100: 11–100th clone, 101:1000: 101–1000 clone. **F** Dot plots representing mean sharing values (F value) in the TRA, TRB, TRG, TRDV1-4, and TRDV1-6 repertoire of neonate (blood, blue dots (N-PB), $n = 3$) and adult animals (blood (A-PB), $n = 3$) and spleen (A-SP, $n = 4-6$). Repertoires of samples within an age group are compared pairwise and the mean level of sharing is plotted in the graph. Lines indicate means. **B** Boxplots display the interquartile range (25–75th percentile) with the median line. Whiskers extend to the minimum and maximum values within a certain distance from the quartiles, with outliers shown as larger dots. **E** The error bars represent the 95% confidence interval around the mean value (height of the bar). **A**, **C**, **F** Data was analyzed using Kruskal-Wallis with Dunn’s multiple test for pairwise comparisons. **B**, **E** Data was analyzed using Kruskal-Wallis with Holm correction for pairwise comparisons. Source data are provided as a Source Data file.

$\gamma\delta$ T cells from naked mole-rat diverge into two distinct transcriptomic profiles

To gain insights into the potential physiological role of $\gamma\delta$ T cells and the associated significance of our observations regarding the $\gamma\delta$ TCR, we employed TRUST4, a tool for extracting CDR3 sequences from single-cell sequencing data. We utilized a publicly available dataset

obtained from the blood of naked mole-rats, comprising samples from three young adult animals (around 3 years old) and three middle-aged animals (around 11 years old)⁴¹. After extracting CDR3 sequences and filtering for TCR-positive cells, six distinct clusters were identified that correlated well with previously identified clusters from the original study⁴¹ (Fig. 6A, Supplementary Fig. 7A, B). The TRG and TRD

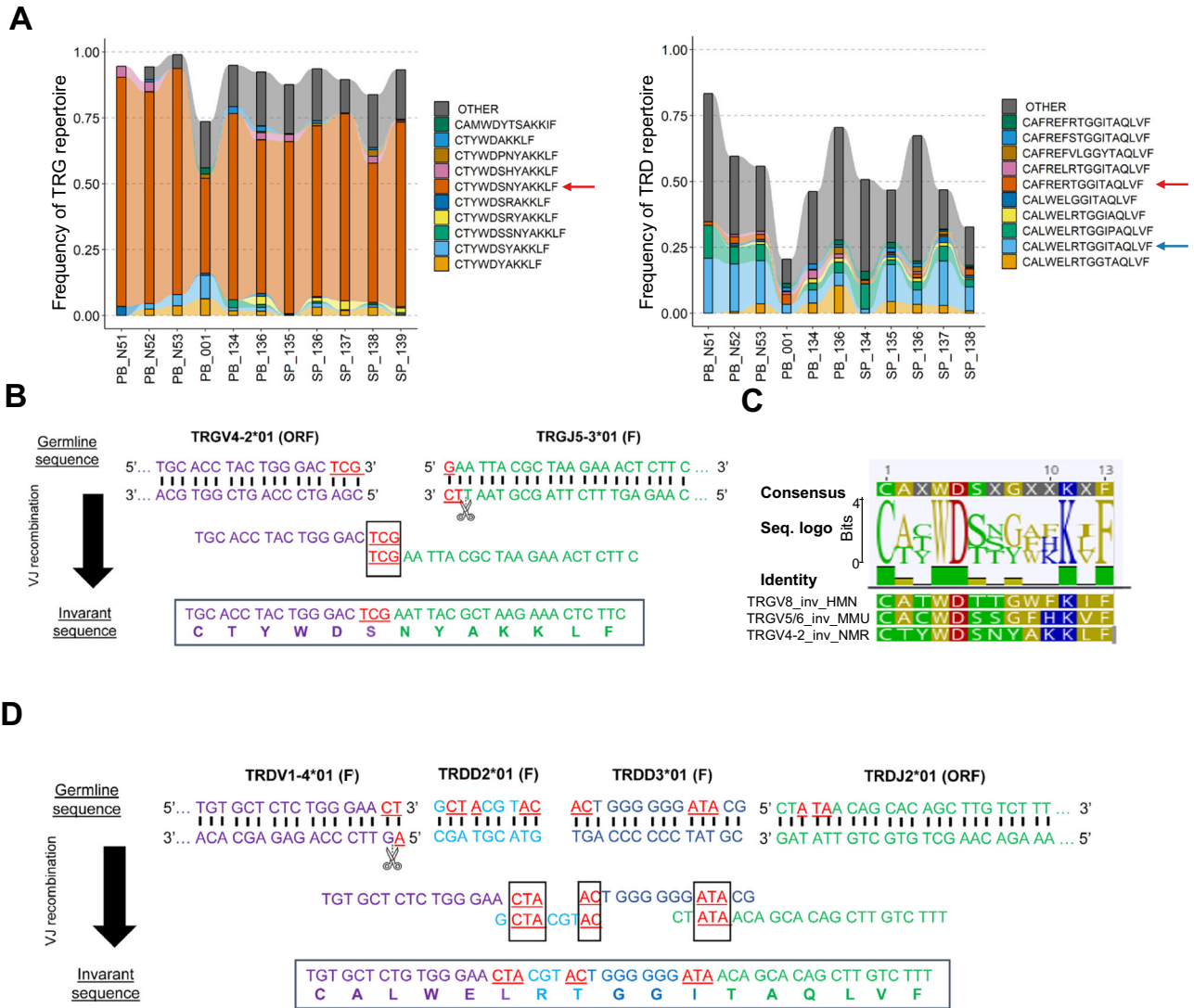


Fig. 5 | Invariant naked mole-rat TRGV4-2 and TRDV1-4 CDR3 sequences can be generated by a mechanism of short homology repeat recombination. A Alluvial plots displaying the frequency of top 10 shared gamma (left) and delta (right) TCR sequences in naked mole-rat animals. The red arrow in the left plot indicates the most abundant public TRG sequence. Blue and red arrows in the right plot indicate the most abundant public TRDV1-4 and TRDV1-6 sequences respectively. Sample identification numbers of individual naked mole-rats ($n = 11$) are indicated at the x axis; see Supplementary Table 1 for more information of each ID number. **B** Recombination using short homology repeats between TRGV4-2 and TRG5-3. **C** Sequence alignment and consensus logo plot (top) of the naked mole-rat TRGV4-2 invariant CDR3 sequence (NMR suffix), human TRGV8 invariant CDR3 sequence (HMN suffix) and the mouse TRGV5/TRGV6 invariant CDR3 sequence (MMU suffix). Amino acid key colors: dark yellow = non-polar aliphatic/aromatic, red = polar negative, blue = polar positive, green = polar neutral. **D** Recombination using short homology repeats between *TRDV1-4*, *TRDD2*, *TRDD3* and *TRDJ2*. **B**, **D** Overlapping nucleotides are shown in red. Source data are provided as a Source Data file.

CDR3 sequences were highly prevalent in cluster 5, which aligned with the T-cell $\gamma\delta$ cluster reported in the original study based on gene expression profiles⁴¹ (Fig. 6B, Supplementary Fig. 7A). Surprisingly, TRG and TRD CDR3 sequences were also enriched in cluster 4, previously⁴¹ identified as a T-regulatory (T-reg)-like cluster, and a few sequences were present in cluster 3, corresponding to the original cluster⁴¹ of CD8 T cells (Fig. 6B, Supplementary Fig. 7A, B). $\gamma\delta$ T cells in cluster 5 exhibited differential expression of genes associated with cytotoxic lymphocytes (*TBX21*, *IL2RB*, *CX3CR1*, *CCL5*, *NKG7*, *SLAMF7*, *CCR5*, *FCGR3A*), and interestingly, some NK receptors (*KLRD1*, *KLRG1*), suggesting that cells from this cluster may possess the functional characteristics of bona fide NK cells in naked mole-rats (Fig. 6C, Supplementary Fig. 7C). Further confirmation of the NK-like phenotype of cluster 5 cells came from results of the hypergeometric enrichment test applied to the differentially expressed genes, as mouse NK-gene sets ranked among the top-associated terms in the analysis

(Supplementary Fig. 7D). Intriguingly, cells from cluster 4 displayed a transcriptome enriched with molecules associated with diverse functionalities, including members of the tumor necrosis factor receptor superfamily (*TNFRSF4*, *TNFRSF18*), as well as a set of genes associated with IL-17 production in mouse and human $\gamma\delta$ T cells (*AHR*, *CD44*, *MAF*, *BLK*)⁴²⁻⁴⁴ (Fig. 6C, Supplementary Data 1).

In sum, based on CDR3 as a natural barcode, we identified two functionally different $\gamma\delta$ T cell clusters among blood naked mole-rat T cells: one associated with NK killing machinery while the other appears to be more functional diverse and include expression of type 3-associated genes.

