

Evolution of T cells in the cancer-resistant naked mole-rat

Tzuhua D. Lin^{1,†}, Nimrod D. Rubinstein^{1,†,*}, Nicole L. Fong¹, Megan Smith¹, Wendy Craft¹, Baby
Martin-McNulty¹, Rebecca Perry³, Martha A. Delaney², Margaret A. Roy¹, Rochelle Buffenstein^{1,3,*}

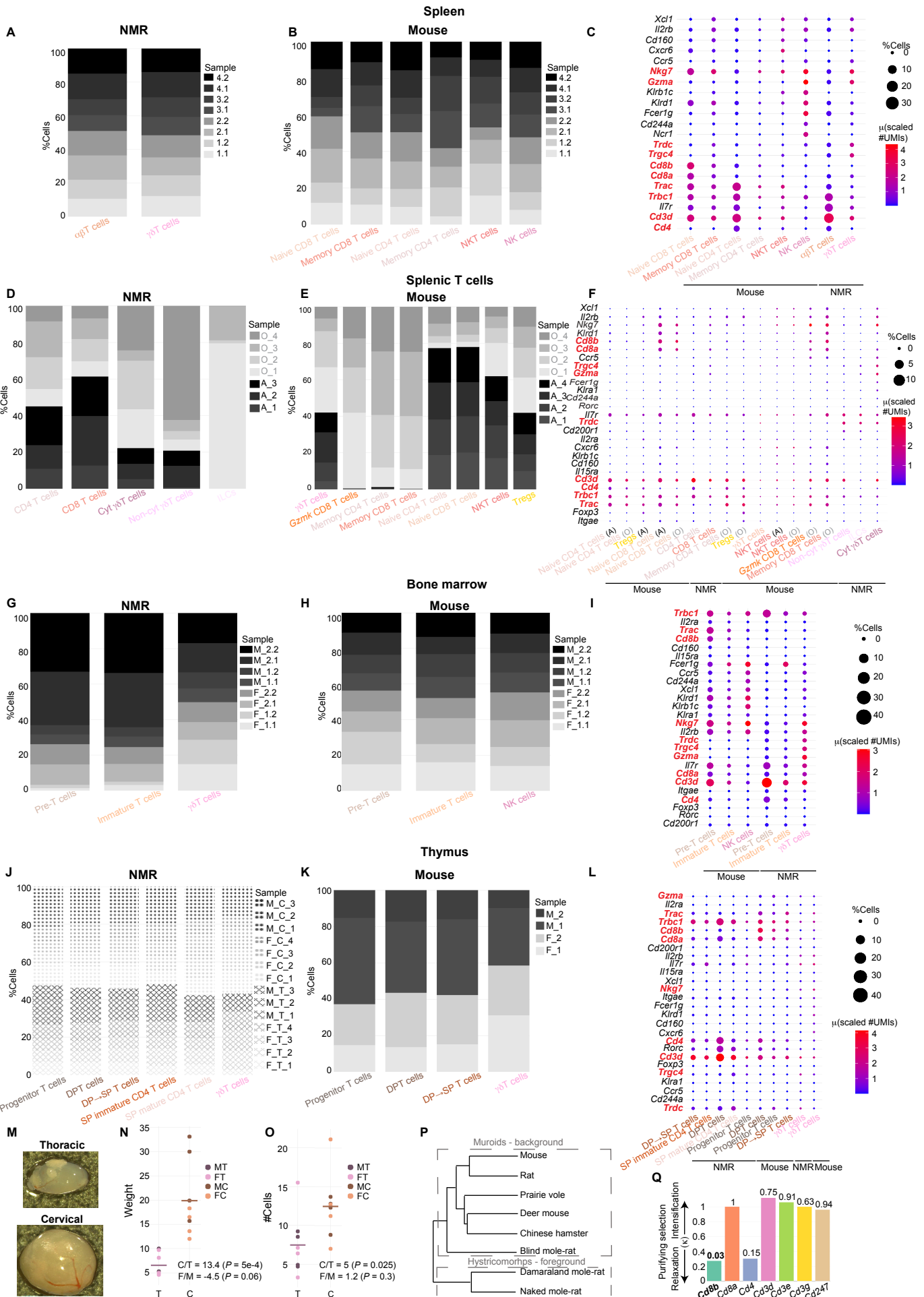


Fig. S1: ScRNA-seq of circulating T-cell subsets reveals that the NMR has a large population of $\gamma\delta$ T cells

