

## EVOLUTIONARY BIOLOGY

## Cathelicidin antimicrobial peptides mediate immune protection in marsupial neonates

Jongbeom Park<sup>1†</sup>, Wenfan Ke<sup>1,2</sup>, Aellah Kaage<sup>1‡</sup>, Charles Y. Feigin<sup>1§</sup>, Aaron H. Griffing<sup>1,3</sup>, Yuri Pritykin<sup>2,4</sup>, Mohamed S. Donia<sup>1\*</sup>, Ricardo Mallarino<sup>1\*</sup>

Marsupial neonates are born with immature immune systems, making them vulnerable to pathogens. While neonates receive maternal protection, they can also independently combat pathogens, although the mechanisms remain unknown. Using the sugar glider (*Petaurus breviceps*) as a model, we investigated immunological defense strategies of marsupial neonates. Cathelicidins—a family of antimicrobial peptides expanded in the genomes of marsupials—are highly expressed in developing neutrophils. Sugar glider cathelicidins reside in two genomic clusters, and their coordinated expression is achieved by enhancer sharing within clusters and long-range physical interactions between clusters. Functionally, cathelicidins modulate immune responses and have potent antibacterial effects, sufficient to provide protection in a mouse model of sepsis. Evolutionarily, cathelicidins have a complex history, with marsupials and monotremes uniquely retaining both clusters among tetrapods. Thus, cathelicidins are critical mediators of marsupial immunity, and their evolution may reflect the life history–specific immunological needs of these animals.

## INTRODUCTION

Prenatally protected in the sterile maternal womb, mammalian offspring are suddenly exposed to a plethora of microbes at birth (1, 2). While such exposure allows colonization of commensal microbes, it also poses a considerable threat as sepsis caused by infections can lead to rapid neonatal mortality (3). Notably, susceptibility to such infections may vary among different mammalian lineages as a function of their life history. For instance, precocial mammals, such as cattle and guinea pigs, are often equipped with most of the essential immune components at birth (4, 5). This relative immunological maturity at birth enables precocial mammals to swiftly respond to pathogens. In contrast, altricial mammals, such as mice and rats, are born with a less developed immune system due to their short gestation period (6, 7), requiring more extensive maternal care compared to their precocial relatives.

Among mammals, marsupials are an extreme case of altricial birth. Marsupials diverged from eutherians around 160 million years ago and constitute a unique lineage with characteristic reproductive and morphological traits (8, 9). Females have short pregnancies and give birth to highly immature young that reside inside a pouch where they complete their physical development. A consequence of this short gestation period is that several key immune components are absent in marsupial neonates. Namely, while the hematopoietic niche has already migrated from the liver to the bone marrow during eutherian fetal development, the marsupial neonate liver is still an active site of hematopoiesis (10, 11). Moreover, unlike eutherian neonates, marsupial newborns lack lymphoid organs such as thymus

and lymph nodes, making them incapable of mounting an adaptive immune response. This presents a challenge because the maternal pouch is in contact with the environment and is known to harbor a wide range of bacterial species, including many pathogenic ones (12, 13).

Despite their immunological immaturity, marsupial neonates can survive the nonsterile environment of the pouch in part through maternal protection. This includes the transfer of immunoglobulins through milk as well as the secretion of antimicrobial compounds from specialized pouch glands (14–17). In addition, various immune-related tissues in developing marsupials, including bone marrow and spleen, contain granulocytes and lymphocytes, indicating that neonates themselves can actively fend off pathogen attacks (10). Our current understanding of marsupial neonatal immune protection, particularly neonate-mediated immunity, remains incomplete for several reasons. First, previous research on marsupial immunity has largely relied on work conducted in wild animals (15, 18–20), making it challenging to obtain accurately timed samples or perform controlled experiments. In addition, the scarcity of suitable antibodies has hindered efforts to identify cell types and the cellular composition of marsupial neonatal hematopoietic tissue using histology and microscopy (10, 21–23). In addition, to our knowledge, there has been no unbiased experiment conducted to identify key protective genes expressed by neonatal immune cells. Here, using our captive colony of marsupial sugar gliders (*Petaurus breviceps*) (Fig. 1A) and combining transcriptomics, epigenomics, functional assays, and comparative genomics, we set out to dissect the mechanisms underlying neonate immune protection in marsupials.

## RESULTS

## Neutrophils are the most abundant immune cells in the sugar glider neonate liver

While histological studies have identified few immune cells in developing marsupial neonates, the exact composition and identity of these immune cells during hematopoiesis, as well as the repertoire of key protective genes expressed by each cell type, are unknown. To fill this gap, we conducted single-cell RNA sequencing (scRNA-seq) of the day of birth (postnatal day 0, P0) sugar glider liver, which is

<sup>1</sup>Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA.

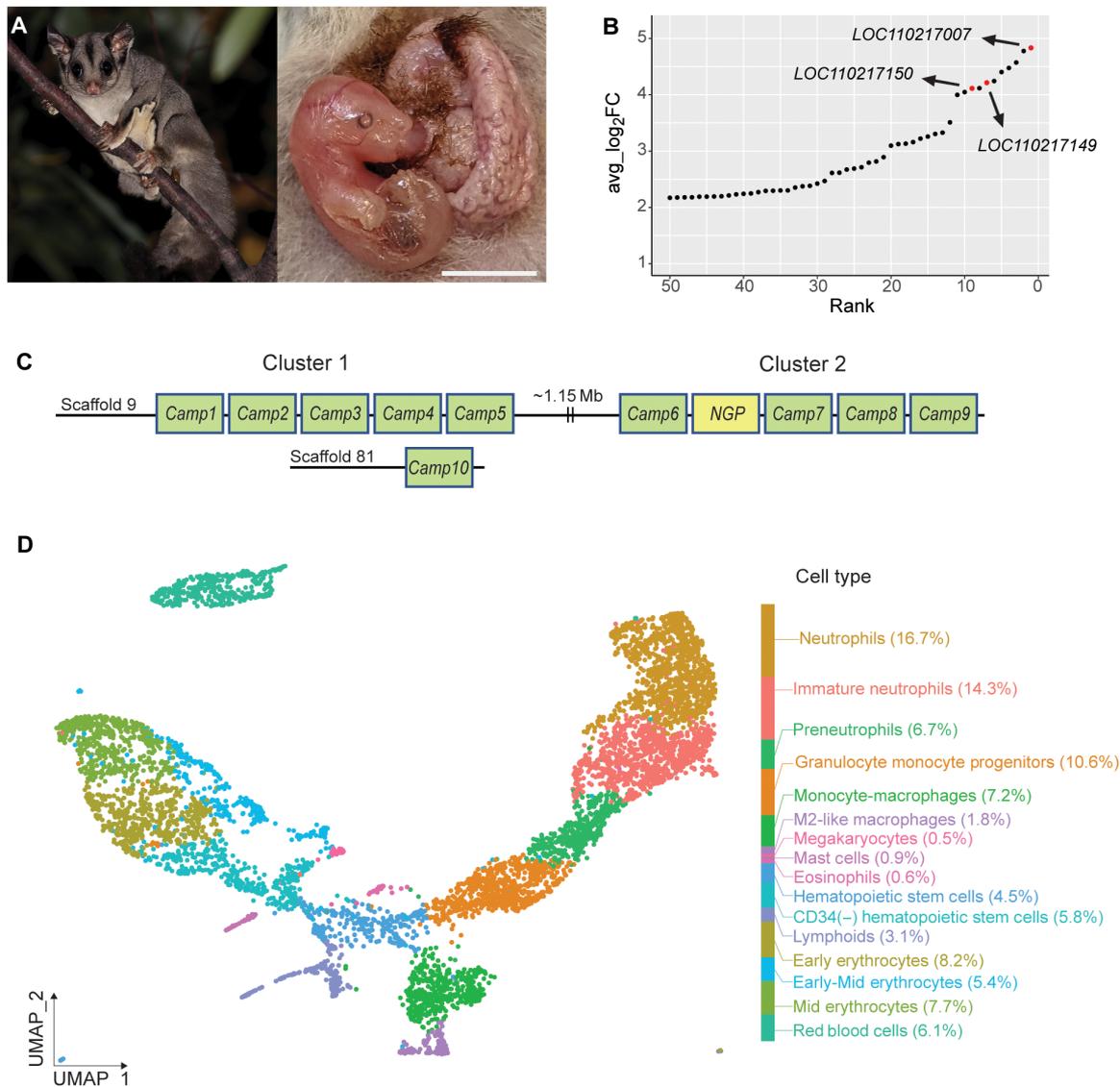
<sup>2</sup>Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08540, USA. <sup>3</sup>Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA. <sup>4</sup>Department of Computer Science, Princeton University, Princeton, NJ 08544, USA.

\*Corresponding author. Email: donia@princeton.edu (M.S.D.); rmallarino@princeton.edu (R.M.)

†Present address: Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA.

‡Present address: Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA.

§Present address: Department of Environment and Genetics, La Trobe University, Bundoora, VIC 3086, Australia.



**Fig. 1. Cathelicidins are highly and specifically expressed in sugar glider neutrophils.** (A) An adult sugar glider [left, photo by P. Kavanagh, originally published under a CC BY 2.0 license (<https://creativecommons.org/licenses/by/2.0/deed.en>)] and a P4 joey laying atop the everted maternal pouch (right). Scale bar, 0.5 cm. (B) Graphical representation of the top differentially expressed genes in neutrophil lineage cells compared to the rest of the liver cells. Genes were ranked by log<sub>2</sub>FC values. Cathelicidin genes are denoted in red. (C) A schematic of the genome structure of sugar glider cathelicidins. Reciprocal blast with leadbeater's possum suggests that *Camp10* belongs to cluster 1. (D) Uniform manifold approximation and projection (UMAP) plot showing cell type clustering and composition of the sugar glider neonatal (P0) liver.

the main site of neonatal hematopoiesis at this stage (10). We detected 6028 cells grouped into 16 clusters, which we then annotated using established gene expression markers (fig. S1). Our analysis revealed that neutrophil lineage cells (i.e., neutrophils, immature neutrophils, preneutrophils, and granulocyte monocyte progenitors) were the most abundant cell type, comprising 48.4% of the entire cell population. Neutrophils are a type of white blood cell and are an essential part of the innate immune system, the body's first line of defense against infections (24). Red blood lineage cells were the second most abundant cell type, comprising 27.4% (fig. S1). We identified a small lymphoid cell population (3.2% of cells), a finding that contrasts with previous reports suggesting that T and B cells were not present in the liver of marsupial neonates (fig. S1) (10, 11, 25, 26).

Lastly, through histological analysis, we confirmed that hematopoiesis takes place within the neonatal liver (fig. S2). Thus, by characterizing the cell type composition of the marsupial liver at single-cell resolution, our results suggest that marsupials rely heavily on neutrophils for neonatal protection.

### Cathelicidins are highly expressed in marsupial neutrophils

The finding that developing neutrophils are the most abundant cells in marsupial neonatal hematopoietic tissue prompted us to focus on this lineage and investigate its function. We therefore conducted differential expression analyses comparing all neutrophil lineage cells to other cell types in the liver. Among the 596 genes that were differentially expressed ( $P$  value < 0.0001, Log<sub>2</sub>Fold > 1), three of the

top 10 most highly up-regulated genes in neutrophil lineage cells (*LOC11021700*, *LOC110217149*, and *LOC110217150*), including the most highly up-regulated one (*LOC11021700*), were identified as cathelicidin antimicrobial peptides (Camps) (Fig. 1B and data table S1). Cathelicidins constitute a key component of the innate immune system due to their ability to disrupt microbial membranes and modulate immune responses (27, 28).

Notably, compared to humans and mice, marsupial genomes have a much larger repertoire of cathelicidin genes. The peptides translated from these genes are known to play key roles in regulating immune responses of different marsupial species (18, 19, 29–32). To gain more insights into the evolution and function of marsupial cathelicidins, we next characterized the expression of these genes in sugar gliders. To do this, we first systemically annotated all members of this gene family in the sugar glider genome and then performed expression-based validation (see Materials and Methods) (table S1 and data table S2) (33). In total, we identified 10 Camp genes, labeled *Camp1* to *Camp10*, as well as the sugar glider ortholog of *Ngp* (neutrophilic granule protein), a cathelicidin-related gene. These genes resided in two clusters located 1.15 Mb apart and were distinguished by the phylogenetic affinity of their constituent genes (Fig. 1C, figs. S3 and S4, and table S2). Cluster 1 contains *Camp1* to *Camp5* and *Camp10*, whereas cluster 2 contains *Camp6* to *Camp9* and *Ngp*. While *Camp2* and *Camp5* have stop codons in exon 2, indicating that they are likely pseudogenes, the remaining genes encode proteins that range from 100 to 266 amino acids in length and 11.1 to 70.9% in identity (table S3).