Naked mole-rat NK-like effector $\gamma\delta$ T are enriched for invariant TCR sequences

Analysis of the single-cell TCR data revealed that the two primary $\gamma\delta$ T cell clusters (clusters 4 and 5) contained distinct TCR ecosystems: cells

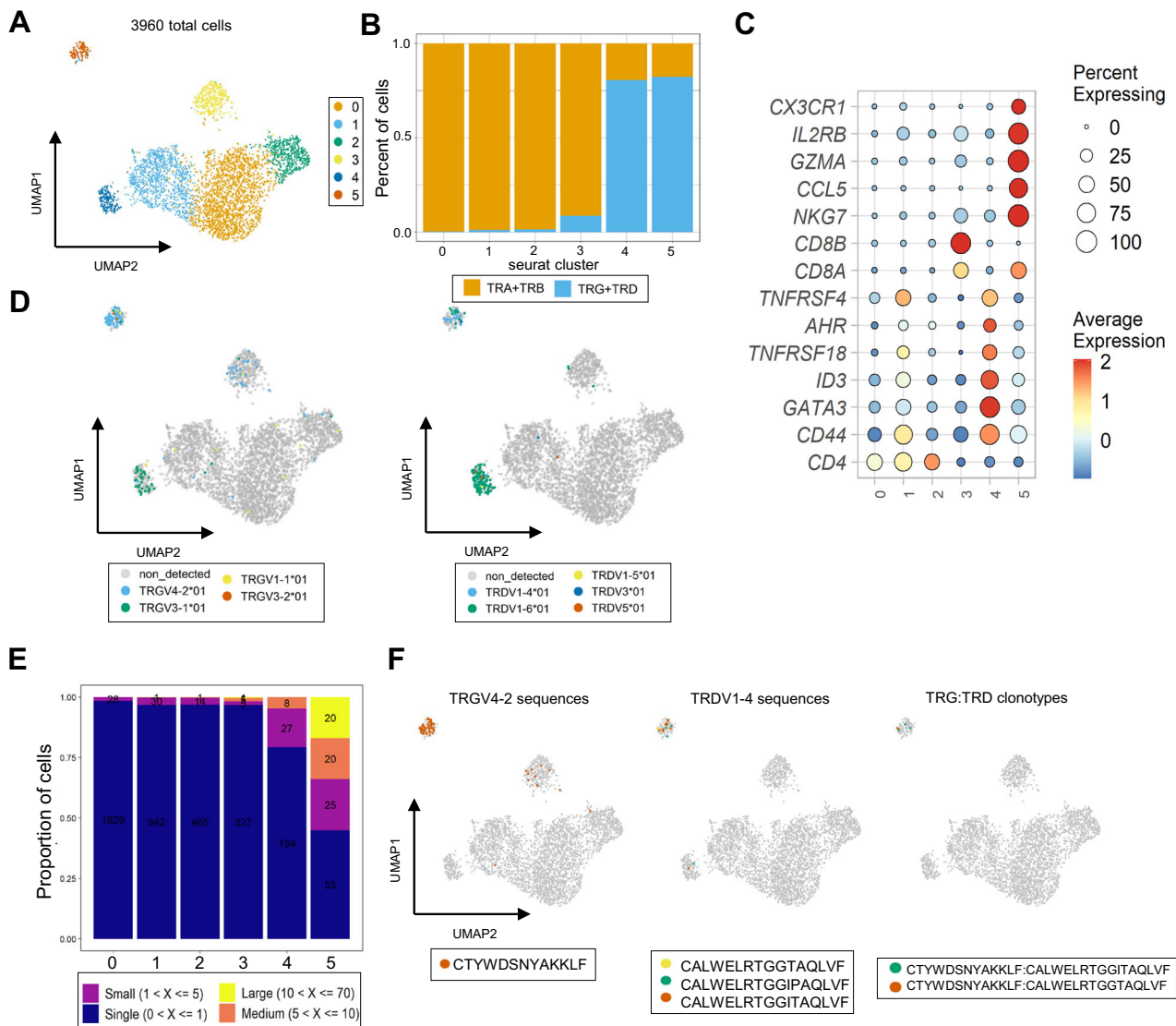


Fig. 6 | Naked mole-rat NK-like effector $\gamma\delta$ T are enriched for invariant TCR sequences. **A** UMAP plot from blood T cells of 4 young and 4 middle-aged adult animals. The plot only shows single cells where TCR/CDR3 transcripts were detected. **B** Bar plots indicating the relative frequency of TCR $\alpha\beta$ and TCR $\gamma\delta$ cells in the different clusters identified in **A**. Cells are classified in one of the two categories if at least one of both constitutive TCR chains (TRA or TRB and TRG or TRD) is detected in their transcriptome. **C** Dot plot heatmap displaying row-scaled Log2-fold change (logFC) expression values of selected DEGs in the different clusters identified in **A**. Each dot represents the average expression profile of all the cells in

the cluster. **D** UMAP plots (as generated under **A**) displaying TRGV (left) and TRDV (right) usage. **E** Clonality analysis of single cell TCR-seq data of the clusters identified in **A**. All TCR sequences (TRA, TRB, TRG and TRD) are classified in different categories according to their occurrences in the single cell dataset. Numbers in the bar plots refer to the number of TCR sequences in each category. **F** UMAP plots (as generated under **A**) highlighting specific TRGV4-2 (left) and TRDV1-4 sequences (middle) and TRG:TRD paired clonotypes (right). Source data are provided as a Source Data file.

from cluster 5 exhibited enrichment for TRGV4-2- and TRDV1-4-containing CDR3 sequences, while cluster 4 was predominantly composed of TRGV3-1- and TRDV1-6-containing CDR3 transcripts (Fig. 6D), in line with our bulk $\gamma\delta$ TCR sequence data (Fig. 4B, Fig. 5). This variation in V gene usage was accompanied by differences in TRGJ and TRGC usage, with TRGJ5-3 and TRGC5 being prevalent in the TRGV4-2-enriched cluster 5, and TRGJ1-2 and TRGC1 in the TRGV3-1-enriched cluster (Supplementary Fig. 8A). Notably, these two TRGV-TRGJ-TRGC pairings are located in separate cassettes of the naked-mole rat TRG locus, although intriguingly, TRGV3-1-TRGJ1-2 are paired with the non-functional TRGC1 gene (Fig. 1A). All TRD sequences were paired with the same TRDJ gene (TRDJ2) and used the only available TRDC gene (Supplementary Fig. 8A, Fig. 1C). Differences in the level of TRG and TRD detection between clusters 4 and 5 (Fig. 6D) can be related to the fact the different types of $\gamma\delta$ T cells can express different $\gamma\delta$ TCR levels⁴⁵⁻⁴⁸.

Clonality analysis of the TCR repertoires demonstrated that both cluster 4 and cluster 5 were enriched for expanded TCR sequences (Fig. 6E). Consistent with this observation, diversity was reduced in the TRG repertoire of both clusters 4 and 5 (Supplementary Fig. 8B). For TRD, however, this low level of diversity was limited to cluster 5 (Supplementary Fig. 8B). Visualization in the UMAP plot of cells expressing the invariant TRGV4-2 CTYWDSNYAKKLF sequence and the invariant TRDV1-4 CALWELRTGGTAQLVF sequence clearly demonstrated a localization bias towards the cluster 5 exhibiting NK-like effector features (Fig. 6F left and middle panel) and confirmed the pairing of the two invariant sequences (Fig. 6F right panel). Interestingly, we detected expression of the invariant TRGV4-2 sequence in the CD8 cluster (cluster 3); however, TRD detection was low, and the only detected sequences were non-invariant TRDV1-4-containing CDR3 sequences (Fig. 6D, F). Smaller expanded TRDV1-6 and TRGV3-