Bar-charts of the cell proportions assigned to each NK and T -cell subset from spleens of (A) four NMRs and (B) four mice (two technical replicates for each sample in each species). Bars are relatively uniformly represented by all samples; (C) gene-by-cell-subset heatmap of the NK and T -cell subsets of the two species. Bar-charts of the cell proportions assigned to each T-cell subset from spleens of (D) three adult and four old NMRs and (E) four adult and four old mice. NMR ILC and the mouse CD4 and CD8 subsets are non-uniformly represented by the samples; (F) gene-by-cell-subset heatmap of the T-cell subsets of the two species. Bar-charts of the cell proportions assigned to each T-cell subset from bone marrows of (G) two male and two female NMRs and (H) two male and two female mice (two technical replicates for each sample in each species). Bars are relatively uniformly represented by all samples; (I) gene-by-cell-subset heatmap of the T-cell subsets of the two species. Bar-charts of the cell proportions assigned to each T-cell subset from thymuses of (J) seven NMRs (thoracic and cervical) and (K) four mice (thoracic). Bars are relatively uniformly represented by all samples; (L) gene-by-cell-subset heatmap of the T-cell subsets of the two species. Heatmaps are color-coded by expression levels and diameter-coded by the proportion of the gene-expressing cells in the cell subset. Genes identifying $\gamma\delta$ T-cell subsets are labeled in red. Comparison between the NMR thoracic and cervical thymic tissues showing that the cervical tissue is larger in (M) size; (N) weight; and (O) number of cells. Quantification of the intensity of purifying selection operating on (P) hystricomorph relative to muroid (Q) T-cell co-receptor genes, revealing the significant relaxation of purifying selection in the hystricomorph *Cd8b*. Values above bars are multiple-hypothesis-adjusted *p* values of the hystricomorph vs. muroid relaxation in purifying selection (RELAX; Methods). **Abbreviations:** A – Adult; C – Cervical; Cyt – Cytotoxic; DP – Double Positive; F – Female; M – Male; NK – Natural Killer; NKT – Natural Killer T; Non-cyt – Non-cytotoxic; O – Old; P – P-value; SP – Single Positive; T – Thoracic. Source data are provided as source data file.