Next, we reanalyzed our scRNA-seq data from neonatal liver to establish the specific cell types in which the different cathelicidins and *Ngp* are expressed (Fig. 1D). Most of these genes (i.e., *Camp1*, *Camp3*, *Camp4*, *Camp6*, *Camp8*, *Camp9*, *Camp10*, and *Ngp*) are expressed exclusively in neutrophil lineage cells (Fig. 2A and fig. S4). Among neutrophil lineage cells, we observed differences in expression levels, with *Camp3*, *Camp4*, *Camp10*, *Camp6*, and *Ngp* showing high, overlapping expression levels in a large proportion of cells, whereas *Camp1*, *Camp8*, and *Camp9* are expressed at relatively lower levels and in a reduced number of cells (Fig. 2A and figs. S5 and S6). *Camp7* is exclusively expressed by M2-like macrophage cells, suggesting that its transcriptomic regulation, and likely its function, are distinct from the other cathelicidin gene family members (Fig. 2A and fig. S7).

Last, we characterized the dynamics of cathelicidin expression in hematopoietic tissues throughout the lifetime of sugar gliders by performing longitudinal bulk RNA-seq on liver tissue from P0, P10, and adult (>2 years) sugar gliders. In addition, we sampled bone marrow from adult sugar gliders, as this tissue is the main site of hematopoiesis in adult marsupials (10). We found that five of the cathelicidins (*Camp3*, *Camp4*, *Camp10*, *Camp6*, and *Camp7*) as well as *Ngp* are expressed in moderate to high levels in the liver of P0 and P10 joeys, while their expression in this tissue is reduced in adults [read per kilobase per million reads (RPKM) >50] (fig. S7). Notably, these same genes, as well as *Camp9*, are highly expressed in adult bone marrow (fig. S7).

Together, our transcriptomic analyses indicate that cathelicidins have divergent expression patterns, with a subset of them—*Camp3*, *Camp4*, *Camp10*, *Camp6*, and *Ngp*—showing high expression levels in overlapping neutrophil lineage cells, while the expression of another one—*Camp7*—is restricted to macrophage cells. Moreover, our longitudinal analysis indicates that these same genes continue to

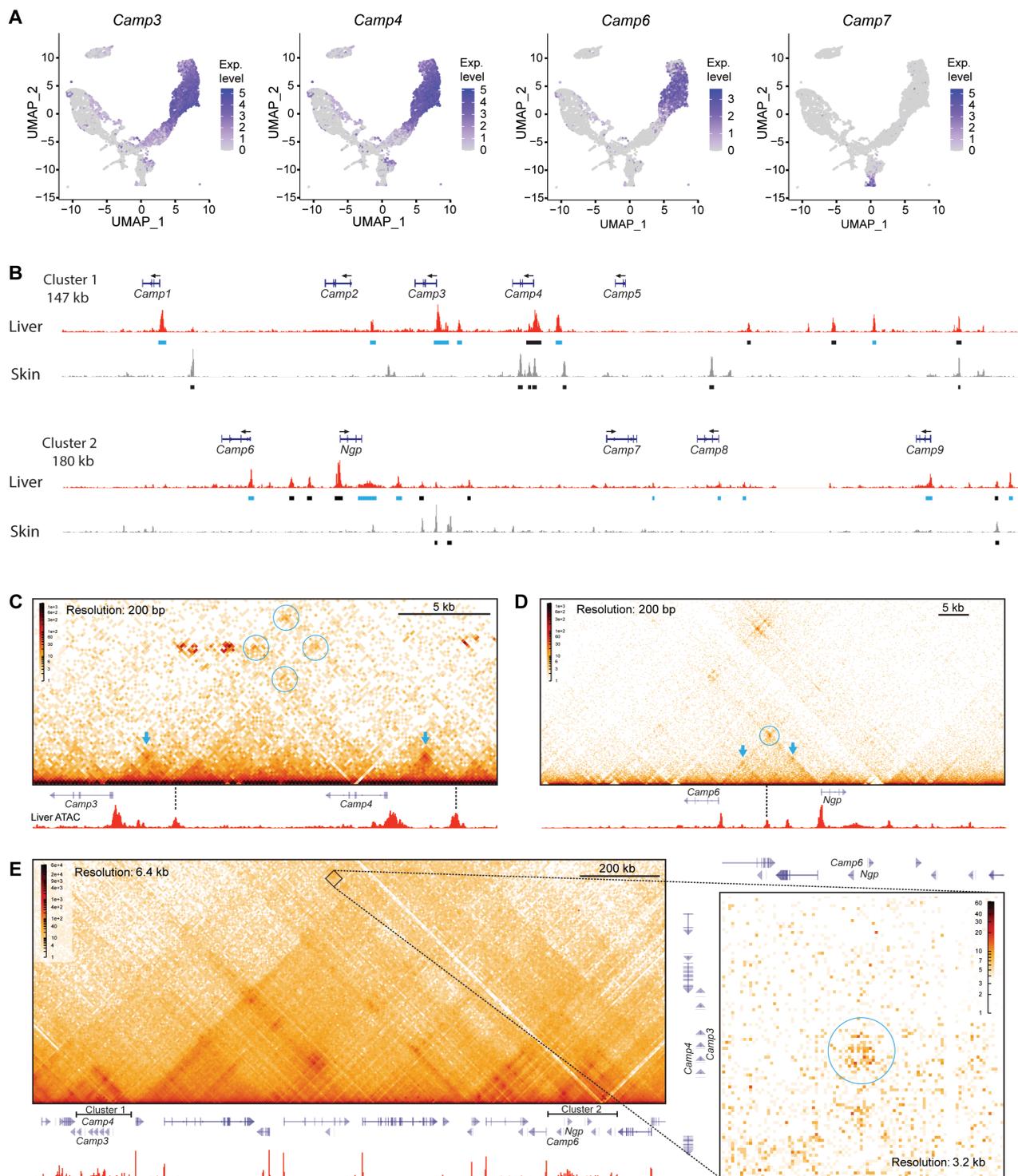
be expressed at high levels throughout the lifetime of sugar gliders, first in the liver and subsequently in the bone marrow.

### Coexpression of cathelicidins is driven by enhancer sharing

Considering that a subset of cathelicidin genes (i.e., *Camp3*, *Camp4*, *Camp6*, *Camp10*, and *Ngp*) were highly coexpressed in developing neutrophils, we next sought to determine whether these genes share common regulatory elements to this end, we conducted an assay for transposase-accessible chromatin sequencing (ATAC-seq) in P10 sugar glider liver, an approach that allows for the identification of open chromatin regions and constitutes a useful strategy for identifying putative cis-regulatory elements (CREs) (34, 35). We identified a total of 10 and 14 peaks located in clusters 1 and 2, respectively (Fig. 2B). To further filter our data, we reanalyzed ATAC-seq data from sugar glider skin tissue (36), which, with the exception of *Camp7*, does not express cathelicidin genes (33). Thus, we could exclude shared open chromatin regions between these tissues as plausible regulators of cathelicidin gene expression. Examination of candidate CREs showed that *Camp3*, *Camp4*, *Camp6*, and *Ngp*, all of which were highly expressed in liver neutrophils, had a total of seven liver-specific open peaks in nearby intergenic regions (Fig. 2B).

We next examined whether any of these putative CREs physically interact with the promoters of *Camp3*, *Camp4*, *Camp6*, and *Ngp* to regulate the expression of these genes. To achieve this, we sampled bone marrow tissue from adult sugar gliders and performed Region Capture Micro-C (RCMC), a chromosome conformation capture technique that enables the detection of interactions between genomic loci at high resolution (37). Among the putative CREs found in cluster 1, two of them displayed strong contact interactions with the promoters of *Camp3* and *Camp4* (Fig. 2C). Notably, these two putative CREs contain a ~600-base pair (bp) region with 81.6% sequence identity, suggesting that they originated via duplication. Similarly, we identified one putative CRE in cluster 2 showing a strong contact interaction with the promoters of *Camp6* and *Ngp* (Fig. 2D). Overall, our results indicate that at least one putative CRE within each cluster displays contact interactions with the promoters of two different genes, providing examples of enhancer sharing. In addition to within cluster interactions, our RCMC data revealed a long-range (>1 Mb) contact interaction between clusters. Specifically, we found evidence of contacts between genes found in cluster 1 (i.e., *Camp3* and *Camp4*) and genes found in cluster 2 (i.e., *Camp6* and *Ngp*) (Fig. 2E). These results suggest that genes across both clusters can be coregulated.

Motivated by the finding that enhancer sharing can account for the coordinated expression of cathelicidin genes, we next sought to identify transcription factors driving such patterns. We analyzed the regulatory regions of *Camp3* and *Camp4*, two neighboring genes sharing regulatory elements and coexpressed at high levels in the same cells (Fig. 2, A and C, and fig. S6). First, we carried out motif enrichment analysis on putative CREs and then cross-referenced the results with our gene expression data, as well as with publicly available ChIP data for human and mouse (figs. S8 and S9 and data table S3). Through this filtering strategy, we identified five transcription factors (*Spi1*, *Fli1*, *Runx1*, *C/ebpδ*, and *C/ebpe*), which were robustly expressed in P0 livers (fig. S8). To test whether these transcription factors could bind to putative CREs, we performed a set of luciferase assays using our immortalized line of sugar glider dermal fibroblasts (36). Specifically, we cotransfected each of the five transcription factors with luciferase reporter vectors containing the putative



**Fig. 2. Expression and regulation of sugar glider cathelicidins.** (A) UMAP plots displaying the expression of selected cathelicidin genes. (B) ATAC-seq traces of sugar glider liver and skin tissue. Shown are genomic regions corresponding to clusters 1 and 2. Black boxes denote putative CREs, and blue boxes denote putative CREs containing *Cebp* family binding sites. Arrows indicate the direction of transcription. (C to E) Contact maps of sugar glider adult bone marrow. Displayed are genomic regions containing *Camp3* and *Camp4* (C), *Camp6* and *Ngp* (D), and both cathelicidin clusters (E). Heatmaps visually represent the frequency of physical interactions between different regions of the genome, with darker colors representing higher frequencies than lighter colors. Blue arrows denote an interaction between enhancers and their proximal promoters. Circles indicate interactions among multiple enhancers and promoters.

*Camp3* CRE, which was used as a representative sequence. Quantification of fluorescence after 48 hours revealed that, of the different transcription factors tested, *C/ebpδ* and *C/ebpε* drove significantly higher levels of luciferase activity, compared with the green fluorescent protein (GFP) control vector (fig. S8 and table S4). All highly expressed cathelicidins—*Camp3*, *Camp4*, *Camp6*, and *Ngp*—contained a *C/ebp* transcription factor binding motif in their promoter or in a nearby intergenic peak (Fig. 2B). Notably, however, the putative shared CRE found to interact with both *Camp6* and *Ngp* in the RCMC analysis described above (Fig. 2D) did not contain a *C/ebp* motif (Fig. 2B). Therefore, additional transcription factors may be involved in facilitating local enhancer-promoter interactions regulating *Camp6* and *Ngp*.

Together, our open chromatin data coupled to our RCMC data indicate that cathelicidin coexpression can take place via enhancer sharing. Moreover, interactions between regulatory sequences and different cathelicidin genes can take place both within and between clusters. Lastly, our luciferase data suggest that, while *C/ebp* transcription factors are key regulators of cathelicidin expression, the presence of additional transcription factors is likely required to coordinate the regulatory control of members of the gene family.

### The function of sugar glider cathelicidins

The large repertoire of cathelicidin genes observed in marsupials, compared to mice and humans, raises the intriguing possibility that this gene family has been fine-tuned by natural selection such that different genes have acquired distinct roles. To test this, we next performed a variety of in vitro and in vivo assays aimed at characterizing the function of sugar glider cathelicidins. First, we investigated whether cathelicidins, like their eutherian homologs, exhibit antibacterial properties, and the extent to which different genes vary in this ability. In vivo, cathelicidins are cleaved into mature peptides by serine proteases (38, 39). Therefore, we predicted the most plausible cleavage sites and then chemically synthesized the mature peptides for seven sugar glider cathelicidins (CAMP1, CAMP3, CAMP4, CAMP7, CAMP8, CAMP9, and CAMP10) and for NGP (Fig. 3A and table S2) (27, 40, 41). We did not synthesize mature peptides of CAMP6 because of its considerable length (137 AA) or of the putative pseudogenes *Camp2* and *Camp5*. Most of the synthesized peptides were predicted to have a helical secondary structure, like other canonical cathelicidin peptides (fig. S10).