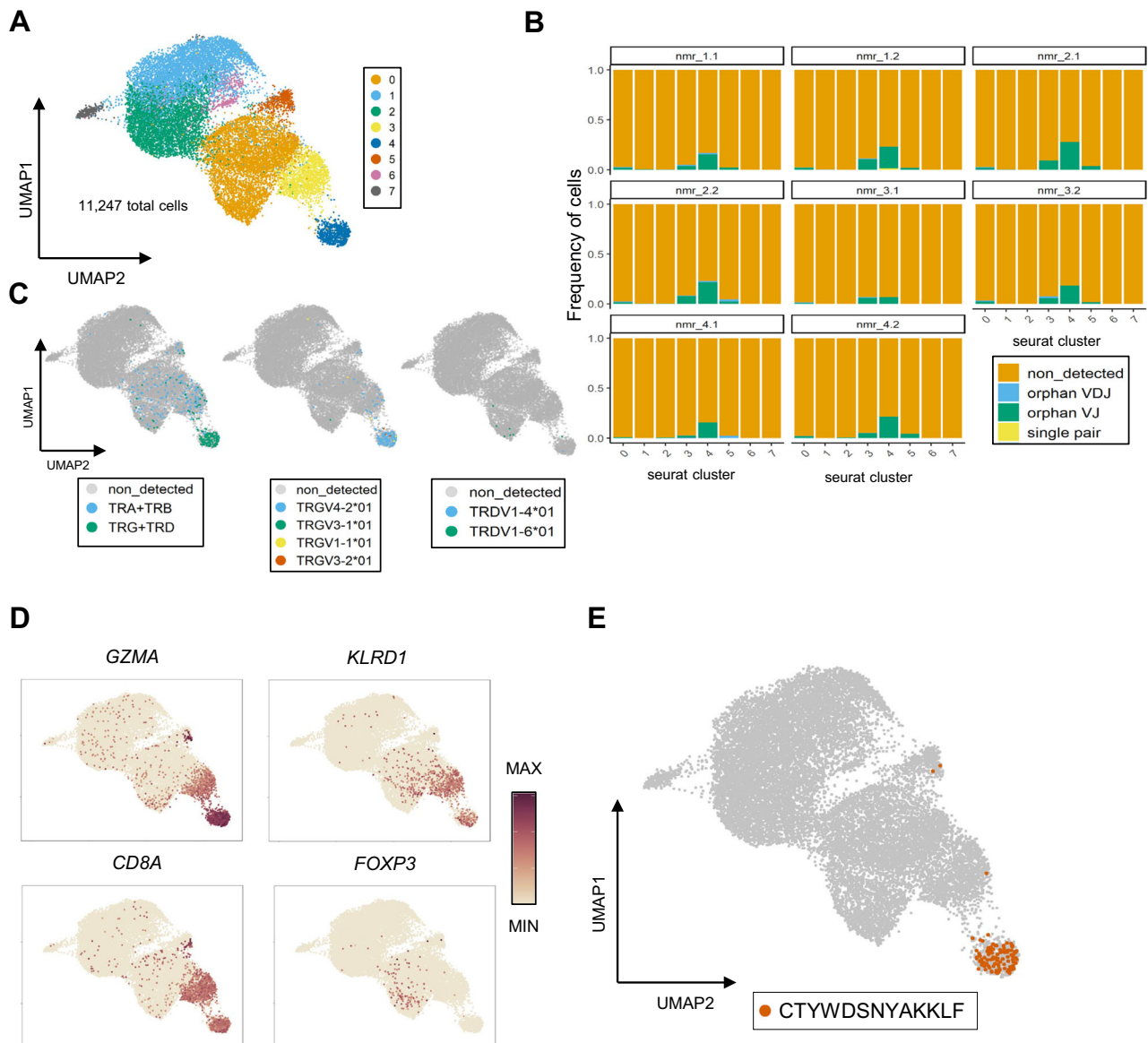


Fig. 7 | Spleen NK-like effector $\gamma\delta$ T cells also contain invariant TCR γ sequences. **A** UMAP plot from spleen single cell T cells of 4 young-adult animals. **B** Bar plots indicating frequencies of detected TCR transcripts in the total T cell clusters identified in (A). Orphan_VDJ = cells with single TRB or TRD chain, orphan VJ: cells with single TRA or TRB chain, single pair: cells with a single TRA-TRB or TRG-TRD pair, non_detected: cells where TRUST4 did not extract any CDR3 information. Each

animal has 2 replicate experiments labeled with the .1 and .2 suffix. **C** UMAP plots (as generated under (A)) displaying receptor subtype (left), TRGV (middle), and TRDV (right) usage (each graph shows the data of a single animal). **D** UMAP plots (as generated under (A)) displaying expression levels of selected genes. **E** UMAP plot (as generated under (A)) highlighting the cells that express the invariant TRGV4-2 sequence. Source data are provided as a Source Data file.

1 sequences were detected in the case of $\gamma\delta$ T cells from cluster 4 compared with the invariant constituents of the NK-like TCR, with some of the TRDV1-6 sequences constituting private expansions (Supplementary Fig. 8C). Intriguingly, we detected two cells expressing the previously reported invariant TRDV1-6 sequence (Fig. 5A right panel, Supplementary Fig. 6A, B, Supplementary Fig. 8C) that clustered in cluster 5 (Supplementary Fig. 8C). Further analysis of the paired TRG-TRD data confirmed that cells from cluster 4 (enriched for *TRDV1-6*) preferentially paired with TRGV3-1 sequences in different animals, while cells in cluster 5 (enriched for *TRDV1-4*) exhibited pairing with TRGV4-2 sequences (Supplementary Fig. 8D). Lastly, the TRAV1-TRAJ33 MAIT-like sequence (Supplementary Fig. 6D) also clustered within the region where NK-like cells were present (Supplementary Fig. 8E).

In order to extend our findings in the blood to the spleen compartment, we re-analyzed a public available dataset from spleens of

four young-adult animals (23–25 months old)⁶. After applying a similar pipeline as the one used before with the blood dataset, we observed lower levels of TCR detection in all T-cell clusters compared to the blood experiment (Fig. 7A, B). The majority of $\gamma\delta$ T cells were localized in cluster number 4 and a few $\gamma\delta$ T cells were scattered inside cluster 3 (Figs. 7A, C left panel). Cluster 4 corresponded to an NK-like effector cluster (Figs. 7A, D, Supplementary Data 2), and, similar to our findings in the blood dataset, the TRGV4-2 CTYWDSNYAKKLF CDR3 sequence was highly enriched in this cluster (Fig. 7C middle panel, Fig. 7E). Due to the extremely low detection levels, we were unable to draw conclusions regarding TRDV usage in this spleen cluster 4 (Fig. 7C right panel). Additionally, cells from cluster 3, which also contained some TRGV4-2 sequences, exhibited characteristics of conventional CD8 T cells (Fig. 7A, C, D), again aligning with our observations in the blood dataset (Fig. 6A, D, Supplementary Fig. 7A). Notably, the few TRDV1-6-containing sequences clustered far away from the TRGV4-2 cells,

within a region where the T-regulatory gene *FOXP3* is expressed (Fig. 7C, D).

Thus overall, the NK-like effector $\gamma\delta$ T cells in the periphery (blood, spleen) express an invariant γ 4-2/ δ 1-4 TCR, while in the blood another more variable, both at functional and TCR level, $\gamma\delta$ T cell cluster is present.

Naked mole-rat invariant $\gamma\delta$ TCR sequences are generated in the thymus until mid-life

To gain insights into the natural dynamics of the $\gamma\delta$ TCR repertoire across different age groups, we conducted TCR sequencing of thoracic and cervical thymic tissue samples⁷ from neonates and young-adult (around 3 years old) animals. Despite the limited number of samples, we observed consistent trends in the general properties of the CDR3 regions as seen in the blood and spleen bulk TCR-seq results such as lower junctional diversity and shorter CDR3 sequences compared to the adult counterparts (Supplementary Fig. 9A). Analysis using Tree

maps revealed that in both age groups, the TRG and TRD repertoires exhibited oligoclonal patterns, whereas the TRA and TRB repertoires displayed high levels of polyclonality (Fig. 8A, B, Supplementary Fig. 9B). The invariant TRGV4-2 and TRDV1-4 sequences were detected in both age groups (Fig. 8A, B). This suggests that the production of these invariant sequences is sustained until 3 years of age, which contrasts with observations in mice and humans^{9,43,49}. However, it cannot be excluded that the detected invariant sequences originated from the re-circulation of peripheral T cells towards the thymus⁵⁰. To confirm the sustained production of these invariant TCR chains, we utilized a publicly available single-cell dataset obtained from the thoracic and cervical thymi of 2 young-adult (around 3 years old) and 2 middle-aged (around 11 years old) adult animals⁷. Cells from T cell clusters in the original thoracic and cervical thymic datasets integrated well with each other (Fig. 8C, D, Supplementary Data 3). The overall detection of TCR CDR3 transcripts using TRUST4 was at a similar level to what we observed in the blood single-cell dataset (Supplementary