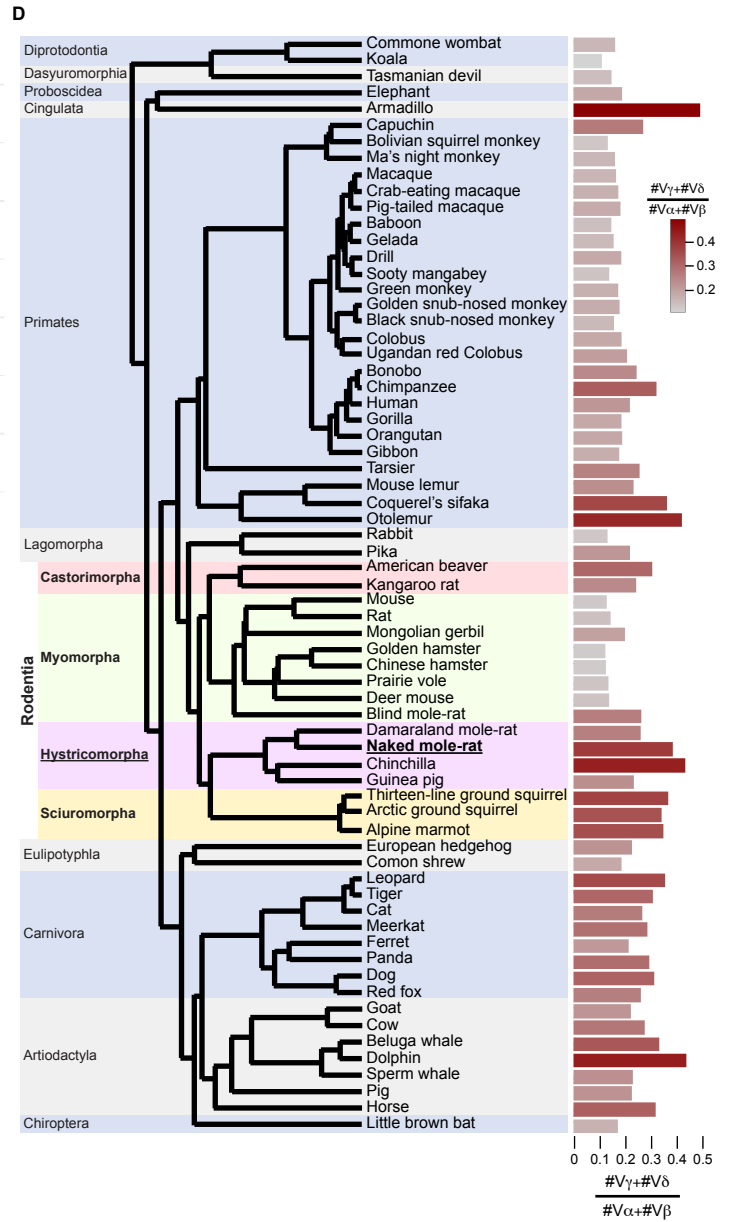
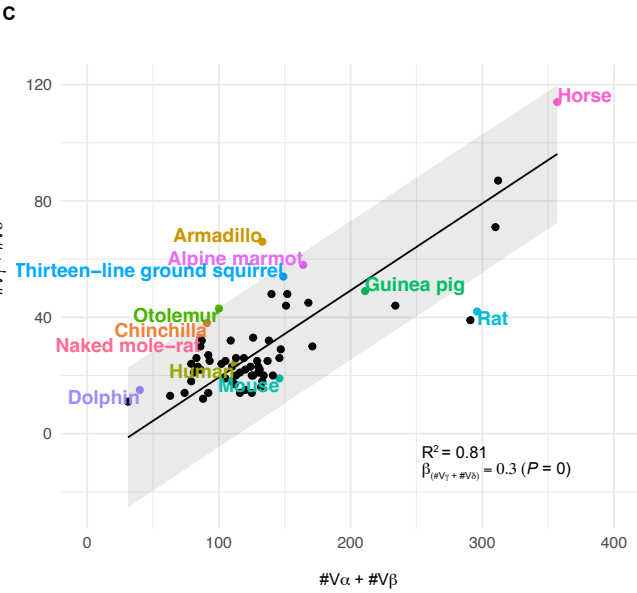
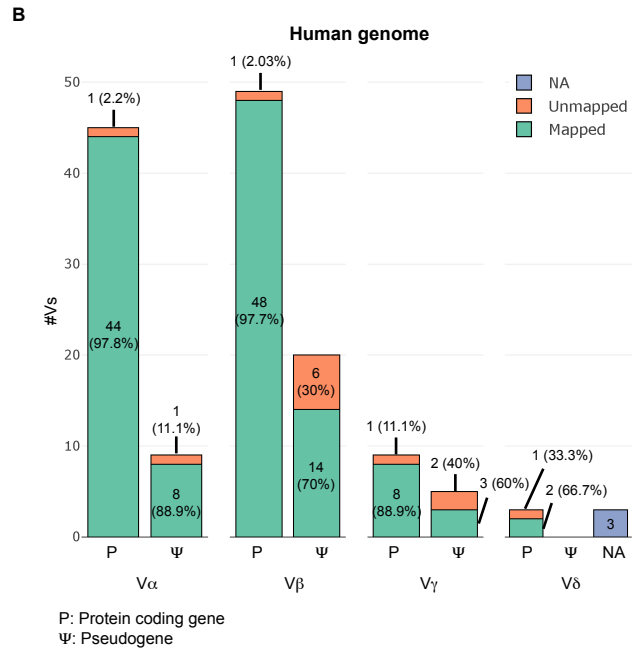
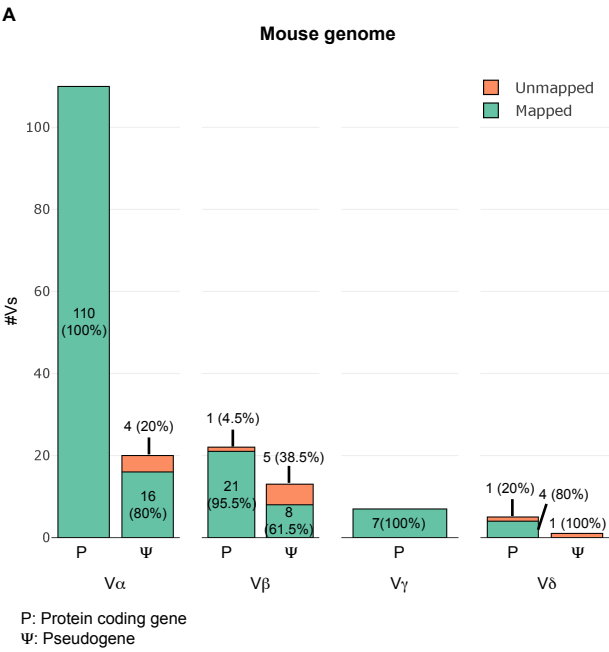
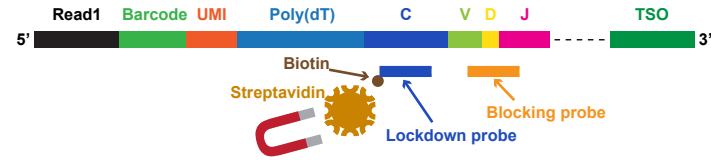