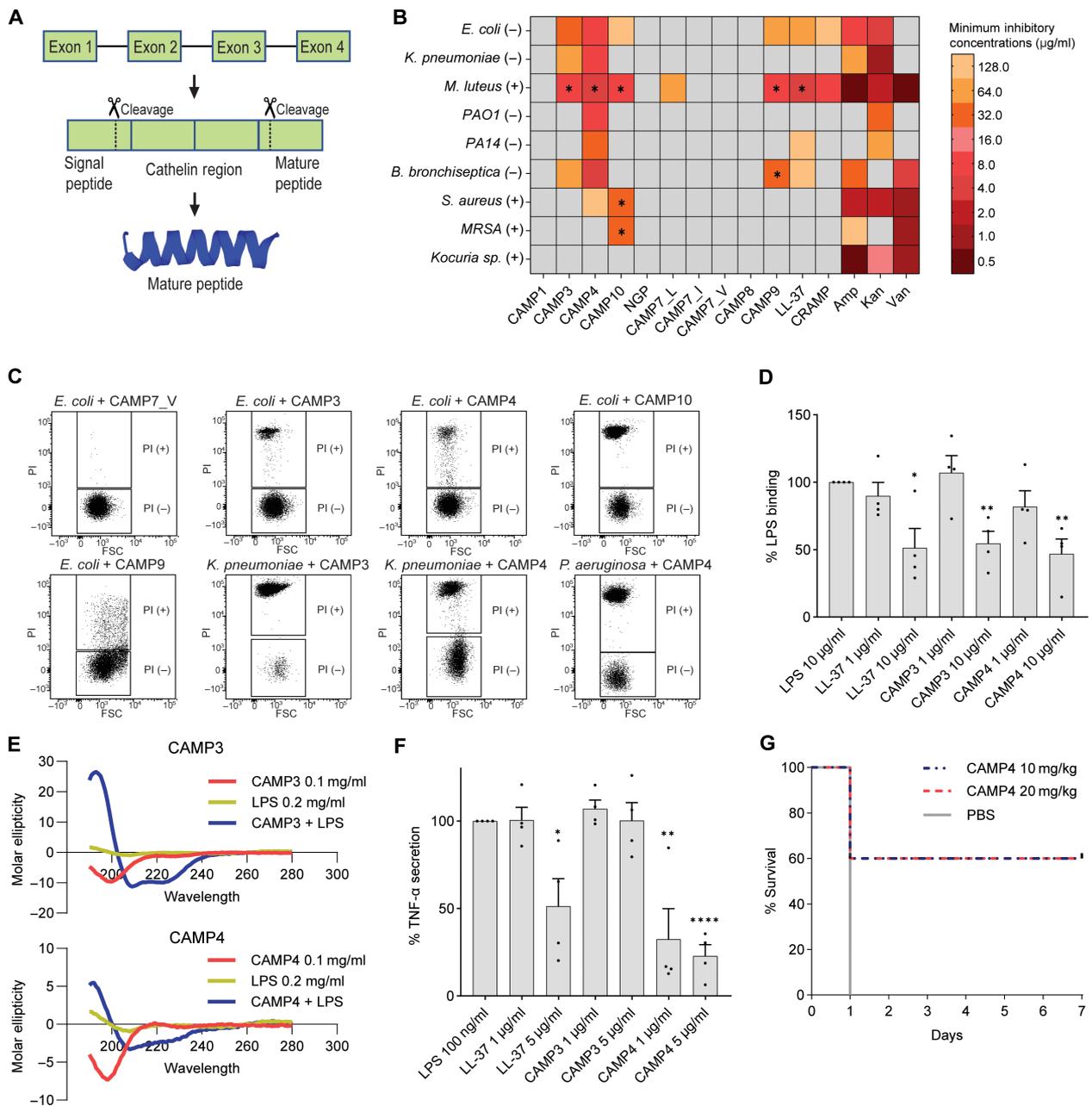
We then performed broth micro-dilution antibacterial assays against *Escherichia coli*, *Klebsiella pneumoniae*, *Micrococcus luteus*, *Pseudomonas aeruginosa*, *Bordetella bronchiseptica*, and *Staphylococcus aureus*, common mammalian pathogens that were previously found in the pouch of wild marsupials (12, 13), as well as against *Kocuria* sp., which was isolated from the pouch and skin swabs of female sugar gliders within our colony (Fig. 3B and table S5). For comparison, we included synthesized peptides corresponding to the human (LL-37) and mouse cathelicidins (CRAMP), as well as the antibiotics ampicillin, kanamycin, and vancomycin, all of which are known to have strong antibacterial activity (42, 43). We found that CAMP3, CAMP4, CAMP9, and CAMP10 exhibited strong antibacterial activities against various bacteria, while CAMP1, CAMP7, CAMP8, and NGPs had low or no antibacterial activities (Fig. 3B and table S5). Among the peptides with strong antimicrobial activity, CAMP4 was the strongest, showing marked efficacy against *E. coli*, *K. pneumoniae*, *M. luteus*, *B. bronchiseptica*, and *P. aeruginosa* with minimum inhibitory concentrations (MICs) as low as 4 μg/ml, which are lower or comparable to

the values of both human and mouse cathelicidins against these microbes (Fig. 3B and table S5). In addition, the MIC of CAMP4 against particular bacterial species was comparable to that of common antibiotics (e.g., ampicillin against *E. coli*) (Fig. 3B and table S5). Notably, none of the cathelicidins showed any antibacterial activity against the sugar glider isolate *Kocuria* sp. (Fig. 3B and table S5), suggesting that these potentially symbiotic microbes are resistant to the host's AMPs. These results indicate that multiple sugar glider cathelicidins exhibit strong antibacterial properties against common mammalian pathogens, albeit to a varying extent depending on the cathelicidin and bacterial species being tested.

While broth micro-dilution assays indicate that sugar glider cathelicidins inhibit the growth of bacteria, it remains unknown whether these peptides have direct bactericidal activities. To address this, we performed a propidium iodide (PI) absorption assay (44). If an antibiotic agent kills bacteria, PI will penetrate bacterial membrane and intercalate with DNA, emitting a fluorescent signal that can be measured via flow cytometry (Fig. 3C). We applied this assay to *E. coli*, *K. pneumoniae*, and *P. aeruginosa* PAO1, using cathelicidin peptides that showed antibacterial activities against each of the bacterial species in the broth microdilution assay (*E. coli*: CAMP3, CAMP4, CAMP9, and CAMP10; *K. pneumoniae*: CAMP3 and CAMP4; *P. aeruginosa* PAO1: CAMP4), along with the non-antibacterial cathelicidin, CAMP7, as a negative control (Fig. 3C). We found that all four antibacterial cathelicidins were able to kill bacteria, while CAMP7 was not (Fig. 3C). Since antimicrobial cathelicidins are often cytotoxic to mammalian cells (45, 46), we performed cell viability assays and found that the MIC ranges for the strongest cathelicidins (i.e., CAMP3, CAMP4, and CAMP10) were only minimally cytotoxic to mammalian cells, with cell viabilities higher than 90% for both immortalized sugar glider fibroblasts and J774.1 murine macrophages (fig. S11).

In addition to direct microbicidal activity, human and mouse cathelicidins (i.e., LL-37 and CRAMP) have immunomodulatory properties, including chemotaxis, anti- and proinflammatory effects, and the promotion of T helper 17 differentiation (47–50). To establish whether sugar glider cathelicidins were capable of secondary activities in addition to the direct antibacterial property, we first incubated murine macrophages (J774.1) with fluorescein isothiocyanate (FITC)-conjugated lipopolysaccharide (LPS) in the presence of cathelicidins or a vehicle control and measured the resulting fluorescence intensity using flow cytometry. Among the peptides tested, CAMP3 and CAMP4 significantly reduced the binding of LPS to macrophages (Fig. 3D and fig. S12). This effect was caused by direct binding of CAMP3 and CAMP4 to LPS, as determined by circular dichroism analyses (Fig. 3E) (51). Next, we used enzyme-linked immunosorbent assay (ELISA) to measure the amount of tumor necrosis factor-α (TNF-α, a proinflammatory cytokine) released by murine macrophages upon LPS incubation, in the presence of LPS-binding cathelicidins or a vehicle control. Our results showed that CAMP4 significantly reduced TNF-α secretion compared to the vehicle control (Fig. 3F). Overall, our results show that cathelicidin peptides decrease the expression of a proinflammatory cytokine by inhibiting the binding of LPS to mammalian cells and thereby exhibit LPS-binding activity.

Last, we sought to evaluate the in vivo efficacy of cathelicidins in a murine sepsis model (52, 53). To this end, we tested whether CAMP4, the peptide that showed the strongest response in our antibacterial assays as well as potent LPS-binding activity, enhanced the survival of mice infected with *E. coli*. We intraperitoneally injected  $5 \times 10^7$  colony-forming units (CFUs) of *E. coli* into laboratory mice



**Fig. 3. Function of sugar glider cathelicidins.** (A) Schematic representation of cathelicidin peptide processing. Cathelicidins, composed of a signal peptide, a cathelin region, and a mature peptide, get cleaved into mature peptides by proteolytic proteins. (B) Heatmap showing the minimal inhibitory concentrations (MICs) of CAMPs against different types of bacteria. Asterisks denote the existence of a higher peptide concentration where percent inhibition was lower than 90%. Plus and minus signs denote Gram-stain status of the different bacteria. (C) Flow cytometry analysis of PI incorporation. Upon bacterial membrane disruption, PI intercalates with DNA, emitting a fluorescence signal. X axis and Y axis represent forward scatter and fluorescence intensity of PI, respectively. (D) Bar graph representing the fluorescence intensity, as measured via flow cytometry, of macrophages coincubated with LPS-FITC and different peptides. Y axis represents the percent of the median fluorescence intensity relative to the PBS control. Statistical significance was assessed using an unpaired t test ( $n = 4$ ;  $*P < 0.05$ ,  $**P < 0.01$ ). (E) Graphs showing the shift in the circular dichroism spectra of CAMP3 and CAMP4 upon incubation with LPS. The existence of negative bands around 220 nm and the positive band at 193 nm indicate the formation of an  $\alpha$  helix caused by the interaction with LPS. (F) Bar graph showing the amount of TNF- $\alpha$  secreted from LPS-stimulated macrophages upon coincubation with peptides. Values are displayed relative to the PBS control. Statistical significance was assessed using an unpaired t test ( $n = 4$ ;  $*P < 0.05$ ,  $**P < 0.01$ ,  $****P < 0.0001$ ). (G) Result from the in vivo mouse survival assay. Mice infected with *E. coli* were treated with CAMP4 or PBS control ( $n = 5$ ).

and treated them with one of two different concentrations of CAMP4 or a vehicle control [phosphate-buffered saline (PBS)]. Strikingly, we found that while all mice treated with PBS died within 7 days, mice treated with both *Camp4* (10 and 20 mg/kg) showed 60% survival rate (three of five mice survived) (Fig. 3G). This demonstrated that *Camp4* is effective in protecting mice from lethal sepsis, further corroborating our *in vitro* results.

### Evolution of sugar glider cathelicidins

The immunological significance of cathelicidins and the large number of these genes in marsupial genomes (18, 29, 32) prompted us to investigate the evolutionary history of this gene family. First, we examined patterns of divergence among sugar glider cathelicidins. We conducted pairwise comparisons of amino acid sequences among all sugar glider cathelicidins encoding functional proteins and found that these genes have experienced marked sequence diversification in the fourth exon, which is the region encoding the mature effector peptide (Fig. 4A). Moreover, comparisons among *Camp3*, *Camp4*, and *Camp10*, the three most similar genes, indicated that the nucleotide sequence identity of the fourth exon (37.8%) was significantly lower than those of three introns (85.5, 91.7, and 84.7%, respectively) (Fig. 4B).

Motivated by the extensive number of genes observed among marsupial cathelicidins, we expanded our evolutionary analysis to search for broader evolutionary patterns among tetrapods. To achieve this goal, we annotated cathelicidins and *Ngp* genes in publicly available genomes of 24 additional mammals (table S6). In addition, we included cathelicidin sequences of four nonmammalian tetrapods as outgroups (table S6). Then, we aligned sequences along with syntenic marker genes for cluster identification (Fig. 4C). Our analysis revealed that marsupials and monotremes are the only tetrapods with two clusters of cathelicidins, while eutherian mammals and nonmammalian tetrapods each have a single cluster (Fig. 4C). Notably, the cathelicidin cluster present in eutherian mammals is distinct from the cluster found in nonmammalian tetrapods, as indicated by their syntenic relationships with surrounding genes and the phylogenetic relationships of their constituent cathelicidins. This observation suggests that a duplication event of the cathelicidin gene cluster took place in the mammalian ancestor before the divergence between monotremes and therian mammals. Marsupials and monotremes have retained both the ancestral tetrapod and novel mammalian clusters, while eutherian mammals lost the “ancestral” cluster (cluster 2 in sugar gliders) shared with nonmammalian tetrapods, retaining only the “mammalian” cluster (cluster 1 in sugar gliders). Further pairwise sequence comparisons among sugar glider cathelicidins revealed that genes within the mammalian cluster show higher identity than those within the ancestral cluster, reinforcing that the mammalian cluster has likely arisen more recently (Fig. 4D).

Last, our comparative annotation and synteny map suggests that *Ngp* is a gene that exists only in mammals. We carried out a phylogenetic analysis of 151 tetrapod cathelicidin and 28 *Ngp* amino acid sequences and found that all *Ngp* sequences formed a monophyletic group (fig. S13 and data table S4). Furthermore, when we blasted amino acid sequences of representative mammalian *Ngp* genes against nonmammalian protein database, we found that the top hits were cathelicidins, indicating that *Ngp* evolved from a cathelicidin in early mammal or pre-mammalian synapsid diversification (table S7).