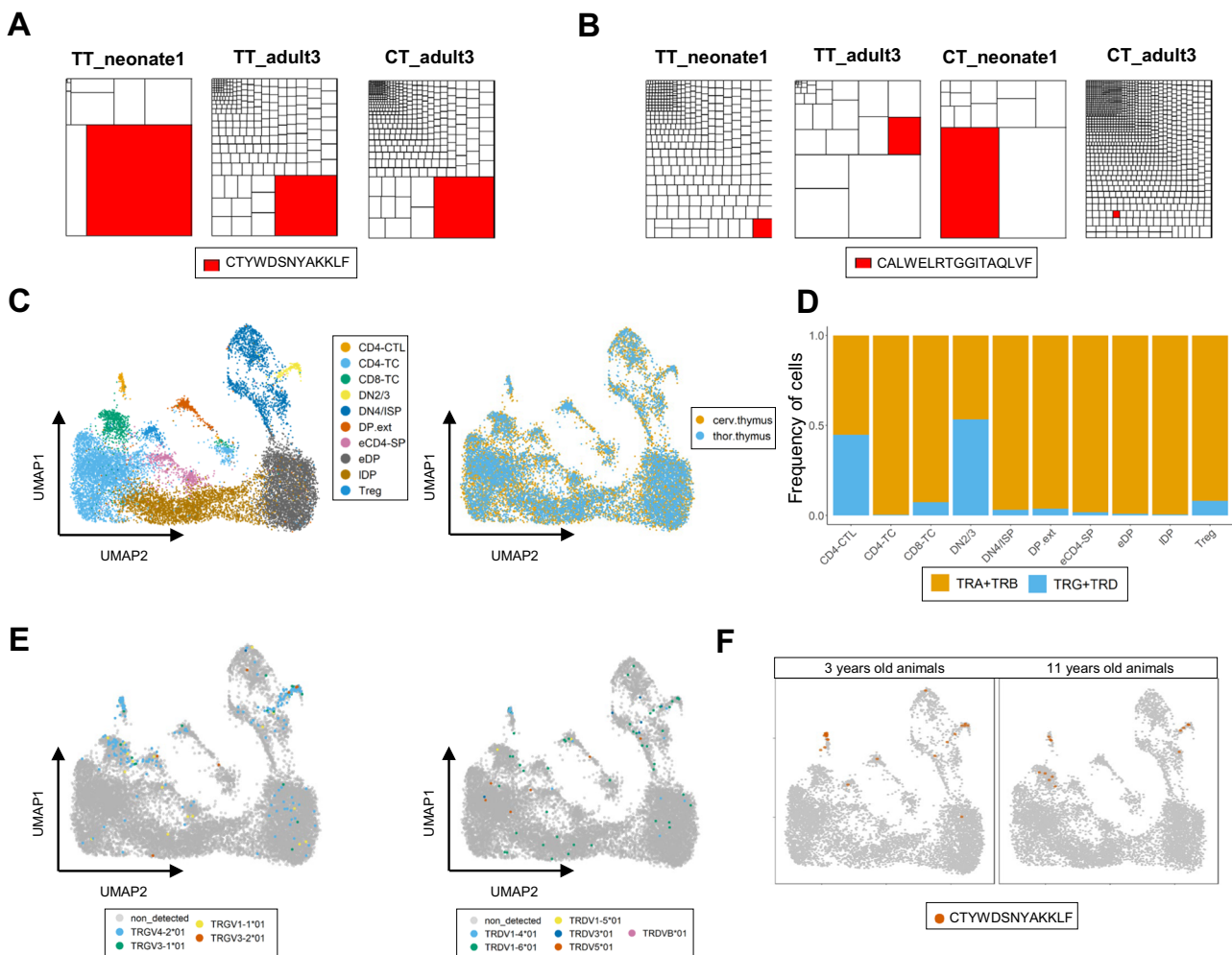


Fig. 8 | Naked mole-rat invariant $\gamma\delta$ TCR sequences are generated in the thymus until mid-life. **A** Tree maps representing the TRG repertoire in neonate and adult thymic samples highlighting in red the prevalence of the invariant TRGV4-2 CDR3 sequence. **B** Tree maps representing the TRD repertoire in neonate and adult thymic samples highlighting in red the prevalence of the invariant TRDV1-4 CDR3 sequence. For (**A**, **B**): each Tree map represents the repertoire data of the indicated naked mole-rat ID (see Supplementary Table 1 for more information for each ID number). **C** UMAP plot colored based on identified clusters from thymic (thoracic and cervical) single cells (left) and colored by organ origin of the cells (right). Data are obtained from a public dataset with sequencing data from 4 animals (2 young-adult and 2 middle-aged adult individuals). **D** Bar plot indicating the relative frequency of TCR α and TCR β cells in the different thymic clusters identified under (**C**). Cells are

classified in one of the two categories if at least one of both constitutive TCR chains (TRA or TRB and TRG or TRD) is detected in their transcriptome. Cells without TCR transcript detection (non_detected) are not depicted in the plot (**E**) UMAP plots (as generated under (**C**)) displaying TRGV (left) and TRDV (right) usage in the single cell thymic data. **F** UMAP plot as generated under (**C**) highlighting the thymocytes that express the invariant TRGV4-2 sequence across the ontogeny. Abbreviations on (**C**): DN2/3 (double negative); DN4/ISP (immature single positive); eDP (early double-positive), IDP (late double-positive), eCD4-SP (early CD4 single-positive); CD4-TC (CD4⁺ T-lymphocyte); CD8-TC (CD8⁺ T-lymphocyte); DP.ext (extrathymic/circulating DPs); Treg (regulatory T cell); CD4-CTL (CD4⁺ cytotoxic T-lymphocyte). Source data are provided as a Source Data file.

Fig. 9C). However, the relative abundance of $\gamma\delta$ T cells in T-reg and NK-like effector clusters (referred to as CD4-CTL in the original paper's annotation) was lower (Fig. 8D, E). Interestingly, a significant proportion of $\gamma\delta$ TCR transcripts, primarily TRG chains, were detected in the double negative (DN)2/3 cluster, which corresponds to T-cell precursors (Fig. 8D, E). In contrast, the detection of TRA rearrangements in this cluster was relatively low compared to TRG, TRB, and TRD transcripts, which follows the stereotypical sequential loci rearrangement described in human and mouse thymocyte development (Supplementary Fig. 10A)¹³. Consistent with the findings in blood and spleen, TRGV4-2 transcripts were particularly abundant in the NK-like effector cluster (annotated as "CD4-CTL", Figs. 8C, E left panel). Due to lower detection of TRD transcripts, the presence of TRDV1-4 sequences in that cluster was limited (Fig. 8E right panel). As observed in other anatomical contexts, invariant TRGV4-2 sequences were highly enriched in the NK-like effector thymic cluster at 3 and 11 years and, interestingly, they were already present in the progenitor DN2/3 cluster at both 3 and 11 years of age (Fig. 8F), indicating sustained production of the invariant CTYWDSNYAKKLF sequence until middle-aged adulthood. Similarly, while the levels of detection of the invariant TRDV1-4 sequence were quite low, they were only detected in the CD4-CTL cluster (Supplementary Fig. 10B left panel). Furthermore, akin to T cells in blood, sequences of the TRAV1-TRAJ33 MAIT-like sequence were present in the NK-like cluster, with some sequences also detected in the maturing cluster eCD4-TC (Supplementary Fig. 10B right panel). Expression of naked mole-rat *MRI*, the MHC-like molecule involved in the development and function of MAIT cells in mice and human⁴⁰, could be observed (Supplementary Fig. 10C) and it showed high homology with its mouse and human counterparts (Supplementary Fig. 10D, E).

Discussion

The naked mole-rat lack NK cells⁶. Here, we identify public invariant $\gamma\delta$ T cells that are pre-programmed during development in the thoracic and cervical thymi of the naked mole-rat with NK-like effector functions, before going to the periphery (blood, spleen).

The invariant TRG chain is formed by the rearrangement in the genome (V(D)J recombination) of the *TRGV4-2* gene with the *TRGJ5-3* gene while the invariant TRD chain is formed by the rearrangement of TRDV1-4/TRDD2/TRDD3/TRDJ2. We identified in all these sets of genes short homology repeats at the coding ends, also known as microhomology domains, that likely drive the formation of the invariant TRG and TRD chains. These invariant sequences dominate the naked mole-rat $\gamma\delta$ TCR repertoire and by single-cell analysis we formally show that these chains pair and thus assemble a complete invariant V γ 4-2/V δ 1-4 $\gamma\delta$ TCR. Such short homology repeats are used by early life $\gamma\delta$ T cells both in mouse and human⁸. In these species the generation of these TCRs is limited to the fetal period, when the expression of TdT is low or absent, thus allowing the recombination of the short homology repeats to occur as there is not disturbance by the random addition of N nucleotides at the coding ends of the genes^{8,51}. Remarkably, in contrast to mouse and human^{9,43,49}, in the naked mole-rat this invariant V γ 4-2/V δ 1-4 TCR is being generated at least till adult life (11 years old). This might be due to the relative low level of N additions that are seen at that age: for example for the TRG chain this is around 0.5 for blood/spleen and 2 for thymus in the naked mole-rat, while for human adult blood/post-natal thymus (including a 9 year-old thymus) the number of N additions is 5, thus up to 10 times higher^{49,52}.

We propose that the naked mole-rat invariant V γ 4-2/V δ 1-4 TCR recognize a common molecular signal shared by naked mole-rat cells undergoing transformation which then activates the NK-like killing effector program (including release of perforin and granzymes) that is intrinsically present in the peripheral V γ 4-2/V δ 1-4 $\gamma\delta$ T cells prior to signal encounter. Such functional programming is also observed in the mouse and human fetal thymic invariant $\gamma\delta$ T cells^{14,43,49}, but only in the