Fig. S2: Mammalian phyletic pattern of TCR regions reveals large diversity of γ and δ , relative to α and β , variable regions in the NMR genome

Benchmarking the detection of variable TCR regions in 67 mammalian genomes using the **(A)** mouse and **(B)** human genomes and their corresponding annotations where each figure shows the number annotated variable regions, from all four TCR loci, that were mapped, unmapped, and uniquely found (NA) by our approach (Methods), revealing a low false detection rate; **(C)** the association between the sum of γ and δ and the sum of α and β variable TCR regions and the strong fit of the phylogenetic least squares model to these data, which is why the diversity of γ and δ variable TCR regions relative to α and β variable TCR regions is more robustly expressed as the residuals of that fit than **(D)** the ratio of the two displayed to the side of the phyletic pattern of 67 mammalian genomes (Methods), color-coded by orders, where the Rodentia order is further color-coded by suborder. **Abbreviations:** V – Variable. Source data are provided as source data file.

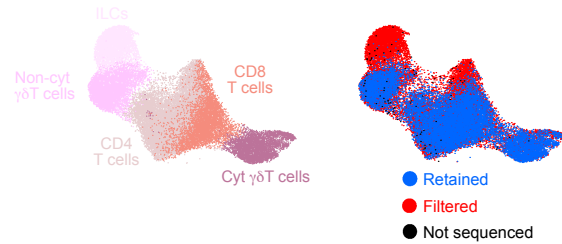
A

3'-barcoded T-cell receptor cDNA

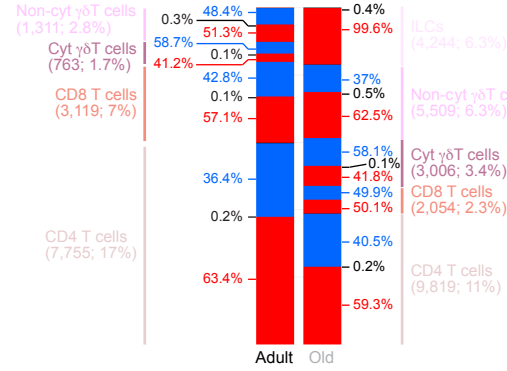


B

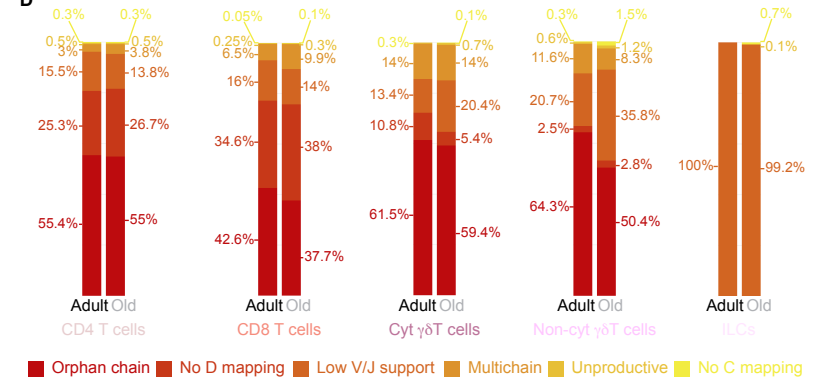
Naked mole-rat



C

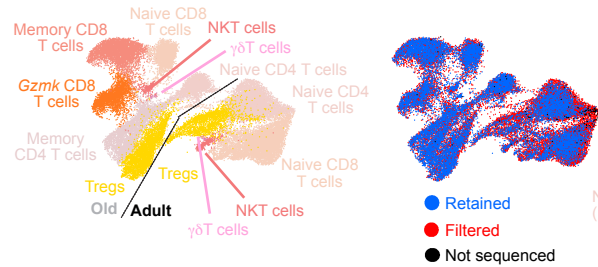


D

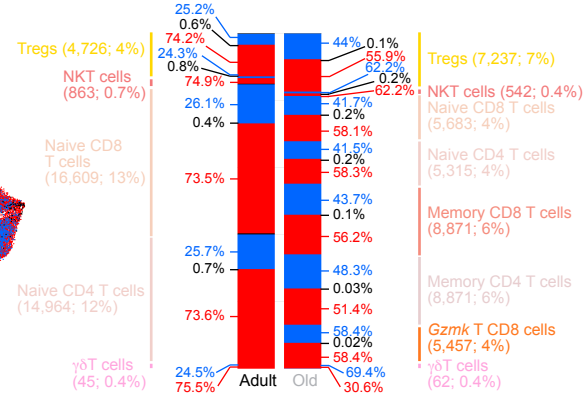


E

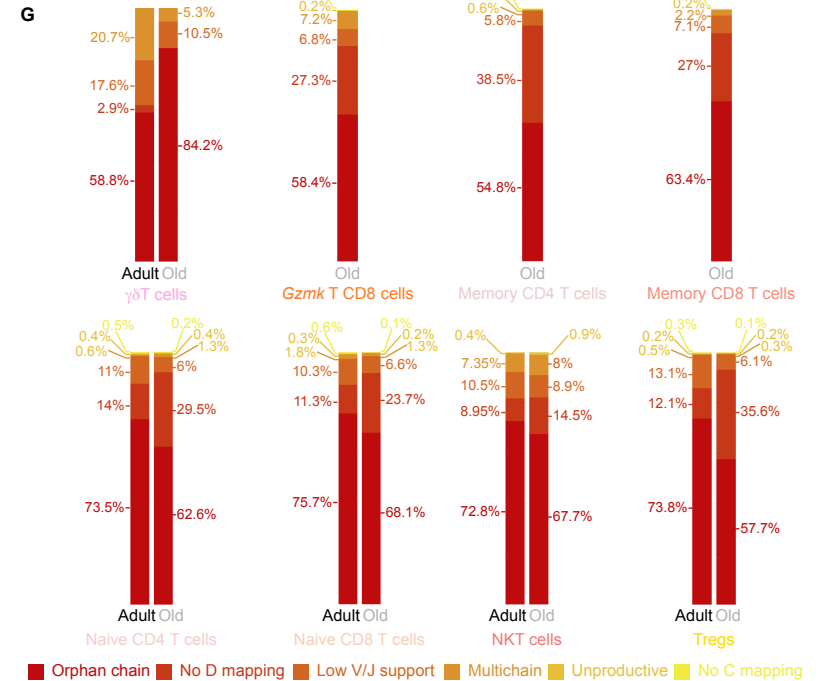
Mouse



F



G



H

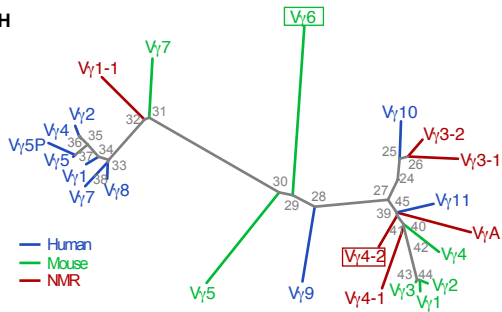