Overall, our phylogenetic characterization of cathelicidin and *Ngp* genes indicates that mammalian cathelicidins have experienced

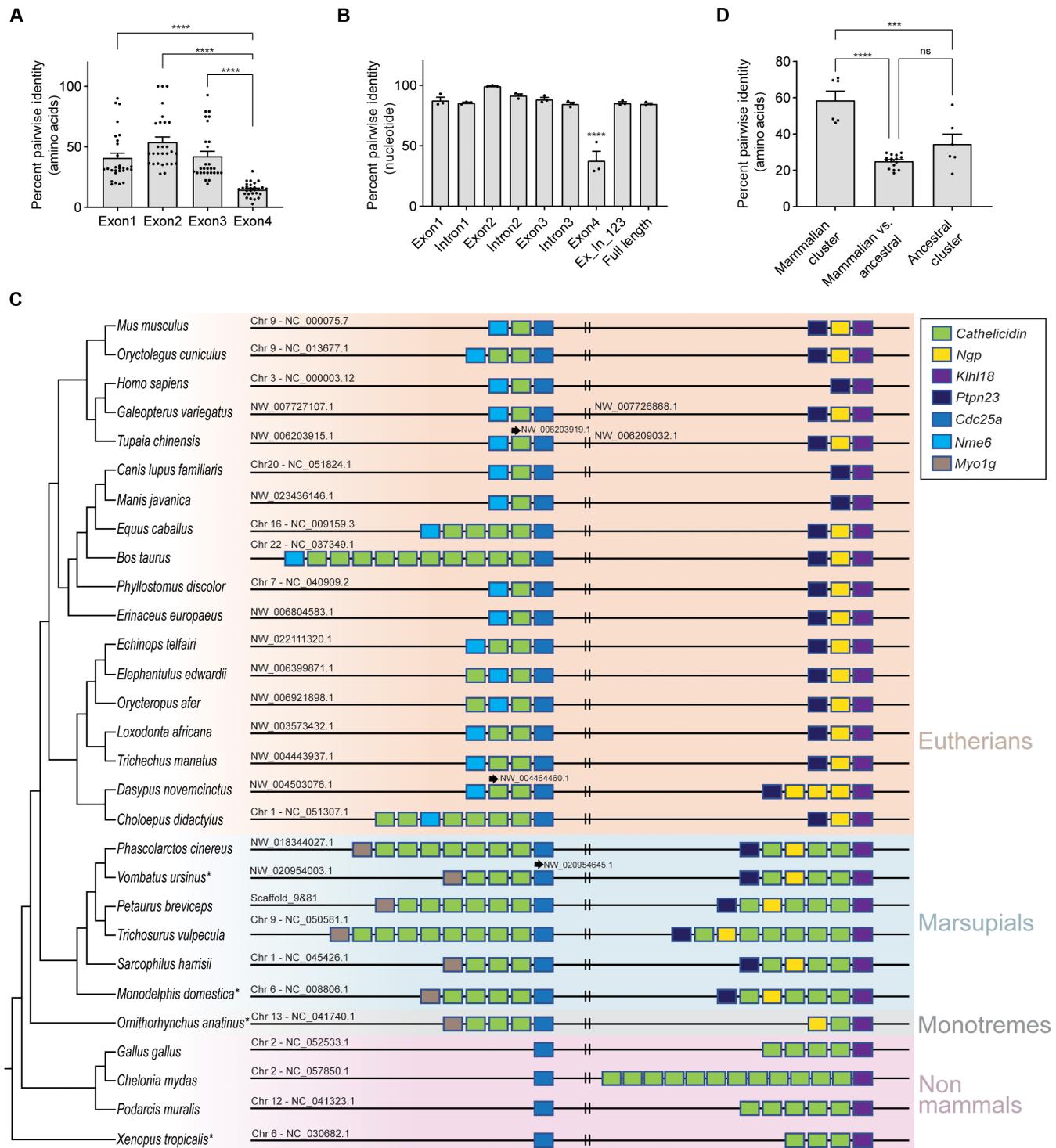
multiple evolutionary events, including both duplication and loss of gene clusters, as well as the derivation of a gene, *Ngp*. The genomic expansion and the retention of two different cathelicidin clusters suggest that cathelicidins likely played a crucial role in shaping marsupial immune defenses throughout evolution.

### DISCUSSION

At birth, the transition to a microbe-rich extrauterine environment presents mammalian neonates with major challenges due to their naïve immune systems. These challenges are even greater in marsupials, which complete the development of lymphoid organs postnatally (10). The exceptionally altricial state of marsupial neonates increases their reliance on innate immunity during the first days and weeks of life. Through single-cell transcriptomic analysis of the P0 sugar glider liver, we have characterized the cellular composition and the repertoire of genes expressed in the neonatal marsupial hematopoietic tissue. We present evidence that several cathelicidin genes are highly coexpressed in neutrophils and that this process is partly regulated by enhancer sharing and long-range physical interactions between loci found in the two clusters. Moreover, we show that sugar glider cathelicidins have LPS-binding properties, can directly kill bacteria by disrupting their cell walls, and are able to protect mice from lethal sepsis. Lastly, our comparative genomic analysis revealed that marsupials and monotremes—both of which give birth to highly altricial neonates (54)—retain both ancestral and mammalian cathelicidin clusters, while eutherians and nonmammalian tetrapods have a single cluster.

By studying sugar gliders—a species suitable for captive breeding—and by combining transcriptomics, epigenomics, functional experiments, and comparative genomics, we demonstrate that cathelicidins represent a key component of the early postnatal immunological defense system in marsupials. Since their discovery in insects, eukaryotic antimicrobial peptides have been extensively studied as primary defense molecules (55). However, in vertebrates, particularly in mammals, these peptides are often considered secondary accessories to the adaptive immune system. As marsupials constitute a system that bridges the gap between invertebrate models lacking adaptive immunity and eutherian models with well-developed adaptive immune systems, our conclusions provide evidence that antimicrobial peptides can be primary actors of immune defenses in mammals.

At birth, marsupials and eutherians exhibit marked immunological differences (10, 56). Recent gene expression profiling across various organs reveals that the cathelicidin gene is highly expressed in the liver and myeloid cells of newborn mice (P0) (57, 58), suggesting that cathelicidins may play roles in eutherian neonatal innate immunity. Our findings suggest that differences between cathelicidins in sugar gliders and mice are not primarily due to differences in expression but arise from two key factors. First, immune cell composition differs considerably at birth: The blood of P0 mice has a higher proportion of lymphocytes compared to neutrophils (59), whereas the liver of P0 sugar gliders has 15 times more neutrophils than lymphocytes. As a result, cathelicidins may play a more prominent role in neonatal immune protection in marsupials than in eutherians. Second, cathelicidins in sugar gliders and mice differ in gene copy number and sequence. While mouse genomes contain a single cathelicidin gene, marsupial genomes have multiple copies, which exhibit extensive sequence diversification, particularly in the fourth exon. Although having more copies does not necessarily result in more diverse



**Fig. 4. Evolution of the cathelicidin gene family. (A)** Percent sequence identity (amino acids) among the exons of eight sugar glider cathelicidins. Statistical significance was assessed using a one-way analysis of variance (ANOVA) with Tukey's multiple comparisons ( $****P < 0.0001$ ). **(B)** Percent sequence identity (nucleotide) among introns and exons of *Camp3*, *Camp4*, and *Camp10*. Statistical significance was assessed using a one-way ANOVA, and post hoc pairwise comparisons were carried out using a Bonferroni correction;  $****P < 0.0001$ . **(C)** Synteny map of cathelicidin gene clusters across mammals and nonmammalian vertebrates. Asterisks denote species in which cathelicidins were present outside the shown region. **(D)** Percent sequence identity (amino acids) within and between cathelicidin clusters. Statistical significance was assessed using a one-way ANOVA, and post hoc pairwise comparisons were carried out using a Bonferroni correction ( $***P < 0.001$ ,  $****P < 0.0001$ ).

functions, our findings in sugar gliders suggest that this genomic diversity may offer marsupials a broader range of defenses compared to species with lower variation in copy number.

While our study constitutes a comprehensive characterization of the evolution, regulation, and function of marsupial cathelicidins, there are several questions that remain to be addressed and key future directions to pursue. First, although our functional studies used bacterial species previously identified in other marsupials and in our captive colony, it would be crucial to also characterize the pathogens found in the sugar glider's native environment. This would allow us to test whether sugar glider cathelicidins are effective against the pathogens they naturally encounter and ultimately help identify selective pressures driving the evolution of antimicrobial peptides in this species (60). Second, as a complement to our predictions of mature peptides, carrying out *in vivo* peptidomic profiling approaches will be critical to elucidate the full range of peptides generated by that sugar glider cathelicidin genes (61). Third, while many sugar glider cathelicidins have LPS-binding and antibacterial functions, others, such as *Camp6*, *Camp7*, and *Ngp*, do not, despite being highly expressed. Hence, functionally characterizing the genes that make up cluster 2 may yield fascinating insights. For example, because of its unique expression pattern, *Camp7* is likely to have noncanonical functions different from those of other cathelicidin genes. Fourth, while our assays demonstrate that sugar glider cathelicidins are functionally important, the extent to which these genes protect sugar glider neonates *in vivo* remains unknown. Incorporating genome editing approaches in sugar gliders will eventually allow us to systematically study the function of each cathelicidin gene through loss-of-function experiments (62). In addition, investigating whether systemic infection of sugar glider neonates with microbial pathogens promotes the expression of cathelicidins in liver granulocytes constitutes a promising research avenue. Fifth, while our functional characterization primarily focused on cathelicidin genes expressed by neutrophils, other antimicrobial peptides are likely to contribute to neonatal immune defenses. Thus, since antimicrobial peptides are also expressed in nonleukocyte cells (e.g., epithelial and mucosal cells), future efforts should focus on characterizing immunological factors produced by tissues exposed to external environments (63). Lastly, contrary to previous reports, our study identified a small lymphoid population in the liver of marsupial neonates. Sampling the marsupial liver at additional time points could provide insights into the dynamics of these cells and help determine how cell type composition changes over time.

In addition to motivating these questions, our results pave the path toward studying foundational concepts in mammalian gene regulation, ecology, and evolution. For example, the high-resolution regulatory interactions between the two cathelicidin gene clusters provide a framework for identifying genomic elements, transcription factors, and coexpression determinants among members of a single gene family. Moreover, mammalian cathelicidin evolution represents a system to investigate the relationship between a rapidly evolving gene family and the immunological needs of a taxon. Our study has identified a correlation between gene duplication events and a key life history trait. As different mammalian species exhibit a broad spectrum of developmental maturity at birth (54), comparative genomic analysis across a range of altricial and precocial taxa will further illuminate the relationship between life history and the evolution of immune-related gene families.

## MATERIALS AND METHODS

### Sugar glider husbandry

A breeding colony of sugar gliders was maintained as previously described (33). Briefly, captive adult sugar gliders were mostly housed as male–female pairs, with few animals being housed as a group of one male and two females. The breeding colony was kept under a 12-hour light and 12-hour dark cycle with temperatures and humidities range from 20° to 27°C and 30 to 70%, respectively. Females were checked for pouch young by visual inspection as well as gentle palpation of the maternal pouch. Pouch young were removed from the pouch while the mother was briefly anesthetized with isoflurane. All procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at Princeton University (2155 and 3002).

### Single-cell RNA sequencing

A P0 sugar glider neonate was euthanized by decapitation, and its liver was carefully dissected from the animal, washed in ice-cold PBS, and digested with 0.02% collagenase II for 30 min at 37°C with gentle rotation. Primary cells were filtered with a 70- $\mu$ m cell strainer and then centrifuged for 4 min at 800g at 4°C. The resulting cell pellet was suspended in 1 $\times$  PBS containing 0.04% fetal bovine serum (FBS). Cell viability and number were assessed with a hemocytometer using trypan blue. A scRNA-seq library was prepared using the Chromium Single-Cell 3-prime Library Prep Next GEM (v3.1) and was sequenced on an Illumina NovaSeq 6000 (28  $\times$  90 bp, paired-end), generating 508 M reads. The resulting reads were mapped against the sugar glider genome using CellRanger-7.0.0. Cells containing more than 200 genes, but less than 10% of total mitochondrial genes, were kept for downstream analysis. Unique Molecular Identifier (UMI) count data of surviving cells were normalized using Seurat v4.3.0 with SCTransform (64). We performed principal component analysis and ran Uniform Manifold Approximation and Projection (UMAP) dimensional reduction technique ( $n = 30$  dimensions) using Seurat's default parameters. Neighbors and clusters were identified with 30 dimensions and 1.0 resolution, respectively.

Clusters were annotated on the basis of the following marker genes: red blood cell lineages: *Hemgn*, *Klf1*, *Kit*, *Band3*, *Gpa*, and hemoglobin subunits (65, 66); hematopoietic stem cell clusters: *Cd34* and *Znf521* (67, 68); neutrophil lineage cells: *S100A8*, *Gfi1*, *Mpo*, *Prtn3*, *Elane*, *Ltf*, *Ngp*, *Csf3r*, *Prex1*, and *Pde4b* (69–71); lymphoid cells: lymphoid marker genes *Cd3* subunits, *Blk*, and *Jchain* (72, 73); monocyte-macrophage lineages: *Maf*, *Mafb*, and *Cd80* (74–78); megakaryocytes: *Tuba8* and *Gp1ba* (79, 80); mast cells: *LOC110197119* (mas-related G protein-coupled receptor member H-like) and *Ms4a2* (81, 82); and eosinophils: *Epx* and *Il5ra* (83, 84).

Differential expression analysis was conducted using the FindMarkers() function of the Seurat package. Normalized gene expression values were visualized using FeaturePlot(). Coexpression plots were visualized using FeatureScatter().