naked mole-rat, this NK effector program persists following the production of the V γ 4-2/V δ 1-4 thymocytes at least till 11 years. This allows the continuous generation of "fresh" invariant $\gamma\delta$ TCR NK-like effector T cells that can actively control the formation of cancer cells in the periphery. Such continuous production would prevent exhaustion of $\gamma\delta$ T cells as has been seen for example when attempting to activate multiple times human V γ 9V δ 2 T cells in clinical trials²⁷. Because of this, efforts are made to generate $\gamma\delta$ T cell cellular products for adoptive cellular immunotherapy starting from human cord blood that results in larger expansions compared to adult peripheral blood because of their less differentiated nature⁵³. In human, the prevalent TRDV gene used in early life is *TRDV2*, which can pair with *TRGV9* to form innate phosphoantigen-reactive V γ 9V δ 2 T cells or pair with other V γ chains in order to generate V γ 9-V δ 2+ $\gamma\delta$ T cells⁸. Human V γ 9V δ 2 $\gamma\delta$ T cells can also be made in the post-natal thymus, but with a different type of V γ 9V δ 2 TCR and at much lower frequencies^{54,55}. Remarkably, while in the mouse no orthologues for the human *TRGV9* and *TRDV2* genes could be found, we showed here, using unrooted phylogenetic analysis, that clear orthologue genes exist in the naked mole-rat. However, these were categorized as pseudogenes, and we could indeed not find any usage of these TRGV and TRDV genes within expressed CDR3 sequences. In parallel, no members of the BTN3 group of butyrophilins, important for phosphoantigen-reactivity of the human and alpaca V γ 9V δ 2 TCR^{12,32,33,56}, could be identified in the naked mole-rat, which is in line with the proposed co-evolution of functional TRGV9, TRDV2, and BTN3 genes^{12,57}. It is therefore unlikely that the naked mole-rat possess phosphoantigen-reactive $\gamma\delta$ T cells. We propose that the role of innate $\gamma\delta$ T cells is rather taken by the abundant subset of invariant V γ 4-2/V δ 1-4 $\gamma\delta$ T cells. The naked mole-rat *TRGV4-2* gene was most similar to mouse *Trgu3* (annotated as an open-reading frame, IMGT) and *Trgu1/Trgu2* that are used in the adaptive mouse $\gamma\delta$ TCR repertoire¹⁴. There were also some similarities with the human *TRGV11* gene, but this gene is a pseudogene⁵⁸. Thus, there appears to be rapid evolution of the TRGV genes, which is also observed for the functional TRGV genes (with the noted exception of *TRGV9*) within primates⁵⁹ and other mammals⁶⁰. Despite the absence of obvious similarities between naked mole-rat, mouse, and human regarding the TRGV genes that are used in the formation of invariant $\gamma\delta$ TCRs, it was interesting to observe that at the CDR3 level there was a high homology, both at the level of CDR3 length and amino acid physicochemical properties, of the naked mole-rat TRGV4-2 CDR3 sequence CTYWDSNYAKKLF with the human fetal-derived TRGV8 CDR3 CATWDTTGWFKIF^{36,43,49} and mouse fetal-derived Trgv5/v6 CDR3 CACWDSGFKHVF^{37,38,52}. This may indicate that these invariant CDR3 γ sequences may be involved in the recognition of similar molecular moieties, that are for example expressed on malignant cells. The strong bias of the naked mole-rat TCR repertoire towards an invariant clonotype may also reflect the lower need for a high polyclonal TCR repertoire due to a possible lower pathogen diversity in the underground environment. It has been suggested that naked mole-rat's unique living conditions have led to the acquisition of intrinsic anti-cancer mechanisms as a secondary effect. Similarly, the pathogen-deprived environment in which they reside may have facilitated the emergence of invariant TCR mechanisms for protection, resembling the human fetal context where the pathogen-privileged utero environment also relies on an invariant $\gamma\delta$ TCR repertoire to combat congenital infections that breach its protected sterile environment^{36,61,62}.

Besides the invariant V γ 4-2/V δ 1-4 NK-like effector cells, we identified in the blood of the naked mole-rat also a $\gamma\delta$ cell cluster that showed a more variable TCR and functional repertoire. While there was enrichment for the usage of TRGV3-1 and TRDV1-6, the CDR3 regions were more variable than in the V γ 4-2/V δ 1-4 subset, especially for the TRDV1-6 CDR3. TRGV3-1 paired with a different TRGJ (TRGJ1-2) than the "invariant" TRGV4-2 and, intriguingly, with the *TRGCI* gene which was annotated as a constant pseudogene because of the lack of exon 2.

However, since no other V gene could be observed in the TRGV3-1/TRGJ1-2/TRGCI-expressing cells, we conclude that this TRG chain is likely the functional chain that pairs with the TRDV1-6 chain. Of note, a similar usage of a TRGC lacking exon 2 in an expressed TRG chain has been described in salmon⁶³. More generally, different TRGC can show different exon-intron organizations, which is mainly linked to the different number of exon 2 copies, of which the biological significance remains unclear⁶⁰. Functionally, the V γ 3-1/V δ 1-6 $\gamma\delta$ T cells were clearly different from the NK-like program observed in the invariant V γ 4-2/V1-4 cells. They express a range of genes without clear association to a particular programmed function, although some genes linked to IL-17 production were enriched.

We could not detect TCR sequences that are similar to the ones described to be present in invariant NKT cells in other species, which coincided with the absence of *CD1D* from the naked mole-rat genome, the antigen-presenting element that is essential for the development and activation of these unconventional $\alpha\beta$ T cells⁴⁰. In contrast, a low level (0–0.4% of TRA sequences) of a MAIT TRA sequence was detected, together with *MRI*, the presenting element of vitamin B metabolites to the MAIT TCR⁴⁰. Thus, while the blind mole-rat (*Spalax galili*) lacks both iNKT and MAIT cells⁶⁴, the naked mole-rat conserved low levels of MAIT cells. The absence of iNKT and only low levels of MAIT cells in the naked mole-rat could explain the high prevalence of $\gamma\delta$ T cells, since in the mouse these three unconventional T cell types compete for the same niche⁶⁵ and we showed that the few MAIT cells detected in our single-cell data cluster functionally together with the invariant V γ 4-2/V δ 1-4 $\gamma\delta$ T cell cluster suggesting the $\gamma\delta$ and MAIT cells also share a functional niche in the naked mole-rat.

In summary, we have identified and characterized $\gamma\delta$ T cell biology in the naked mole-rat providing a framework to investigate the intrinsic anticancer biology of this species from an immunological $\gamma\delta$ T cell-based perspective.

Methods

TCR locus annotation

Two chromosome-level *Heterocephalus glaber* genome assemblies, the naked mole-rat maternal (GCA_944319715.1) and paternal (GCA_944319725.1), were obtained from the National Library of Medicine (NCBI) Assembly database to study the TR loci⁶⁶. Maternal and paternal assemblies were derived from the gonad tissue of a 5-year-old male naked mole-rat generated via the 10x Linked reads sequencing and Nanopore long-read sequencing respectively. The annotation of the T-Cell Receptor (TR) loci of the *Heterocephalus glaber* was conducted following the comprehensive IMGT[®] biocuration pipeline⁶⁷, which is guided by the axioms and concepts of IMGT-ONTOLOGY. These concepts include “IDENTIFICATION,” “DESCRIPTION,” “CLASSIFICATION,” “NUMEROTATION,” “LOCALIZATION,” “ORIENTATION,” and “OBTENTION”; in order to maintain the accuracy and consistency of the annotation process by providing a controlled vocabulary and annotation rules⁶⁸.

To identify and extract the TR loci from the corresponding chromosomal assembly, a comparison was performed between them and the TR reference sequences of *Homo sapiens* using the Basic Local Alignment Search Tool (BLAST)⁶⁹. The delimitation for each locus was determined by detecting the IMGT[®] Bornes, conserved non-TR-coding genes located upstream of the first or downstream of the last gene of an IG or TR locus²⁸. The IMGT[®] 5' bornes for the TRB, TRG, and TRAD locus are *MOXD2*, *AMPH*, and *ORIOG3* respectively, while their IMGT[®] 3' bornes are *EPHB6*, *STARD3NL*, and *DADI* respectively. Each IMGT[®] annotated locus sequence was submitted to the Third Party Annotation (TPA) database and was assigned a unique NCBI accession number. In the naked mole-rat maternal assembly, the accession numbers are as follows: BK064754.1 for TRA/TRD, BK064759.1 for TRB, and BK064756.1 for TRG. In the naked mole-rat paternal assembly, the accession numbers are: BK064758.1 for TRB and BK064755.1 for TRG.

Note that we did not submit a naked mole-rat annotation for the TRAD locus based on the paternal assembly because of the presence of gaps, indicative of lower quality sequencing in that genome region.

The IMGT[®] nomenclature for the TR genes, was based on the “CLASSIFICATION” concept of the IMGT-ONTOLOGY⁶⁸. TR genes of the naked mole-rat were classified into a subgroup based on sequence identity with human and/or mouse TR subgroups, calculated using IMGT V-QUEST⁷⁰, Clustal Omega⁷¹, and Nphylgeny⁷². Variable genes with a V-REGION identity score above 70%, while those with scores below 70% were assigned to new subgroups. TR variable genes were marked by the subgroup number, and if there were multiple genes belonging to the same subgroup, they were followed by a hyphen and a number indicating their relative position in the locus. Genes with highly degenerated V-REGION that cannot be classified into a subgroup were marked with a letter instead of a subgroup number. Additionally, new alleles were annotated when the coding part of each gene (V-REGION, J-REGION, D-REGION or C exons) had an identity score of $\geq 99\%$ in different assemblies. The functionality of genes and alleles was defined according to the “IDENTIFICATION” axiom of IMGT-ONTOLOGY. An allele was considered as functional (F) if its coding region had an open reading frame without stop codons or any defects in the splicing sites, recombination; as open reading frame (ORF) if its coding region had an open reading frame without stop codons but showed alterations in the splicing sites, recombination signals or conserved amino acids; and as pseudogene (P) if the coding region had stop codon(s) and/or frameshift mutation(s)⁷³.

The annotated data of the current analysis were entered into the IMGT[®] reference directory and IMGT[®] databases, including IMGT/LIGM-DB⁷⁴, IMGT/GENE-DB⁷⁵, IMGT/3Dstructure-DB⁷⁶ and IMGT/2Dstructure-DB²⁸ by using a set of standardized labels described in the “DESCRIPTION” axiom of IMGT-ONTOLOGY. The data also entered in IMGT[®] sequence analysis tools (IMGT/V-QUEST (6), IMGT/HighV-QUEST⁷⁷, and IMGT/DomainGapAlign⁷⁸). Finally, all the genomic annotated information of the TR loci of the naked mole-rat was consolidated into the IMGT[®] Repertoire (<http://www.imgt.org/IMGTrepertoire/>), which encompasses various web resources such as: Locus representation, Locus description, Locus in genome assembly, Locus gene order, Locus Borne, Gene tables, Potential germline repertoire, Protein displays, Alignments of alleles, Colliers de Perles, and germline [CDR1-IMGT.CDR2-IMGT.CDR3-IMGT] lengths. The proposed nomenclature is considered provisional based on future data being available and involvement of IUIS at the appropriate time/level.