Fig. S3: Single-cell TCR sequencing reveals surprising features of clonotypic diversity in naked mole-rat T-cell subsets

(A) An illustration of the hybridization-capture approach that we developed in order to obtain full-length single-cell TCR sequences utilizing previously generated 10× 3'-barcoded samples (Methods); **(B)** UMAP projections of T cells from spleens of three adult and four old NMRs. Each point represents a cell, the left UMAP is color-coded by T-cell subset and the right UMAP is color-coded by whether their corresponding hybridization-capture sequenced reads were retained, filtered, or not sequenced at all; the stacked bar-charts in **(C)** provide these retention/filtering proportions per each T-cell subset; the stacked bar-charts in **(D)** break down the proportions of filtered hybridization-capture sequenced reads by either one of the six filtering criteria we used (Methods) per each T-cell subset; **(E-G)** are figures corresponding to Fig. S3B-D, for T cells from spleens from the four adult and four old mice; **(H)** phylogenetic reconstruction of the mammalian variable γ TCR regions, pruned to only show the human, mouse, and NMR leaves (Methods). Bootstrap support values (%) are labeled at each split and the boxed leaves are the variable γ TCR regions that comprise the dominant clonotypes found in the NMR and mouse $\gamma\delta$ T-cell subsets. **Abbreviations:** C – Constant; Cyt – Cytotoxic; D – Diversifying; J – Joining; Non-cyt – Non-cytotoxic. Source data are provided as source data file.