### Histology

Postcranial portions of a P6 joey was fixed in 4% paraformaldehyde, washed in 1 $\times$  PBS, and cryoprotected through a series of 15 and 30% sucrose, before embedding in optimal cutting temperature compound (Tissue-Tek 4583). Sagittal sections were cut at 12- $\mu$ m thickness using a Leica CM3050S cryostat, mounted, and subsequently stained using hematoxylin and eosin (85). Stained sections were visualized using a Vectra Polaris slide scanner. Resultant

sections allowed for the visualization of developing liver, heart, and additional viscera.

### Annotation of cathelicidins and *Ngp*

After compiling sequences of 79 mammalian cathelicidin and cathelicidin-like sequences from National Center for Biotechnology Information (NCBI), we blasted protein sequences of all cathelicidins against the previously constructed de novo sugar glider transcriptome (33). For targeted gene annotation, we next blasted both 79 mammalian cathelicidin protein sequences and de novo sugar glider transcript hits with e-values lower than  $10^{-5}$  against the sugar glider genome and extracted sequences 1-mb upstream and downstream of the resulting blast hits.

Using de novo transcript hits, as well as mRNA and protein sequences of 79 mammalian cathelicidins as EST evidence, we ran Maker3 (86) on extracted genome sequences (six rounds) using the Augustus human gene prediction model with snap hmm building (87, 88). To annotate cathelicidin genes and *Ngp* in 24 additional mammalian species (table S6; *Ornithorhynchus anatinus*, *Monodelphis domestica*, *Sarcophilus harrisi*, *Trichosurus vulpecula*, *Vombatus ursinus*, *Phascolarctos cinereus*, *Choloepus didactylus*, *Dasyurus novemcinctus*, *Trichechus manatus latirostris*, *Loxodonta africana*, *Orycteropus afer afer*, *Elephantulus edwardii*, *Echinops telfairi*, *Erinaceus europaeus*, *Phyllostomus discolor*, *Bos taurus*, *Equus caballus*, *Manis javanica*, *Canis lupus familiaris*, *Tupaia chinensis*, *Galeopterus variegatus*, *Homo sapiens*, *Oryctolagus cuniculus cuniculus*, and *Mus musculus*), we used mammalian cathelicidin and NGP sequences as EST evidence and performed the same targeted gene annotation using a single round of the Maker3 run. Resulting annotations were manually compared with NCBI reference annotations to accurately count the number of genes. For four nonmammalian tetrapods (*Xenopus tropicalis*, *Podarcis muralis*, *Chelonia mydas*, and *Gallus gallus*), we primarily relied on the NCBI and ensemble annotations since our Maker3 annotation based on mammalian cathelicidins did not provide accurate annotations. Specifically, focusing on the ancestral cluster locus, we looked for genes annotated as cathelicidins or cathelicidin-like. Then, to find additional cathelicidins, we blasted protein sequences of these genes against the protein database of the respective species.

For further validation of sugar glider cathelicidin and *Ngp* sequences, we generated cDNA from RNA extracted from the lung, liver, stomach, kidney, intestine, and brain of an ~P20 (1.35 g) joey. Polymerase chain reaction (PCR) amplifications followed by Sanger sequencing confirmed the sequences of six cathelicidins and of *Ngp*.

To infer the genomic location of *Camp10*, we first acquired protein sequences of cathelicidins from the leadbeater's possum (*Gymnobelideus leadbeateri*, Genome assembly LBP\_v1:GCA\_011680675.1), a sister species of the sugar glider, by lifting-over gene model from sugar glider cathelicidin annotation (36, 89). We then conducted reciprocal protein blast of cathelicidins from the leadbeater's possum and the sugar glider, as well as a nucleotide blast search of a 7000-bp region of scaffold 81, which contains cathelicidin 10, against the genome of the leadbeater's possum (fig. S3).

### Phylogenetic tree construction

Sugar glider cathelicidin and *Ngp* nucleotide sequences aligned with MAFFT v7.475 were fed into RAxML for the construction of unrooted phylogenetic trees with 1000 bootstrap values and GTRGAMMA model (90). To build a comprehensive phylogeny of cathelicidins and *Ngp*, we compiled 151 cathelicidins and 28 NGP sequences from

NCBI (data table S4). We built a phylogenetic tree with MEGA-X using cathelicidin and NGP sequences aligned with the built-in muscle aligner (100 bootstraps and JTT model) (91). All trees were visualized with FigTree (<https://github.com/rambaut/figtree>).

### Bulk RNA sequencing

For bulk RNA sequencing, sugar glider liver samples were collected from P0 neonates ( $n = 3$ ), ~P10 joeys (0.45, 0.55, and 0.56 g;  $n = 3$ ), and adults ( $n = 3$ ). In addition, we collected bone marrow samples from adult tibia and femur ( $n = 3$ ), as previously described (92). Briefly, hindlimb bones were cut (around the kneecap) and placed into 0.5-ml tubes. Using a sterile pin, a hole was poked in the bottom of the tube, and it was placed inside a 1.5-ml container with 100  $\mu$ l of PBS. The bone marrow sample was then collected by centrifugation (8000g, 30 s at room temperature). The RNeasy mini kit (QIAGEN) was used for RNA extraction following the supplier's protocol. We used 250 ng of RNA to prepare libraries with NEBNext Ultra II Directional RNA Library Prep kit. Libraries were sequenced to a median depth of 27 M reads on the NovaSeq 6000 ( $2 \times 65$  bp). Demultiplexed reads were filtered and trimmed using Trimmomatic 0.39 with parameters: MINLEN:25 AVGQUAL:20 (<https://github.com/usadellab/Trimmomatic>). Resulting reads were mapped against the sugar glider genome using STAR-2.7.8a (<https://github.com/alexdobin/STAR>). RPKM values were calculated on the basis of read counts acquired from featureCounts (<https://subread.sourceforge.net/>). RPKM values were visualized using GraphPad Prism.

### ATAC sequencing

Single-cell suspensions were generated from a liver of a P10 joey as described for scRNA-seq. For each replicate ( $n = 3$ ), library preparation was done using 100,000 cells, following the Omni-ATAC protocol (35). Briefly, following lysis on ice for 3 min, the cells were incubated with TDE1 transposase (Illumina) for 1 hour at 37°C and then purified with the Zymo DNA Clean and Concentrator-5 kit (Zymo). Illumina sequencing adaptors and barcodes were added to the DNA fragments. The resulting libraries were sequenced on NovaSeq SP flowcell as 61-nt reads, generating 59, 66, and 64 M reads for each of the three libraries. Raw ATAC reads were trimmed by NGmerge, mapped to the genome using Bowtie2 (<https://github.com/BenLangmead/bowtie2>), and converted to BAM files using Samtools (<https://htslib.org/>). Duplicate reads were further removed by picard (<https://broadinstitute.github.io/picard/>), and the resulting reads were filtered by samtools. Peaks were called using MACS2 with following parameters: --nomodel -q 0.05 --keep-dup all --shift -100 --extsize 200 -g 2456432000--nolambda (93). IDR (irreproducible discovery rate) was used to assess peak calls concordance (94). Only peaks called in at least two of three pairwise IDR analyses were considered for downstream analyses. Peak data were visualized in IGV (<https://github.com/igvteam/igv>).

### RMC library construction

Bone marrow samples were collected from adult sugar gliders ( $n = 2$ ) as described above. Cells were incubated with 0.2% NaCl for 30 s, and red blood cells were removed by incubating the sample with 1.6% NaCl for 30 s. Micro-C libraries were generated from 3.2 and 3.3 M cells, using previously described methods (37, 95–97). Briefly, the samples were cross-linked with 4% formaldehyde (20 min and 500 rpm at room temperature), followed by the secondary cross-linking in 1 ml of freshly prepared 3 mM disuccinimidyl glutarate (Thermo

Fisher Scientific, 20593) and ethylene glycol bis(succinimidyl succinate) (Thermo Fisher Scientific, 21565) in Phosphate-buffered saline with Tween 20 (40 min, 500 rpm at room temperature). The reaction was quenched by adding 250  $\mu$ l of 2 M tris-HCl (pH 7.5) for 5 min, washed twice with PBS containing 3% bovine serum albumin (BSA), snap-frozen, and stored at  $-80^{\circ}\text{C}$ . For library construction, the samples were digested with micrococcal nuclease (concentration determined by titration) (20 min, 1000 rpm at room temperature) in MB1 buffer [50 mM NaCl, 10 mM tris-HCl (pH 7.5), 5 mM  $\text{MgCl}_2$ , 1 mM  $\text{CaCl}_2$ , 0.2% NP-40, and protease inhibitor cocktail (Sigma-Aldrich, 11697498001)] followed by heat inactivation with 5 mM EGTA at  $65^{\circ}\text{C}$  for 15 min. The samples were washed three times with MB2 buffer [50 mM NaCl, 10 mM tris-HCl (pH 7.5), 10 mM  $\text{MgCl}_2$ , and 1 mM  $\text{CaCl}_2$ ] and then treated with T4 polynucleotide kinase (0.5 U/ $\mu$ l; New England Biolabs M0201) (30 min, 1000 rpm at  $37^{\circ}\text{C}$ ) in End-prep buffer [1 $\times$  NEB buffer 2.1 (New England Biolabs, B7202), 2.5 mM adenosine triphosphate (ATP), and 5 mM dithiothreitol]. After T4 PNK treatment, the samples were incubated with Klenow Fragment (New England Biolabs M0210) at a final concentration of 0.5 U/ $\mu$ l (15 min, 1000 rpm at  $37^{\circ}\text{C}$ ). End-labeling cocktails were added to the samples to make the final concentration 50  $\mu$ M Biotin-dATP (Jena Bioscience NU-835-BIO14), 50  $\mu$ M biotin-deoxycytidine triphosphate (Jena Bioscience, #NU-809-BIOX), 50  $\mu$ M deoxyguanosine triphosphate, 50  $\mu$ M deoxythymidine triphosphate, 0.3 $\times$  T4 DNA ligase buffer (New England Biolabs B0202), and BSA (80  $\mu$ g/ml). The samples were incubated (45 min, 1000 rpm at room temperature) and heat inactivated with 5 mM EDTA (15 min at  $65^{\circ}\text{C}$ ). The samples were washed three times with MB3 buffer [40 mM tris-HCl (pH 7.5) and 10 mM  $\text{MgCl}_2$ ] and then treated with T4 DNA ligase (20 U/ $\mu$ l; New England Biolabs, M0202) (16 hours, 400 rpm at  $16^{\circ}\text{C}$ ) in ligation buffer [1 $\times$  T4 DNA ligase buffer (New England Biolabs, B0202) and BSA (200  $\mu$ g/ml)]. After ligation, the samples were treated with Exonuclease III (4 U/ $\mu$ l; New England Biolabs, M0206) in 1 $\times$  NEB buffer 1 (New England Biolabs, B7001) (30 min, 1000 rpm at  $37^{\circ}\text{C}$ ), followed by addition of ribonuclease A (0.5 mg/ml; Thermo Fisher Scientific, EN0531) (30 min, 1000 rpm at  $65^{\circ}\text{C}$ ). The samples were then treated with proteinase K (0.5 mg/ml) in 1% SDS (overnight at  $65^{\circ}\text{C}$ ). After proteinase K treatment, DNA libraries were extracted and purified from samples using standard phenol:chloroform extraction and ethanol precipitation. Libraries were sheared to 200-bp fragments using Covaris ME220 (duration, 130 s; peak power, 70 W; duty factor, 20%; and cycles per burst 1000). The libraries were then bound to Dynabeads MyOne Streptavidin C1 beads (Thermo Fisher Scientific, 65001) and washed following the manufacturer's instructions. The libraries were end repaired and ligated to adaptors using NEBNext Ultra II library prep kit (New England Biolabs, E7645). Next, the libraries were amplified with NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 2) (New England Biolabs, E7780) and KAPA HiFi Hot-Start ReadyMix (KAPA Biosystems, KK2601) and then purified with Ampure XP beads (Thermo Fisher Scientific, #NC9959336). The Micro-C libraries were then subject to quality control with Qubit and bioanalyzer. For RCMC, the custom probe sets were designed using the sequence of the target capture region (HiC\_scaffold\_9: 107950000-109750000; HiC\_scaffold\_81: 114775). Micro-C libraries were hybridized to the probe sets, pulled down, and amplified using Twist Standard Hyb and Wash Kit v2 (Twist Biosciences 105560). The final RCMC libraries were cleaned up with Ampure XP beads (Thermo Fisher Scientific, #NC9959336) and subjected to

paired-end sequencing on an Illumina Novaseq S1 100 nt Flowcell, yielding a total of 446 M reads.

### RCMC data processing

Micro-C data were aligned to custom genomes with BWA-MEM using parameters -S -P -5 -M (<https://github.com/lh3/bwa>). The resulting BAM files were parsed, sorted, de-duplicated, filtered, and split with Pairtools (<https://github.com/open2c/pairtools>). We removed pairs when only one half of the pair could be mapped or when the MAPQ score was less than three. The resulting files were indexed with Pairix (pairtools function). The files from replicates were merged with Pairtools (<https://github.com/open2c/pairtools>) before generating 100-bp contact matrices using Cooler (98). Last, balancing and mcool file generation was performed with Cooler's Zoomify tool and visualized on Hiclass (99).