Phylogenetic tree construction and gene alignment

Maximum Likelihood trees for each TCR locus and BTN(L)/SKINT proteins were reconstructed with the human, mouse, and naked mole-rat datasets, using the MEGA (Molecular Evolutionary Genetics Analysis) software tool (version 11.0.13)⁷⁹. More specifically, the TCR loci phylogenetic trees were based on nucleotide multiple sequence alignments (MSA), generated with representatives of V-REGION sequences from each subgroup, with the exception of the highly degenerate pseudogenes (TRAVA, TRBVA, TRAVB, etc). The MSA was performed with the online multiple sequence alignment program MAFFT (version 7)⁸⁰, with default parameters, featuring automatic algorithm selection, a 200PAM/K=2 scoring matrix, a gap opening penalty of -1.53, and a gap extension penalty of 0.0, to achieve high alignment accuracy and consistency. The tree visualization was performed using iTOL v6⁸¹, and the different human subgroups were represented with distinct colors.

The BTN(L)/SKINT trees were reconstructed using MAFFT (version 7) with amino acid multiple sequence alignments of the BTN(L) and SKINT proteins extracted from NCBI. Default parameters were applied in MAFFT, using automatic algorithm selection, the BLOSUM62 scoring matrix, a gap opening penalty of -1.53, and a gap extension penalty of 0.0. A single representative sequence for each

protein family was selected for inclusion in the analysis. The tree was visualized with iTOL v6 and the colors were assigned manually indicating the orthologous proteins identified using the NCBI orthology database.

For the alignment and comparison of *MRI* genes, the Clustal Omega multiple sequence alignment package (version 1.2.4)⁷¹ was used, with default parameters, utilizing the Gonnet transition matrix, a gap opening penalty of 6 bits and a gap extension penalty of 1 bit for the gene sequence alignment. The results visualization was achieved with Jalview (version 2.11.3.2)⁸². In the alignment of invariant CDR3 γ regions from human, mouse, and naked mole-rat Geneious Prime (version 2023.1.2) was used to group sequence by similarity using Clustal Omega method (version 1.2.4).

Animals and sample collection

Tissue samples (blood, spleen, and cervical/thoracic thymi) were collected from animals at different developmental stages: neonates 1st day of birth, young-adult animals (around 3 years old), and middle-aged adults (around 11 years old). The primary factor motivating the choice of spleen and blood for investigation of the TCR repertoire in peripheral tissues was the feasibility of obtaining these tissues in the naked mole-rat at various developmental timepoints, including neonates. The cervical and thoracic thymi were included in the study as they are the T cell-generating organs in the naked mole-rat. All neonates in the study were littermates, while young adults were sourced from three different litters: adults 1, 2, and 3 were littermates, adults 5 and 6 were also littermates, and adult 4 was unrelated to the others (IDs correspond to Supplementary Table 1). The only middle-aged animal (Adult 7) was from an independent litter. After collection tissue samples were lysed in RLT + buffer and stored at -80°C for further analyses. Ethical and legal approval was obtained by the University of Rochester Committee on Animal Resources (UCAR), and animal experiments were performed in accordance with guidelines instructed by UCAR with protocol numbers 2009-054 (naked mole rat) and 2017-033 (mouse). Naked mole rats were housed at the University of Rochester colonies, breeding and detail described elsewhere⁸³.

Bulk TCR-sequencing and analysis

RNA was isolated from tissues single-cell suspensions with the RNeasy Micro Kit (Qiagen). cDNA was generated performing a template switch anchored RT-PCR. RNA was reverse transcribed via a template switch cDNA reaction using in the same tube **1**) a common TRGC primer with degenerate nucleotides specific for all the possible functional TRGC genes identified with IMGT annotation rules (5'-GAAAGRTATGTTC-CAGCCTY-3'), **2**) a TRDC primer (5'-GAGACAAGCAACCTTTGTTCCA TT-3') to amplify CDR3 δ sequences, **3**) a template-switch adapter (5'-AAGCAGTGGTATCAACGCAGAGTACATrGrGrG-3') and **4**) the SuperScript II RT enzyme (Invitrogen). For CDR3 α and CDR3 β sequences, we used oligodT primer (5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3') instead of specific locus primers to generate cDNA. The cDNA was then purified using AMPure XP beads (Agencourt). Subsequent amplification of the TRG and TRD region was achieved using a combination of TRGC specific primers including **1**) a TRGC2 primer (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAAGGGCTTGGGGAAATGTT-3'), **2**) a TRGC3 primer (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GCTTGGGGAAATGTCTGTGTC-3') and **3**) a TRGC5 primer (5'-GTCTC GTGGGCTCGGAGATGTGTATAAGAGACAGTGGGGAAATGTCTCCAT CTG-3'). These combination of TRG primers was mixed as well with a specific TRDC primer (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGA-GACAGTGGTTTGAAGGAGGCTGAAT-3', adapter in italic) and a primer complementary to the template-switch adapter (5'-TCGTCGGC ACGTCAGATGTGTATAAGAGACAGAAGCAGTGGTATCAACGCAG-3', adapter in italic) with the KAPA Real-Time Library Amplification Kit (Kapa Biosystems). In the case of TRA and TRB sequence the KAPA enzyme and the primer complementary to the template-switch

adapter were combined with a TRAC specific primer (5'-GTCTC GTGGGCTCGGAGATGTGTATAAGAGACAGCTTTCAGCTGGTACACGG CA-3') and a TRBC specific primer (5'-GTCTCGTGGGCTCGGAGATG TGTATAAGAGACAGATGGCTCAAACACAGCCACC-3'). After purification with AMPure XP beads, an index PCR with Illumina sequencing adapters was performed using the Nextera XT Index Kit. This second PCR product was again purified with AMPure XP beads. HTS of the generated amplicon products was performed on an Illumina MiSeq platform using the V2 300 kit, with 300 bp at the 39 end (read 2) and 150 bp at the 59 end (read 1) (at the GIGA center, University of Liège, Liège, Belgium). Default parameters were used except to assemble TRDD gene where three instead of five consecutive nucleotides were applied as assemble parameter. Raw sequencing reads from fastq files (read 1 and read 2) were aligned to the constructed IMGT reference for V, D, and J genes from naked mole-rat (based on the long-read chromosome assemblies as described under "TCR locus annotation") using MiXCR software version 4.1.2⁸⁴. Default parameters were used except to assemble TRDD gene where three instead of five consecutive nucleotides were applied as assemble parameter. CDR3 sequences were then exported and analyzed using VDJtools software version 1.2.1 using default settings⁸⁵. Sequences out of frame and containing stop codons were excluded from the analysis. The degree of TCR repertoire overlap between two different samples was analyzed using the overlap F metrics. The *F* value is the geometric mean of relative overlap frequencies and is thus calculated as follows: $F_{ij} = \sqrt{f_{ij}f_{ji}}$; where $f_{ij} = \sum_{k=1}^N \delta_{ik}$ is the total frequency of clonotypes that overlap between samples *i* and *j* in sample *i*. Tree maps were created using the Treemap Package (version 2.4–3). Visualization of gene usage, public clonotype frequencies, and clonality were generated using Immunarch R package (version 1.0.0)⁸⁶. Hierarchical clustering based on V gene usage was done based on Jensen–Shannon divergence. Visualization of gene usage, public clonotype frequencies and clonality were generated using Immunarch R package (version 0.9.0)⁸⁶. Hierarchical clustering based on V gene usage was done based on Jensen–Shannon divergence.

TCR extraction from public single-cell RNA-seq datasets

In order to construct single cell TCR-seq libraries from public naked mole-rat single cell RNA-seq data^{6,7,41} sequencing fastq files were processed using TRUST4 software (version 1.0.10)⁸⁷. Naked mole-rat TCR reference from IMGT was used as -f and --ref parameters when using "run-trust4" function. TRUST4 output in AIRR format was used as input for Scirpy package (version 0.13.0)⁸⁸ to re-construct single cell paired TCR repertoires based on barcode sequences.