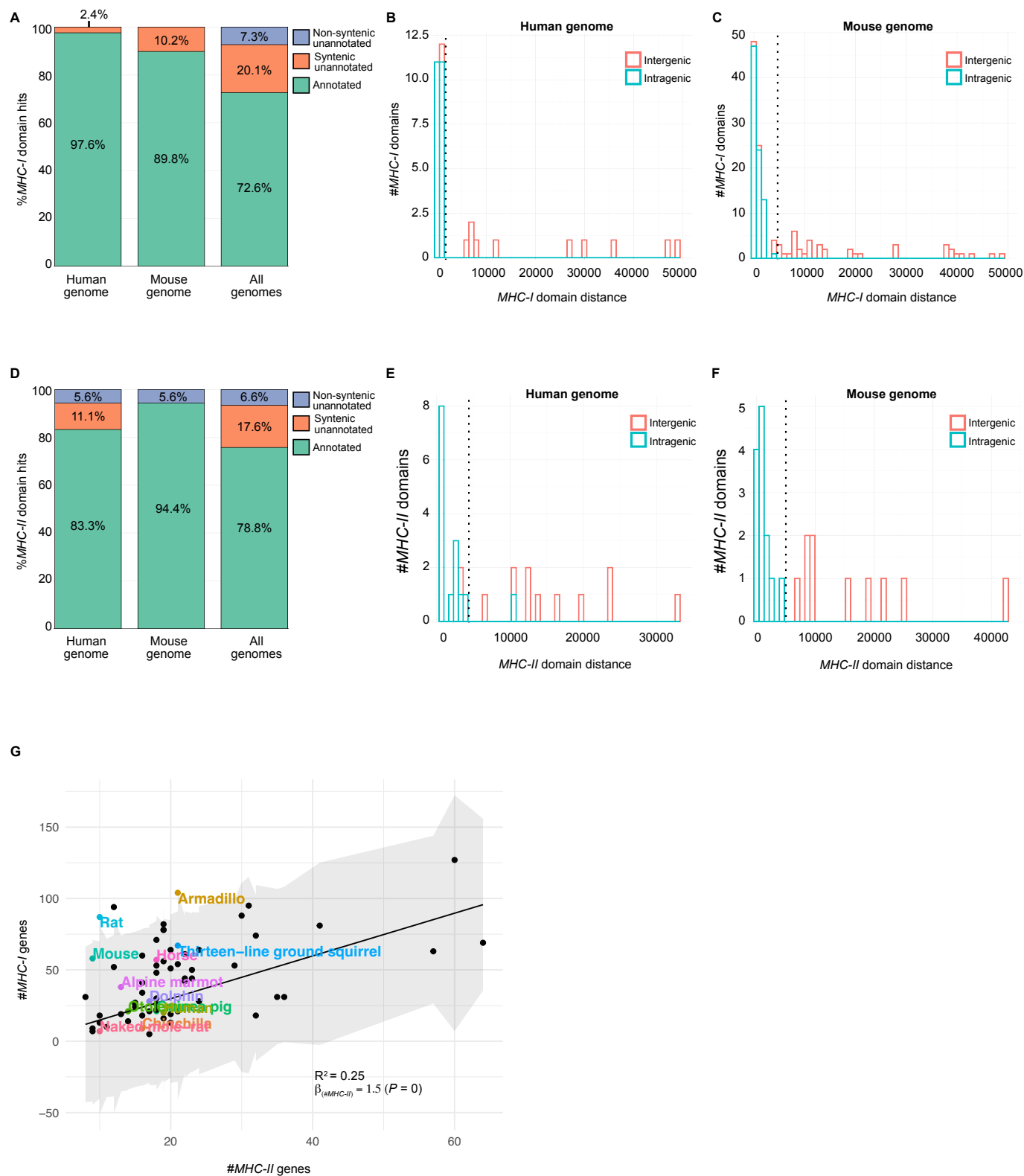


Fig. S4: Mammalian phyletic pattern of *MHC* gene families reveals a small *MHC-I*, relative to *MHC-II*, gene-family size in the naked mole-rat genome