### Transcription factor motif analysis

FIMO (find individual motif occurrences) motif scan was first conducted against putative enhancers of *Camp3* and *Camp4* as the expression levels, sequence identity of both genes, and enhancers are similar to each other, suggesting the existence of shared regulatory mechanism (100). Featureplots and violinplots from scRNA-seq datasets as well as RPKM values from RNA-seq datasets were cross-referenced to filter candidate transcription factors. Specifically, we restricted our analysis to transcription factors that had an RPKM value of  $>1$ . *C/ebp $\delta$* , *C/ebp $\epsilon$* , *Spi1*, *Fli1*, *Runx1*, *Foxp2*, and *Nfatc3* stood out as initial candidates, but *Foxp2* and *Nfatc3* were excluded as they did not bind to human and mouse cathelicidin locus (fig. S8) (101). Further FIMO motif scans were conducted on additional putative regulatory peaks to search for *C/ebp* family binding sites.

### Luciferase assays

Sequences of a putative enhancer region of *Camp3* was PCR amplified from sugar glider genomic DNA (table S1) and cloned into the pGL4.23 luciferase enhancer reporter vector using In-fusion cloning (Takara). Sequences for the different transcription factors and *Camp3* promoter were obtained either commercially (*Camp3* promoter, *Spi1*, *Fli1*, *Runx1*, *Irf1*, and *C/ebp $\epsilon$* ; Twist Biosciences) or by extracting RNA from the bone marrow of an adult sugar glider and reverse transcribing it using qScript cDNA SuperMix (Quantabio) (*C/ebp $\delta$* ) (table S1). Transcription factor fragments were then cloned into the pCMV-GFP vector. A total of 2000 immortalized sugar glider fibroblast cells were seeded in a white 96-well plate (PerkinElmer) and transfected with the experimental constructs (200 ng) as well as a control pGL4.74 renilla reporter vector (20 ng) using 0.3  $\mu$ l of Lipofectamine (Invitrogen). After 48 hours of incubation at  $37^{\circ}\text{C}$ , the cells were harvested and analyzed with the DualGlo Luciferase Assay System (Promega). Luminescence was measured by Tecan Spark microplate reader. Experiments were performed in six technical replicates of two biological replicates except for a GFP control, which has five replicates due to low renilla transfection efficiency in one of the two biological replicates. Experiments were statistically analyzed using an unpaired *t* test in GraphPad Prism.

### Pouch swabbing

Sugar glider pouches were swabbed for 30 s with cotton swabs pre-moistened with ultrapure water. Swabs resuspended in ultrapure water were plated in LB media and incubated overnight at  $37^{\circ}\text{C}$ . White colonies that repeatedly grew from pouch swabs of multiple females were isolated and subject to DNA extraction (Zymo, D6005). Bacterial

species were identified by Sanger sequencing of the PCR-amplified 16S ribosomal RNA gene (827F: 5'AGA GTT TGA TCC TGG CTC 3', 1492R: 5' TAC GGY TAC CTT GTT ACG ACT 3').

### In vitro broth micro-dilution antibacterial assay

The first alanine or valine found in the fourth exon of CAMP1, CAMP3, CAMP4, NGP, CAMP7, CAMP9, and CAMP10, were predicted to represent the cleavage sites for the mature peptides (27). For CAMP8, because of the short length (15 amino acids) and the lack of alanine and valine in the fourth exon, we set the cleavage site to be the last alanine in the third exon. For CAMP7, because of the short length of the predicted mature peptide after the valine, we synthesized three versions of mature peptides with predicted cleavage sites located, respectively, in the first leucine, the isoleucine, and the valine found in the fourth exon. We did not synthesize mature peptides of CAMP6 because of its extensive length (137 amino acids) nor of the pseudogenes *Camp2* and *Camp5*. Additional furin cleavage sites with RxxR minimal consensus motif were also noted (table S2) (102). Secondary structures of synthesized mature peptides were predicted using PEP-FOLD3 (103).

We used broth micro-dilution assays (104) to measure minimal inhibitory concentration (MIC) of cathelicidins against *E. coli* [American Type Culture Collection (ATCC) 25922], *K. pneumoniae* (ATCC 43816), *M. luteus* (ATCC 4698), *P. aeruginosa* (PAO1, PA14), *B. bronchiseptica* (ATCC 10580), *S. aureus* (ATCC 29213, MRSA), and *Kocuria* sp. Briefly, we added  $10^5$  CFU of bacteria from overnight cultures into each well of a 96-well plate (costar, Corning) with different concentrations of each peptide or antibiotics. Two hundred microliters of cation-adjusted Mueller-Hinton broth was added to each well. The samples were incubated at 37°C, and absorbance was measured at 0 and 16 hours using the Tecan Spark microplate reader (600 nm). For *M. luteus*, we measured absorbance at 43 hours as well. MIC was determined as the lowest concentration that showed no visible change after 16 hours (43 hours for *M. luteus*) incubation for all three replicates and showed >90% mean percent inhibition based on absorbance measurements.

### Cell viability assay

Cell viability was measured using CellTiter-Glo 2.0 Assay (Promega) following the manufacturer's protocol. Briefly, we incubated 10,000 cells in 100  $\mu$ l of Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, sodium bicarbonate (1.5 g/liter), 1 mM sodium pyruvate, and 1% penicillin-streptomycin mix overnight (37°C) with different concentrations of cathelicidins and an SDS control. Subsequently, we added 100  $\mu$ l of CellTiter-Glo 2.0 reagent. After 2 min of mixing and a 10-min incubation at room temperature, we measured luminescence using Tecan Spark microplate reader. Experiments were done in triplicate.

### PI absorption assay

Bacteria were grown overnight to log phase. Bacteria ( $10^6$  CFU) were incubated for 30 min at 37°C with different concentrations of cathelicidin peptides in 200  $\mu$ l of PBS. After a 2-min centrifugation at 8000 rpm, the pellets were washed with 200  $\mu$ l of PBS and resuspended in 1 ml of PBS. PI was added for a final concentration of 1  $\mu$ g/ml. An LSRII flow cytometer (BD Biosciences) was used to record 50,000 events of PI uptake. SDS (0.25 to 1%) was used as a control.

### In vivo mouse infection

Six- to eight-week-old female mice (C57BL/6J, The Jackson Laboratory) were intraperitoneally injected with  $5 \times 10^7$  CFU of *E. coli* (ATCC 25922) in 100  $\mu$ l of PBS. Thirty min after the bacterial injection, the mice were treated with either CAMP4 peptide (10 or 20 mg/ml) or PBS vehicle. Treated mice were checked daily for 7 days. All experiments performed were approved by the IACUC committee at Princeton University (2155 and 3002).

### LPS-binding assay

Mouse macrophage cells (J774A.1; ATCC TIB-67) were grown in supplemented DMEM in 7.5% CO<sub>2</sub> at 37°C. Cells ( $2.5 \times 10^5$ ) were incubated with LPS-FITC (10  $\mu$ g/ml) and different concentrations of antimicrobial peptides in 500  $\mu$ l of PBS for 30 min. After washing the cells using PBS, median fluorescence intensity was determined using the LSRII flow cytometer (BD sciences). The percentage difference from negative control was statistically analyzed with unpaired *t* tests in GraphPad Prism. Experiments were done in quadruplicate.

### Circular dichroism

The secondary structure of CAMP3 and CAMP4, which showed LPS-binding activity in at least one of the two tested concentrations, were determined using circular dichroism spectrometer (Chirascan, Photophysics). Measurements were taken in a rectangular cuvette using an emission range of 180 to 280 nm (in 1-nm increments). Peptides (0.1 mg/ml) and LPS (0.2 mg/ml) in sterile water were used for measurements.

### TNF- $\alpha$ secretion assay

ELISA assays were performed according to manufacturer's instructions (R&D Systems DuoSet). Briefly, a 96-well plate (Corning Costar) was coated with anti-TNF- $\alpha$  and incubated overnight at room temperature. The plate was then washed with wash buffer, blocked for 1 hour with reagent diluent, washed again, and the liquid aspirated. J774A.1 macrophages were incubated with LPS, cathelicidin, vehicle, or a combination of these for 24 hours and 100  $\mu$ l of supernatant from the incubating macrophages or cytokine standard, in sequentially diluted concentrations ranging from 2000 to 32 pg/ml, were applied to the wells and incubated for 2 hours at room temperature before being washed and aspirated. Biotinylated anti-TNF- $\alpha$ , streptavidin-horseradish peroxidase (HRP), and substrate solution were applied in sequence. The plate was then incubated in the dark for 2 hours or 20 min and subsequently washed and aspirated. The HRP-substrate reaction proceeded for 20 min before the stop solution halted it. The plate was then read for optical density at 450 and 540 nm. The percentage difference from the LPS control was statistically analyzed with unpaired *t* tests in GraphPad Prism. Experiments were done in quadruplicate.

### Pairwise sequence identity analysis

Amino acid sequences of eight cathelicidin genes (*Camp1*, *Camp3*, *Camp4*, *Camp6*, *Camp7*, *Camp8*, *Camp9*, and *Camp10*) were aligned with MAFFT v7.475 in a pairwise manner. Resulting alignments were fed into Sequence Manipulation Suite for identity analysis (data table 34) (105). For statistical analysis, separate pairwise analysis was conducted for each exon. In this analysis, we used the truncated, Maker3-annotated version of *Camp6* due to the extensive length of the

predicted protein compared to the rest of the cathelicidin genes. Nucleotide sequences of *Camp3*, *Camp4*, and *Camp10* were further aligned and compared for detailed analysis. Statistical analysis was conducted in GraphPad Prism using one-way analysis of variance (ANOVA) test. Post hoc pairwise comparisons were carried out using a Bonferroni correction.