Single-cell RNA-seq data processing

Publicly available processed Seurat objects (with qc analysis already performed) from single cell RNA-seq data were downloaded from the corresponding public repositories^{6,7,41} (available in the GEO database under accession codes [GSE185721](#), [GSE202903](#) and [GSE132642](#)). The animals used in these public single cell RNA-seq repositories are different animals, derived from different litters, than the ones used for the bulk TCR experiments described in the current study. All downstream single cell analyses were performed using the R package *Seurat* version 4.3.0⁸⁹. Single cell TCR libraries were added as metadata to the Seurat objects. In all 3 datasets T-cell clusters were subset based on the annotation obtained from the original studies and subsequent analysis were performed on them. In the case of blood data due to the high TCR-extraction efficiency of the TRUST4 pipeline, T-cell clusters were further subset based on the presence of TCR transcripts containing a rearranged CDR3 sequence. In all cases after subset of data principal components (PCs) were calculated using "RunPCA" and 30 dimensions were chosen as input for "RunUMAP" function. UMAP representation was used to generate bidimensional coordinates for each cell. The

k-nearest neighbors of each cell were determined using the ‘Find-Neighbors’ function and this knn graph was used to construct the shared nearest neighbor (SNN) graph by calculating the neighborhood overlap (Jaccard index) between every cell and its k.param nearest neighbors. Finally, the ‘FindClusters’ function was used to cluster cells using the Louvain algorithm based on the same PCs as RunUMAP function. Differential gene expression analysis comparing gene expression of each cluster to all the others was performed by the ‘FindAll-Markers’ function using Wilcoxon-Rank sum test method (two-sided). DEGs were selected based on a log₂-fold change (logFC) ≥ 0.2 difference between the average gene expression of the clusters, a percentage of expression equal to or greater than 15% in at least one test cluster (min.pct ≥ 0.15), a difference equal to or greater than 15% in the fraction of detection between the two groups (min.diff.pct ≥ 0.15) and adjusted $p < 0.05$ (based on Bonferroni correction using all genes in the dataset). Diversity and clonality analysis of single cell TCR-seq data were done thanks to *scRepertoire* package (version 1.11.0)⁹⁰.

Hypergeometric test analysis

Differential expressed genes from blood NK-like effector cluster were used as input for enrich function of *enrichR* package (version 3.0.1)⁹¹. Tabula Muris gene sets were used as reference to perform the comparative test.

Statistical analysis

Statistical analyses were performed using GraphPad Prism 8.0.2 software or with Immunarch package depending on the nature of the data. Two-tailed Student’s *t* test were used for normally distributed data; Mann-Whitney U and Wilcoxon matched paired tests were used for nonparametric data. Differences between more than two groups were analyzed using Kruskal–Wallis and Dunn’s post-tests for nonparametric data.

Data availability

The raw fastq sequencing data of TRA, TRB, TRG, and TRD bulk TCR repertoires data generated in this study have been deposited in the in the SRA public database under accession code [PRJNA998815](https://www.ncbi.nlm.nih.gov/sra/PRJNA998815). Source data are provided with this paper.

References

- Seluanov, A., Gladyshev, V. N., Vijg, J. & Gorbunova, V. Mechanisms of cancer resistance in long-lived mammals. *Nat. Rev. Cancer* **18**, 433–441 (2018).
- Tian, X. et al. High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature* **499**, 346–349 (2013).
- Hadi, F. et al. Transformation of naked mole-rat cells. *Nature* **583**, E1–E7 (2020).
- Hadi, F., Smith, E. S. J. & Khaled, W. T. Naked mole-rats: resistant to developing cancer or good at avoiding it? *Adv. Exp. Med Biol.* **1319**, 341–352 (2021).
- Zhao, J. et al. Reply to: transformation of naked mole-rat cells. *Nature* **583**, E8–E13 (2020).
- Hilton, H. G. et al. Single-cell transcriptomics of the naked mole-rat reveals unexpected features of mammalian immunity. *PLoS Biol.* **17**, e3000528 (2019).
- Emmrich, S. et al. Ectopic cervical thymi and no thymic involution until midlife in naked mole rats. *Aging Cell* **20**, e13477 (2021).
- Papadopoulou, M., Sanchez Sanchez, G. & Vermijlen, D. Innate and adaptive $\gamma\delta$ T cells: How, when, and why. *Immunol. Rev.* **298**, 99–116 (2020).
- Hayday, A. C. Integrative immunology: gamma delta T cells. in *Paul’s Fundamental Immunology*, 1748 (Wolters Kluwer, 2022).
- Morrissey, K. A. et al. Comparison of reptilian genomes reveals deletions associated with the natural loss of $\gamma\delta$ T cells in squamates. *J. Immunol.* **208**, 1960–1967 (2022).
- Hirano, M. et al. Evolutionary implications of a third lymphocyte lineage in lampreys. *Nature* **501**, 435–438 (2013).
- Herrmann, T., Karunakaran, M. M. & Fichtner, A. S. A glance over the fence: using phylogeny and species comparison for a better understanding of antigen recognition by human $\gamma\delta$ T-cells. *Immunol. Rev.* **298**, 218–236 (2020).
- Boehme, L., Roels, J. & Taghon, T. Development of $\gamma\delta$ T cells in the thymus—a human perspective. *Semin. Immunol.* **61–64**, 101662 (2022).
- Vermijlen, D. & Prinz, I. Ontogeny of innate T lymphocytes—some innate lymphocytes are more innate than others. *Front. Immunol.* **5**, 486 (2014).
- Sanchez Sanchez, G., Tafesse, Y., Papadopoulou, M. & Vermijlen, D. Surfing on the waves of the human $\gamma\delta$ T cell ontogenic sea. *Immunol. Rev.* **315**, 89–107 (2023).
- Benveniste, P. M. et al. Generation and molecular recognition of melanoma-associated antigen-specific human $\gamma\delta$ T cells. *Sci. Immunol.* **3**, eaav4036 (2018).
- Vermijlen, D., Gatti, D., Kouzeli, A., Rus, T. & Eberl, M. $\gamma\delta$ T cell responses: how many ligands will it take till we know? *Semin Cell Dev. Biol.* **84**, 75–86 (2018).
- Deseke, M. & Prinz, I. Ligand recognition by the $\gamma\delta$ TCR and discrimination between homeostasis and stress conditions. *Cell Mol. Immunol.* **17**, 914–924 (2020).
- Willcox, B. E. & Willcox, C. R. $\gamma\delta$ TCR ligands: the quest to solve a 500-million-year-old mystery. *Nat. Immunol.* **20**, 121–128 (2019).
- Harly, C. et al. Human $\gamma\delta$ T cell sensing of AMPK-dependent metabolic tumor reprogramming through TCR recognition of EphA2. *Sci. Immunol.* **6**, eaba9010 (2021).
- Herrmann, T. & Karunakaran, M. M. Butyrophilins: $\gamma\delta$ T cell receptor ligands, immunomodulators and more. *Front. Immunol.* **13**, 876493 (2022).
- McKenzie, D. R. et al. Normality sensing licenses local T cells for innate-like tissue surveillance. *Nat. Immunol.* **23**, 411–422 (2022).
- Hayday, A. C. Gammadelta T cells and the lymphoid stress-surveillance response. *Immunity* **31**, 184–196 (2009).
- Gober, H.-J. et al. Human T cell receptor gammadelta cells recognize endogenous mevalonate metabolites in tumor cells. *J. Exp. Med.* **197**, 163–168 (2003).
- Silva-Santos, B., Mensurado, S. & Coffelt, S. B. $\gamma\delta$ T cells: pleiotropic immune effectors with therapeutic potential in cancer. *Nat. Rev. Cancer* **19**, 392–404 (2019).
- Mensurado, S., Blanco-Domínguez, R. & Silva-Santos, B. The emerging roles of $\gamma\delta$ T cells in cancer immunotherapy. *Nat. Rev. Clin. Oncol.* **20**, 178–191 (2023).
- Kobayashi, H. & Tanaka, Y. $\gamma\delta$ T cell immunotherapy—a review. *Pharm. (Basel)* **8**, 40–61 (2015).
- Manso, T. et al. IMG^T databases, related tools and web resources through three main axes of research and development. *Nucleic Acids Res.* **50**, D1262–D1272 (2022).
- Barbee, S. D. et al. Skint-1 is a highly specific, unique selecting component for epidermal T cells. *Proc. Natl Acad. Sci. USA* **108**, 3330–3335 (2011).
- Di Marco Barros, R. et al. Epithelia use butyrophilin-like molecules to shape organ-specific $\gamma\delta$ T cell compartments. *Cell* **167**, 203–218.e17 (2016).
- Jandke, A. et al. Butyrophilin-like proteins display combinatorial diversity in selecting and maintaining signature intraepithelial $\gamma\delta$ T cell compartments. *Nat. Commun.* **11**, 3769 (2020).
- Rigau, M. et al. Butyrophilin 2A1 is essential for phosphoantigen reactivity by $\gamma\delta$ T cells. *Science* **367**, eaay5516 (2020).
- Karunakaran, M. M. et al. Butyrophilin-2A1 directly binds germline-encoded regions of the V γ 9V δ 2 TCR and is essential for phosphoantigen sensing. *Immunity* **52**, 487–498.e6 (2020).