Benchmarking the detection of *MHC-I* protein-coding and pseudo -genes, and unannotated putative pseudogenes in 67 mammalian genomes. **(A)** stacked bar-chart showing the proportion of *MHC-I* domain hits that are annotated, unannotated but are syntenic to annotated *MHC-I* genes, and unannotated and not syntenic to annotated *MHC-I* genes, where the latter two categories form an upper bound to the false positive rate of our approach (Methods).

Distributions of genomic distances between *MHC-I* domains, either of the same gene or genomically adjacent genes, in the **(B)** human and **(C)** mouse genomes (using only annotated *MHC-I* genes). Vertical dotted black lines represent the genomic distance cutoff below which any pair of genomically adjacent *MHC-I* domain hits were assigned to the same *MHC-I* gene.

(D-F) figures corresponding to (A-C) for *MHC-II* genes; **(G)** the association between the number of *MHC-I* and *MHC-II* genes and the weak fit of the phylogenetic least squares model to these data, which is why the ratio between the two was used to express the relative diversity of *MHC-I* (Fig. 4B), rather than the residuals of the model, as applied to the variable TCR regions.

Abbreviations: P – P-value. Source data are provided as source data file.

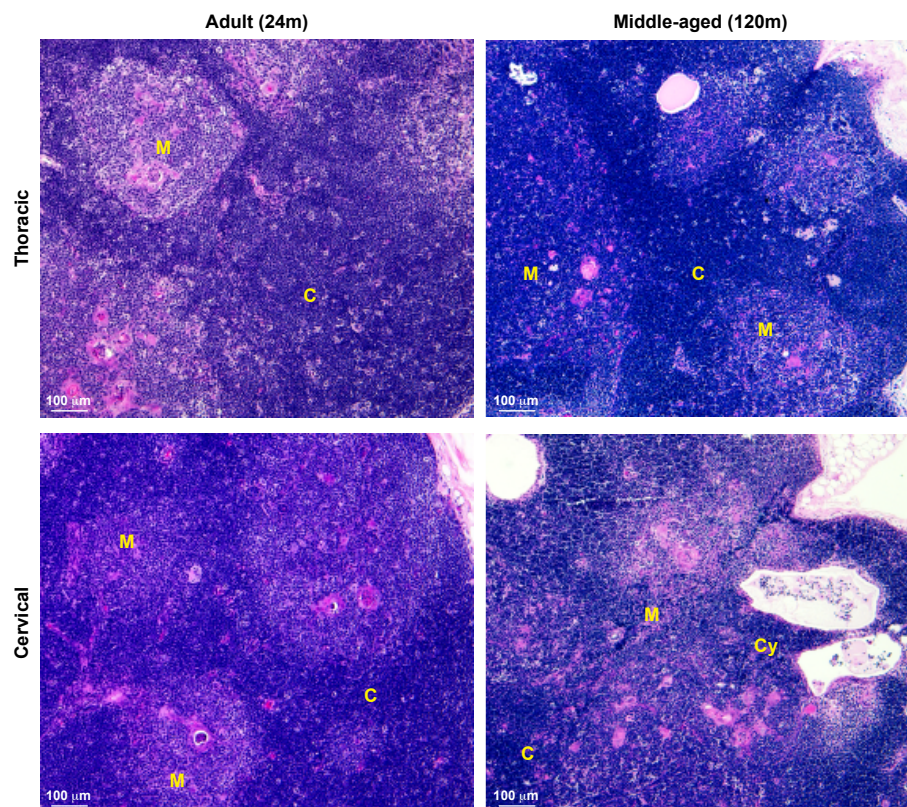


Fig. S5: Thoracic and cervical middle-age NMR thymuses show equivalent microscopic signs indicative of involution

Images of hematoxylin and eosin (H&E) stained sections from thoracic (top) and cervical (bottom) thymic tissues of adult (left) and middle-aged (right) NMRs, showing that at middle age, both thymic tissues present similar signs of indicative of involution, including reduced contrast between cortex and medulla and formation of cysts. **Abbreviations:** C – Cortex; Cy – Cyst; M – Medulla.