## Supplementary Materials

### The PDF file includes:

Figs. S1 to S13

Tables S1 to S7

Legends for data tables S1 to S4

References

### Other Supplementary Material for this manuscript includes the following:

Data tables S1 to S4

## REFERENCES AND NOTES

- M. C. de Goffau, S. Lager, U. Sovio, F. Gaccioli, E. Cook, S. J. Peacock, J. Parkhill, D. S. Charnock-Jones, G. C. S. Smith, Human placenta has no microbiome but can contain potential pathogens. *Nature* **572**, 329–334 (2019).
- A. A. Kuperman, A. Zimmerman, S. Hamadia, O. Ziv, V. Gurevich, B. Fichtman, N. Gavert, R. Strausman, H. Rechnitzer, M. Barzilay, S. Shvalb, J. Bornstein, I. Ben-Shachar, S. Yagel, I. Haviv, O. Koren, Deep microbial analysis of multiple placentas shows no evidence for a placental microbiome. *BJOG* **127**, 159–169 (2020).
- J. E. Lawn, S. Cousens, J. Zupan, Lancet neonatal survival steering team, 4 million neonatal deaths: When? Where? Why? *Lancet* **365**, 891–900 (2005).
- C. C. L. Chase, D. J. Hurley, A. J. Reber, Neonatal immune development in the calf and its impact on vaccine response. *Vet. Clin. North Am. Food Anim. Pract.* **24**, 87–104 (2008).
- A. C. Jones, G. C. Burdige, J. A. Warner, A. D. Postle, J. O. Warner, Ontogeny of circulating leucocytes in the fetal guinea pig. *Biol. Neonate* **70**, 108–115 (2004).
- M. P. Holsapple, L. J. West, K. S. Landreth, Species comparison of anatomical and functional immune system development. *Birth Defects Res. B Dev. Reprod. Toxicol.* **68**, 321–334 (2003).
- S. D. Holladay, R. J. Smialowicz, Development of the murine and human immune system: Differential effects of immunotoxicants depend on time of exposure. *Environ. Health Perspect.* **108**, 463–473 (2000).
- Z.-X. Luo, C.-X. Yuan, Q.-J. Meng, Q. Ji, A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature* **476**, 442–445 (2011).
- M. dos Reis, J. Inoue, M. Hasegawa, R. J. Asher, P. C. J. Donoghue, Z. Yang, Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. Biol. Sci.* **279**, 3491–3500 (2012).
- C. R. Borthwick, L. J. Young, J. M. Old, The development of the immune tissues in marsupial pouch young. *J. Morphol.* **275**, 822–839 (2014).
- J. M. Old, Haematopoiesis in Marsupials. *Dev. Comp. Immunol.* **58**, 40–46 (2016).
- J. E. Deakin, D. W. Cooper, Characterisation of and immunity to the aerobic bacteria found in the pouch of the brushtail possum *Trichosurus vulpecula*. *Comp. Immunol. Microbiol. Infect. Dis.* **27**, 33–46 (2004).
- R. Osawa, W. H. Blanshard, P. G. O'Callaghan, Microflora of the pouch of the koala (*Phascolarctos cinereus*). *J. Wildl. Dis.* **28**, 276–280 (1992).
- Y. Cheng, K. Belov, Antimicrobial protection of marsupial pouch young. *Front. Microbiol.* **8**, 354 (2017).
- K. M. Morris, D. O'Meally, T. Zaw, X. Song, A. Gillett, M. P. Molloy, A. Polkinghorne, K. Belov, Characterisation of the immune compounds in koala milk using a combined transcriptomic and proteomic approach. *Sci. Rep.* **6**, 35011 (2016).
- K. Ambatipudi, J. Joss, M. Raftery, E. Deane, A proteomic approach to analysis of antimicrobial activity in marsupial pouch secretions. *Dev. Comp. Immunol.* **32**, 108–120 (2008).
- G. Bobek, E. M. Deane, Possible antimicrobial compounds from the pouch of the koala, *Phascolarctos cinereus*. *Lett. Pept. Sci.* **8**, 133–137 (2001).
- E. Peel, Y. Cheng, J. T. Djordjevic, S. Fox, T. C. Sorrell, K. Belov, Cathelicidins in the Tasmanian devil (*Sarcophilus harrisii*). *Sci. Rep.* **6**, 35019 (2016).
- C. Petrohilos, A. Patchett, C. J. Hogg, K. Belov, E. Peel, Tasmanian devil cathelicidins exhibit anticancer activity against Devil Facial Tumour Disease (DFTD) cells. *Sci. Rep.* **13**, 12698 (2023).
- R. V. Hewavisenti, K. M. Morris, D. O'Meally, Y. Cheng, A. T. Papenfuss, K. Belov, The identification of immune genes in the milk transcriptome of the Tasmanian devil (*Sarcophilus harrisii*). *PeerJ* **4**, e1569 (2016).
- C. R. Borthwick, L. J. Young, J. M. Old, An examination of the development and localization of key immune cells in developing pouch young of the red-tailed phascogale (*Phascogale calura*). *Anat. Rec.* **302**, 1985–2002 (2019).
- J. M. Old, E. M. Deane, Immunohistochemistry of the lymphoid tissues of the tammar wallaby, *Macropus eugenii*. *J. Anat.* **201**, 257–266 (2002).
- S. W. Hemsley, P. J. Canfield, A. J. Husband, Immunohistological staining of lymphoid tissue in four Australian marsupial species using species cross-reactive antibodies. *Immunol. Cell Biol.* **73**, 321–325 (1995).
- G. L. Burn, A. Foti, G. Marsman, D. F. Patel, A. Zychlinsky, The neutrophil. *Immunity* **54**, 1377–1391 (2021).
- P. A. Cisternas, P. J. Armati, Development of the thymus, spleen, lymph nodes and liver in the marsupial, *Isoodon macrourus* (Northern brown bandicoot, Peramelidae). *Anat. Embryol.* **200**, 433–443 (1999).
- J. M. Old, L. Selwood, E. M. Deane, A developmental investigation of the liver, bone marrow and spleen of the stripe-faced dunnart (*Sminthopsis macroura*). *Dev. Comp. Immunol.* **28**, 347–355 (2004).
- A. E. Shinnar, K. L. Butler, H. J. Park, Cathelicidin family of antimicrobial peptides: Proteolytic processing and protease resistance. *Bioorg. Chem.* **31**, 425–436 (2003).
- M. A. Alford, B. Baquir, F. L. Santana, E. F. Haney, R. E. W. Hancock, Cathelicidin host defense peptides and inflammatory signaling: Striking a balance. *Front. Microbiol.* **11**, 1902 (2020).
- K. A. Daly, M. R. Digby, C. Lefèvre, K. R. Nicholas, E. M. Deane, P. Williamson, Identification, characterization and expression of cathelicidin in the pouch young of tammar wallaby (*Macropus eugenii*). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **149**, 524–533 (2008).
- E. Peel, Y. Cheng, J. T. Djordjevic, D. O'Meally, M. Thomas, M. Kuhn, T. C. Sorrell, W. M. Huston, K. Belov, Koala cathelicidin Phc1Cath5 has antimicrobial activity, including against *Chlamydia pecorum*. *PLOS ONE* **16**, e0249658 (2021).
- E. Peel, Y. Cheng, J. T. Djordjevic, M. Kuhn, T. Sorrell, K. Belov, Marsupial and monotreme cathelicidins display antimicrobial activity, including against methicillin-resistant *Staphylococcus aureus*. *Microbiology* **163**, 1457–1465 (2017).
- K. Belov, C. E. Sanderson, J. E. Deakin, E. S. W. Wong, D. Assange, K. A. McColl, A. Gout, B. de Bono, A. D. Barrow, T. P. Speed, J. Trowsdale, A. T. Papenfuss, Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Res.* **17**, 982–991 (2007).
- C. Y. Feigin, J. A. Moreno, R. Ramos, S. A. Mereby, A. Alivisatos, W. Wang, R. van Amerongen, J. Camacho, J. J. Rasweiler IV, R. R. Behringer, B. Ostrow, M. V. Plikus, R. Mallarino, Convergent deployment of ancestral functions during the evolution of mammalian flight membranes. *Sci. Adv.* **9**, eade7511 (2023).
- J. D. Buenostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- M. R. Corces, A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-Armstrong, S. Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria, S. W. Cho, M. R. Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff, V. I. Risca, A. Kundaje, P. A. Khavari, T. J. Montine, W. J. Greenleaf, H. Y. Chang, An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
- J. A. Moreno, O. Dudchenko, C. Y. Feigin, S. A. Mereby, Z. Chen, R. Ramos, A. A. Almet, H. Sen, B. J. Brack, M. R. Johnson, S. Li, W. Wang, J. M. Gaska, A. Ploss, D. Weisz, A. D. Omer, W. Yao, Z. Colaric, P. Kaur, J. S. Leger, Q. Nie, A. Mena, J. P. Flanagan, G. Keller, T. Sanger, B. Ostrow, M. V. Plikus, E. Z. Kvon, E. L. Aiden, R. Mallarino, Emx2 underlies the development and evolution of marsupial gliding membranes. *Nature* **629**, 127–135 (2024).
- V. Y. Goel, M. K. Huseyin, A. S. Hansen, Region capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nat. Genet.* **55**, 1048–1056 (2023).
- O. E. Sørensen, P. Follin, A. H. Johnsen, J. Calafat, G. S. Tjabringa, P. S. Hiemstra, N. Borregaard, Human cathelicidin, hCAP-18, is processed to the antimicrobial peptide LL-37 by extracellular cleavage with proteinase 3. *Blood* **97**, 3951–3959 (2001).
- A. M. Cole, J. Shi, A. Ceccarelli, Y. H. Kim, A. Park, T. Ganz, Inhibition of neutrophil elastase prevents cathelicidin activation and impairs clearance of bacteria from wounds. *Blood* **97**, 297–304 (2001).
- M. Scocchi, B. Skerlavaj, D. Romeo, R. Gennaro, Proteolytic cleavage by neutrophil elastase converts inactive storage proforms to antibacterial bactenecins. *Eur. J. Biochem.* **209**, 589–595 (1992).
- A. Panyutich, J. Shi, P. L. Boutz, C. Zhao, T. Ganz, Porcine polymorphonuclear leukocytes generate extracellular microbicidal activity by elastase-mediated activation of secreted proteoglycans. *Infect. Immun.* **65**, 978–985 (1997).
- R. L. Gallo, K. J. Kim, M. Bernfield, C. A. Kozak, M. Zanetti, L. Merluzzi, R. Gennaro, Identification of CRAMP, a cathelin-related antimicrobial peptide expressed in the embryonic and adult mouse. *J. Biol. Chem.* **272**, 13088–13093 (1997).