34. Yuan, L. et al. Phosphoantigens glue butyrophilin 3A1 and 2A1 to activate V γ 9V δ 2 T cells. *Nature* **621**, 840–848 (2023).
35. Willcox, C. R. et al. Phosphoantigen sensing combines TCR-dependent recognition of the BTN3A IgV domain and germline interaction with BTN2A1. *Cell Rep.* **42**, 112321 (2023).
36. Vermijlen, D. et al. Human cytomegalovirus elicits fetal gamma-delta T cell responses in utero. *J. Exp. Med.* **207**, 807–821 (2010).
37. Asarnow, D. M. et al. Limited diversity of gamma delta antigen receptor genes of Thy-1+ dendritic epidermal cells. *Cell* **55**, 837–847 (1988).
38. Itohara, S. et al. Homing of a $\gamma\delta$ thymocyte subset with homogeneous T-cell receptors to mucosal epithelia. *Nature* **343**, 754–757 (1990).
39. Harly, C., Robert, J., Legoux, F. & Lantz, O. $\gamma\delta$ T, NKT, and MAIT cells during evolution: redundancy or specialized functions? *J. Immunol.* **209**, 217–225 (2022).
40. Pellicci, D. G., Koay, H.-F. & Berzins, S. P. Thymic development of unconventional T cells: how NKT cells, MAIT cells and $\gamma\delta$ T cells emerge. *Nat. Rev. Immunol.* **20**, 756–770 (2020).
41. Emmrich, S. et al. Characterization of naked mole-rat hematopoiesis reveals unique stem and progenitor cell patterns and neotenic traits. *EMBO J.* **41**, e109694 (2022).
42. Fiala, G. J., Gomes, A. Q. & Silva-Santos, B. From thymus to periphery: molecular basis of effector $\gamma\delta$ -T cell differentiation. *Immunol. Rev.* **298**, 47–60 (2020).
43. Sanchez Sanchez, G. et al. Identification of distinct functional thymic programming of fetal and pediatric human $\gamma\delta$ thymocytes via single-cell analysis. *Nat. Commun.* **13**, 5842 (2022).
44. Tan, L. et al. A fetal wave of human type 3 effector $\gamma\delta$ cells with restricted TCR diversity persists into adulthood. *Sci. Immunol.* **6**, eabf0125 (2021).
45. Sumaria, N. et al. Cutaneous immunosurveillance by self-renewing dermal $\gamma\delta$ T cells. *J. Exp. Med.* **208**, 505–518 (2011).
46. Yokobori, N. et al. CD3 expression distinguishes two gammadelta T cell receptor subsets with different phenotype and effector function in tuberculous pleurisy. *Clin. Exp. Immunol.* **157**, 385–394 (2009).
47. Jagannathan, P. et al. Loss and dysfunction of V δ 2+ $\gamma\delta$ T cells are associated with clinical tolerance to malaria. *Sci. Transl. Med.* **6**, 251ra117 (2014).
48. Van hede, D. et al. Human papillomavirus oncoproteins induce a reorganization of epithelial-associated $\gamma\delta$ T cells promoting tumor formation. *Proc. Natl Acad. Sci.* **114**, E9056–E9065 (2017).
49. Tieppo, P. et al. The human fetal thymus generates invariant effector $\gamma\delta$ T cells. *J. Exp. Med.* **217**, jem.20190580 (2020).
50. Hale, J. S. & Fink, P. J. Back to the thymus: peripheral T cells come home. *Immunol. Cell Biol.* **87**, 58–64 (2009).
51. Zhang, Y. et al. The role of short homology repeats and TdT in generation of the invariant gamma delta antigen receptor repertoire in the fetal thymus. *Immunity* **3**, 439–447 (1995).
52. Papadopoulou, M. et al. Fetal public V γ 9V δ 2 T cells expand and gain potent cytotoxic functions early after birth. *Proc. Natl Acad. Sci.* **107**, 18638–18648 (2020).
53. Ng, J. W. K. et al. Cord blood-derived V δ 2+ and V δ 2- T cells acquire differential cell state compositions upon in vitro expansion. *Sci. Adv.* **9**, eadf3120 (2023).
54. Papadopoulou, M. et al. TCR sequencing reveals the distinct development of fetal and adult human V γ 9V δ 2 T cells. *J. Immunol.* **203**, 1468–1479 (2019).
55. Perriman, L. et al. A three-stage developmental pathway for human V γ 9V δ 2 T cells within the postnatal thymus. *Sci. Immunol.* **8**, eabo4365 (2023).
56. Fichtner, A. S. et al. Alpaca (Vicugna pacos), the first nonprimate species with a phosphoantigen-reactive V γ 9V δ 2 T cell subset. *Proc. Natl Acad. Sci.* **117**, 6697–6707 (2020).
57. Fichtner, A. S., Karunakaran, M. M., Starick, L., Truman, R. W. & Herrmann, T. The armadillo (Dasypus novemcinctus): a witness but not a functional example for the emergence of the butyrophilin 3/V γ 9V δ 2 system in placental mammals. *Front Immunol.* **9**, 265 (2018).
58. Zhang, X. M., Tonnel, C., Lefranc, M. P. & Huck, S. T cell receptor gamma cDNA in human fetal liver and thymus: variable regions of gamma chains are restricted to V gamma I or V9, due to the absence of splicing of the V10 and V11 leader intron. *Eur. J. Immunol.* **24**, 571–578 (1994).
59. Kazen, A. R. & Adams, E. J. Evolution of the V, D, and J gene segments used in the primate gammadelta T-cell receptor reveals a dichotomy of conservation and diversity. *Proc. Natl Acad. Sci. USA* **108**, E332–E340 (2011).
60. Antonacci, R. et al. Evolution of the T-cell receptor (TR) loci in the adaptive immune response: the tale of the TRG locus in mammals. *Genes* **11**, 624 (2020).
61. Ma, L. et al. Effector V γ 9V δ 2 T cell response to congenital *Toxoplasma gondii* infection. *JCI Insight* **6**, e138066 (2021).
62. Dauby, N. & Marchant, A. Fetal Infections: Immune Response to Infections during Fetal Life. In *Fetal Therapy: Scientific Basis and Critical Appraisal of Clinical Benefits* (eds Johnson, A., Oepkes, D. & Kilby, M. D.) 215–223 (Cambridge University Press, Cambridge, 2020).
63. Yazawa, R. et al. Functional adaptive diversity of the Atlantic salmon T-cell receptor gamma locus. *Mol. Immunol.* **45**, 2150–2157 (2008).
64. Izaelsson, M. et al. Distinct organization of adaptive immunity in the long-lived rodent Spalax galili. *Nat. Aging* **1**, 179–189 (2021).
65. Xu, C. et al. Expansion of MAIT cells in the combined absence of NKT and $\gamma\delta$ -T cells. *Mucosal Immunol.* **16**, 446–461 (2023).
66. Kitts, P. A. et al. Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res.* **44**, D73–D80 (2016).
67. Pégorier, P. et al. IMGT® biocuration and comparative analysis of Bos Taurus and Ovis aries TRA/TRD loci. *Genes* **12**, 30 (2020).
68. Giudicelli, V. & Lefranc, M.-P. IMGT-ONTOLOGY 2012. *Front. Genet.* **3**, 79 (2012).
69. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
70. Brochet, X., Lefranc, M.-P. & Giudicelli, V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* **36**, W503–W508 (2008).
71. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
72. Lemoine, F. et al. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* **47**, W260–W265 (2019).
73. Lefranc, M.-P. Immunoglobulin and T cell receptor genes: IMGT(®) and the birth and rise of immunoinformatics. *Front Immunol.* **5**, 22 (2014).
74. Giudicelli, V. et al. IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res* **34**, D781–D784 (2006).
75. Giudicelli, V., Chaume, D. & Lefranc, M.-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res* **33**, D256–D261 (2005).
76. Kaas, Q., Ruiz, M. & Lefranc, M.-P. IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucleic Acids Res* **32**, D208–D210 (2004).
77. Alamyar, E., Duroux, P., Lefranc, M.-P. & Giudicelli, V. IMGT(®) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol. Biol.* **882**, 569–604 (2012).

78. Ehrenmann, F., Kaas, Q. & Lefranc, M.-P. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res* **38**, D301–D307 (2010).
79. Tamura, K., Stecher, G. & Kumar, S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
80. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evolution* **30**, 772–780 (2013).
81. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293–W296 (2021).
82. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
83. Ke, Z., Vaidya, A., Ascher, J., Seluanov, A. & Gorbunova, V. Novel husbandry techniques support survival of naked mole rat (*Heterocephalus glaber*) pups. *J. Am. Assoc. Lab Anim. Sci.* **53**, 89–91 (2014).
84. Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
85. Shugay, M. et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).
86. Samokhina, M. et al. immunomind/immunarch: Immunarch 0.9.0. Zenodo <https://doi.org/10.5281/zenodo.7446955> (2022).
87. Song, L. et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* **18**, 627–630 (2021).
88. Sturm, G. et al. Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics* **36**, 4817–4818 (2020).
89. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
90. Borcherding, N., Bormann, N. L. & Kraus, G. scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Res* **9**, 47 (2020).
91. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

Acknowledgements

This work was supported by an “Actions Blanches” grant (ULB; DV), a WELBIO investigator program (WELBIO-CR-2022 A– 15, D.V.) and by US National Institutes of Health grants to A.S. and V.G. G.S.S. is supported by Télévie/FNRS, WELBIO and “Actions Blanches”. We thank Isabel Kuppens to bring the naked mole-rat to our attention. We thank Alexandre Albani for contribution to the biocuration of the naked mole-rat loci, Veronique Giudicelli, Nika Abdollahi for repertoire analysis and

discussions, Geraldine Folch and Joumana Michaloud for data validations as well as the whole IMGT team.

Author contributions

G.S.S. designed and performed experiments, analyzed data, and wrote the manuscript; S.E. designed and performed experiments; M.G., A.P., and S.K. performed analysis of genomic data; V.G. and A.S. provided samples and resources; D.V. designed and supervised the research, provided resources, analyzed data and wrote the manuscript; all co-authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-48652-z>.

Correspondence and requests for materials should be addressed to David Vermijlen.

Peer review information *Nature Communications* thanks Thomas Herrmann, Thomas Hildebrandt, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024