43. B. Agerberth, H. Gunne, J. Odeberg, P. Kogner, H. G. Boman, G. H. Gudmundsson, FALL-39, a putative human peptide antibiotic, is cysteine-free and expressed in bone marrow and testis. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 195–199 (1995).
44. L. Boulos, M. Prévost, B. Barbeau, J. Coallier, R. Desjardins, LIVE/DEAD® BacLight™: Application of a new rapid staining method for direct enumeration of viable and total bacteria in drinking water. *J. Microbiol. Methods* **37**, 77–86 (1999).
45. Y. R. Bommineni, M. Achanta, J. Alexander, L. T. Sunkara, J. W. Ritchey, G. Zhang, A fowlcidin-1 analog protects mice from lethal infections induced by methicillin-resistant *Staphylococcus aureus*. *Peptides* **31**, 1225–1230 (2010).
46. N. Soundararajan, S. Park, Q. Le Van Chanh, H.-S. Cho, G. Raghunathan, B. Ahn, H. Song, J.-H. Kim, C. Park, Protegrin-1 cytotoxicity towards mammalian cells positively correlates with the magnitude of conformational changes of the unfolded form upon cell interaction. *Sci. Rep.* **9**, 11569 (2019).
47. J. Koziel, D. Bryzek, A. Sroka, K. Maresz, I. Glowczyk, E. Bielecka, T. Kantyka, K. Pyrc, P. Svoboda, J. Pohl, J. Potempa, Citrullination alters immunomodulatory function of LL-37 essential for prevention of endotoxin-induced sepsis. *J. Immunol.* **192**, 5363–5372 (2014).
48. M. G. Scott, D. J. Davidson, M. R. Gold, D. Bowdish, R. E. W. Hancock, The human antimicrobial peptide LL-37 is a multifunctional modulator of innate immune responses. *J. Immunol.* **169**, 3883–3891 (2002).
49. D. Yang, Q. Chen, A. P. Schmidt, G. M. Anderson, J. M. Wang, J. Wooters, J. J. Oppenheim, O. Chertov, LI-37, the neutrophil granule- and epithelial cell-derived cathelicidin, utilizes formyl peptide receptor-like 1 (Fpr1) as a receptor to chemoattract human peripheral blood neutrophils, monocytes, and T cells. *J. Exp. Med.* **192**, 1069–1074 (2000).
50. D. Minns, K. J. Smith, V. Alessandrini, G. Hardisty, L. Melrose, L. Jackson-Jones, A. S. MacDonald, D. J. Davidson, E. Gwyer Findlay, The neutrophil antimicrobial peptide cathelicidin promotes Th17 differentiation. *Nat. Commun.* **12**, 1285 (2021).
51. N. J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2006).
52. A. J. Lewis, C. W. Seymour, M. R. Rosengart, Current murine models of sepsis. *Surg. Infect.* **17**, 385–393 (2016).
53. A. Shrestha, D. Duwadi, J. Jukosky, S. N. Fiering, Cecropin-like antimicrobial peptide protects mice from lethal *E. coli* infection. *PLOS ONE* **14**, e0220344 (2019).
54. K. Ferner, J. A. Schultz, U. Zeller, Comparative anatomy of neonates of the three major mammalian groups (monotremes, marsupials, placentals) and implications for the ancestral mammalian neonate morphology. *J. Anat.* **231**, 798–822 (2017).
55. H. Steiner, D. Hultmark, A. Engström, H. Bennich, H. G. Boman, Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature* **292**, 246–248 (1981).
56. J. M. Old, E. M. Deane, Development of the immune system and immunological protection in marsupial pouch young. *Dev. Comp. Immunol.* **24**, 445–454 (2000).
57. M. Cardoso-Moreira, J. Halbert, D. Valloton, B. Velten, C. Chen, Y. Shao, A. Liechti, K. Ascensão, C. Rummel, S. Ovchinnikova, P. V. Mazin, I. Xenarios, K. Harshman, M. Mort, D. N. Cooper, C. Sandi, M. J. Soares, P. G. Ferreira, S. Afonso, M. Carneiro, J. M. A. Turner, J. L. VandeBerg, A. Fallahshahroudi, P. Jensen, R. Behr, S. Ligo, S. Lindsay, P. Khaitovich, W. Huber, J. Baker, S. Anders, Y. E. Zhang, H. Kaessmann, Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).
58. C. Qiu, B. K. Martin, I. C. Welsh, R. M. Daza, T.-M. Le, X. Huang, E. K. Nichols, M. L. Taylor, O. Fulton, D. R. O'Day, A. R. Gomes, S. Ilcisin, S. Srivatsan, X. Deng, C. M. Distech, W. S. Noble, N. Hamazaki, C. B. Moens, D. Kimelman, J. Cao, A. F. Schier, M. Spielmann, S. A. Murray, C. Trapnell, J. Shendure, A single-cell time-lapse of mouse prenatal development from gastrula to birth. *Nature* **626**, 1084–1093 (2024).
59. J. R. White, H. Gong, T. T. Colaizy, J. G. Moreland, H. Flaherty, S. J. McElroy, Evaluation of hematologic variables in newborn C57/BL6 mice up to day 35. *Vet. Clin. Pathol.* **45**, 87–95 (2016).
60. M. A. Hanson, L. Grollmus, B. Lemaitre, Ecology-relevant bacteria drive the evolution of host antimicrobial peptides in *Drosophila*. *Science* **381**, eadg5725 (2023).
61. M. A. Hanson, B. Lemaitre, Repeated truncation of a modular antimicrobial peptide gene for neural context. *PLOS Genet.* **18**, e1010259 (2022).
62. H. Kiyonari, M. Kaneko, T. Abe, A. Shiraishi, R. Yoshimi, K.-I. Inoue, Y. Furuta, Targeted gene disruption in a marsupial, *Monodelphis domestica*, by CRISPR/Cas9 genome editing. *Curr. Biol.* **31**, 3956–3963.e4 (2021).
63. R. L. Gallo, L. V. Hooper, Epithelial antimicrobial defence of the skin and intestine. *Nat. Rev. Immunol.* **12**, 503–516 (2012).
64. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexis, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
65. L. V. Yang, R. H. Nicholson, J. Kaplan, A. Galy, L. Li, Hemogen is a novel nuclear factor specifically expressed in mouse hematopoietic development and its human homologue EDAG maps to chromosome 9q22, a region containing breakpoints of hematological neoplasms. *Mech. Dev.* **104**, 105–111 (2001).
66. S. Elliott, A. M. Sinclair, The effect of erythropoietin on normal and neoplastic cells. *Biol. Theory* **6**, 163–189 (2012).
67. B. S. Garrison, A. P. Rybak, I. Beeran, B. Heesters, F. E. Mercier, D. T. Scadden, D. Bryder, R. Baron, D. J. Rossi, ZFP521 regulates murine hematopoietic stem cell function and facilitates MLL-AF9 leukemogenesis in mouse and human cells. *Blood* **130**, 619–624 (2017).
68. D. B. AbuSamra, F. A. Aleisa, A. S. Al-Amoodi, H. M. Jalal Ahmed, C. J. Chin, A. F. Abuelela, P. Bergam, R. Sougrat, J. S. Merzaban, Not just a marker: CD34 on human hematopoietic stem/progenitor cells dominates vascular selectin binding along with CD44. *Blood Adv.* **1**, 2799–2816 (2017).
69. M. Evrard, I. W. H. Kwok, S. Z. Chong, K. W. W. Teng, E. Becht, J. Chen, J. L. Sieow, H. L. Penny, G. C. Ching, S. Devi, J. M. Adrover, J. L. Y. Li, K. H. Liang, L. Tan, Z. Poon, S. Foo, J. W. Chua, I.-H. Su, K. Balabanian, F. Bachelier, S. K. Biswas, A. Larbi, W. Y. K. Hwang, V. Madan, H. P. Koeffler, S. C. Wong, E. W. Newell, A. Hidalgo, F. Ginhoux, L. G. Ng, Developmental analysis of bone marrow neutrophils reveals populations specialized in expansion, trafficking, and effector functions. *Immunity* **48**, 364–379.e8 (2018).
70. S. R. Horman, C. S. Velu, A. Chaubey, T. Bourdeau, J. Zhu, W. E. Paul, B. Gebelein, H. L. Grimes, Gfi1 integrates progenitor versus granulocytic transcriptional programming. *Blood* **113**, 5466–5475 (2009).
71. X. Xie, Q. Shi, P. Wu, X. Zhang, H. Kambara, J. Su, H. Yu, S.-Y. Park, R. Guo, Q. Ren, S. Zhang, Y. Xu, L. E. Silberstein, T. Cheng, F. Ma, C. Li, H. R. Luo, Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. *Nat. Immunol.* **21**, 1119–1133 (2020).
72. P. A. Szabo, H. M. Levitin, M. Miron, M. E. Snyder, T. Senda, J. Yuan, Y. L. Cheng, E. C. Bush, P. Dogra, P. Thapa, D. L. Farber, P. A. Sims, Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019).
73. D. Morgan, V. Tergaonkar, Unraveling B cell trajectories at single cell resolution. *Trends Immunol.* **43**, 210–229 (2022).
74. M. Hamada, Y. Tsunakawa, H. Jeon, M. K. Yadav, S. Takahashi, Role of MafB in macrophages. *Exp. Anim.* **69**, 1–10 (2020).
75. L. M. Kelly, U. Englmeier, I. Lafon, M. H. Sieweke, T. Graf, MafB is an inducer of monocytic differentiation. *EMBO J.* **19**, 1987–1997 (2000).
76. V. D. Cuevas, L. Anta, R. Samaniego, E. Orta-Zavalza, J. Vladimir de la Rosa, G. Baujat, Á. Domínguez-Soto, P. Sánchez-Mateos, M. M. Escríbese, A. Castrillo, V. Cormier-Daire, M. A. Vega, Á. L. Corbí, MAFB determines human macrophage anti-inflammatory polarization: Relevance for the pathogenic mechanisms operating in multicentric carpal osteolysis. *J. Immunol.* **198**, 2070–2081 (2017).
77. K. Kang, S. H. Park, J. Chen, Y. Qiao, E. Giannopoulou, K. Berg, A. Hanidu, J. Li, G. Nabozny, K. Kang, K.-H. Park-Min, L. B. Ivashkiv, Interferon- $\gamma$  represses M2 gene expression in human macrophages by disassembling enhancers bound by the transcription factor MAF. *Immunity* **47**, 235–250.e4 (2017).
78. J. Fleischer, E. Soeth, N. Reiling, E. Grage-Griebenow, H. D. Flad, M. Ernst, Differential expression and function of CD80 (B7-1) and CD86 (B7-2) on human peripheral blood monocytes. *Immunology* **89**, 592–598 (1996).
79. H. Fujita, Y. Hashimoto, S. Russell, B. Zieger, J. Ware, In vivo expression of murine platelet glycoprotein Iba. *Blood* **92**, 488–495 (1998).
80. Q. Kimmerlin, C. Strassel, A. Eckly, F. Lanza, The tubulin code in platelet biogenesis. *Semin. Cell Dev. Biol.* **137**, 63–73 (2023).
81. G. Varricchi, A. Pecoraro, S. Loffredo, R. Poto, F. Rivellese, A. Genovese, G. Marone, G. Spadaro, Heterogeneity of human mast cells with respect to MRGPRX2 receptor expression and function. *Front. Cell. Neurosci.* **13**, 299 (2019).
82. G. Cruse, D. Kaur, M. Leyland, P. Bradding, A novel Fc $\epsilon$ RI $\beta$ -chain truncation regulates human mast cell proliferation and survival. *FASEB J.* **24**, 4047–4057 (2010).
83. K. R. Acharya, S. J. Ackerman, Eosinophil granule proteins: Form and function. *J. Biol. Chem.* **289**, 17406–17415 (2014).
84. J. Tavernier, R. Devos, S. Cornelis, T. Tuybens, J. Van der Heyden, W. Fiers, G. Plaetinck, A human high affinity interleukin-5 receptor (IL5R) is composed of an IL5-specific alpha chain and a beta chain shared with the receptor for GM-CSF. *Cell* **66**, 1175–1184 (1991).
85. G. L. Humason, *Animal Tissue Techniques*, (WH Freeman and Company, 1979) pp. 661.
86. B. L. Cantarel, I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sánchez Alvarado, M. Yandell, MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
87. M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
88. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
89. A. Shumate, S. L. Salzberg, LiftOff: Accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).
90. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
91. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

92. S. R. Amend, K. C. Valkenburg, K. J. Pienta, Murine hind limb long bone dissection and bone marrow isolation. *J. Vis. Exp.* 53936 (2016).
93. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoutte, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
94. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments. *AOAS* **5**, 1752–1779 (2011).
95. E. Slobodyanyuk, C. Cattoglio, T.-H. S. Hsieh, Mapping mammalian 3D genomes by Micro-C. *Methods Mol. Biol.* **2532**, 51–71 (2022).
96. T.-H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tijian, X. Darzacq, Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553.e8 (2020).
97. W. Ke, M. Fujitaka, P. Schedl, J. B. Jaynes, Stem-loop and circle-loop TADs generated by directional pairing of boundary elements have distinct physical and regulatory properties. *eLife* **13**, RP94114 (2024).
98. N. Abdennur, L. A. Mirny, Cooler: Scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
99. P. Kerpedjiev, N. Abdennur, F. Lekschas, C. M. Callum, K. Dinkla, H. Strobel, J. M. Luber, S. B. Ouellette, A. Azhir, N. Kumar, J. Hwang, S. Lee, B. H. Alver, H. Pfister, L. A. Mirny, P. J. Park, N. Gehlenborg, HiGlass: Web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
100. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
101. Z. Zou, T. Ohta, S. Oki, ChIP-Atlas 3.0: A data-mining suite to explore chromosome architecture together with large-scale regulome data. *Nucleic Acids Res.* **52**, W45–W53 (2024).
102. G. Thomas, Furin at the cutting edge: From protein traffic to embryogenesis and disease. *Nat. Rev. Mol. Cell Biol.* **3**, 753–766 (2002).
103. A. Lamiable, P. Thévenet, J. Rey, M. Vavrusa, P. Derreumaux, P. Tufféry, PEP-FOLD3: Faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **44**, W449–W454 (2016).
104. I. Wiegand, K. Hilpert, R. E. W. Hancock, Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **3**, 163–175 (2008).
105. P. Stothard, The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**, 1102, 1104 (2000).
106. B. D. Rosen, D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, E. Tseng, T. N. Rowan, W. Y. Low, A. Zimin, C. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghuray, S. D. McKay, F. Thibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDanel, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schultheiss, S. G. Schroeder, A. M. Phillippy, J. B. Cole, C. P. Van Tassel, G. Liu, T. P. L. Smith, J. F. Medrano, De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **9**, giaa021 (2020).
107. B. P. Bentley, T. Carrasco-Valenzuela, E. K. S. Ramos, H. Pawar, L. S. Arantes, A. Alexander, S. M. Banerjee, P. Masterson, M. Kuhlwiilm, M. Pippel, J. Mountcastle, B. Haase, M. Uliano-Silva, G. Formenti, K. Howe, W. Chow, A. Tracey, Y. Sims, S. Pelan, J. Wood, K. Yetzko, J. R. Perrault, K. Stewart, S. R. Benson, Y. Levy, E. V. Todd, H. B. Shaffer, P. Scott, B. T. Henen, R. W. Murphy, D. W. Mohr, A. F. Scott, D. J. Duffy, N. J. Gemmill, A. Suh, S. Winkler, F. Thibaud-Nissen, M. F. Nery, T. Marques-Bonet, A. Antunes, Y. Tikochinski, P. H. Dutton, O. Fedrigo, E. W. Myers, E. D. Jarvis, C. J. Mazzoni, L. M. Komoroske, Divergent sensory and immune gene evolution in sea turtles with contrasting demographic and life histories. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2201076120 (2023).
108. T. S. Kalbfleisch, E. S. Rice, M. S. De Priest Jr., B. P. Walenz, M. S. Hestand, J. R. Vermeesch, B. L. O. Connell, I. T. Fiddes, A. O. Vershina, N. F. Saremi, J. L. Petersen, C. J. Finno, R. R. Bellone, M. E. McCue, S. A. Brooks, E. Bailey, L. Orlando, R. E. Green, D. C. Miller, D. F. Antczak, J. N. MacLeod, Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun Biol.* **1**, 197 (2018).
109. V. C. Mason, G. Li, P. Minx, J. Schmitz, G. Churakov, L. Doronina, A. D. Melin, N. J. Dominy, N. T.-L. Lim, M. S. Springer, R. K. Wilson, W. C. Warren, K. M. Helgen, W. J. Murphy, Genomic analysis reveals hidden biodiversity within colugos, the sister group to primates. *Sci. Adv.* **2**, e1600633 (2016).
110. V. A. Schneider, T. Graves-Lindsay, K. Howe, N. Bouk, H.-C. Chen, P. A. Kitts, T. D. Murphy, K. D. Pruitt, F. Thibaud-Nissen, D. Albracht, R. S. Fulton, M. Kremitzki, V. Magrini, C. Markovic, S. M. Grath, K. M. Steinberg, K. Auger, W. Chow, J. Collins, G. Harden, T. Hubbard, S. Pelan, J. T. Simpson, G. Threadgold, J. Torrance, J. M. Wood, L. Clarke, S. Koren, M. Boitano, P. Peluso, H. Li, C.-S. Chin, A. M. Phillippy, R. Durbin, R. K. Wilson, S. Flicek, E. E. Eichler, D. M. Church, Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).
111. J.-Y. Hu, Z.-Q. Hao, L. Frantz, S.-F. Wu, W. Chen, Y.-F. Jiang, H. Wu, W.-M. Kuang, H. Li, Y.-P. Zhang, L. Yu, Genomic consequences of population decline in critically endangered pangolins and their demographic histories. *Natl. Sci. Rev.* **7**, 798–814 (2020).
112. T. S. Mikkelsen, M. J. Wakefield, B. Aken, C. T. Amemiya, J. L. Chang, S. Duke, M. Garber, A. J. Gentles, L. Goodstadt, A. Heger, J. Jurka, M. Kamal, E. Mauceli, S. M. J. Searle, T. Sharpe, M. L. Baker, M. A. Batzer, P. V. Benos, K. Belov, M. Clamp, A. Cook, J. Cuff, R. Das, L. Davidow, J. E. Deakin, M. J. Fazzari, J. L. Glass, M. Grabherr, J. M. Greally, W. Gu, T. A. Hore, G. A. Huttley, M. Kleber, R. L. Jirtle, E. Koina, J. T. Lee, S. Mahony, M. A. Marra, R. D. Miller, R. D. Nicholls, M. Oda, A. T. Papenfuss, Z. E. Parra, D. D. Pollock, D. A. Ray, J. E. Schein, R. P. Speed, K. Thompson, J. L. VandeBerg, C. M. Wade, J. A. Walker, P. D. Waters, C. Webber, J. R. Weidman, X. Xie, M. C. Zody, Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team, J. A. M. Graves, C. P. Ponting, M. Breen, P. B. Samollow, E. S. Lander, K. Lindblad-Toh, Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007).
113. Y. Zhou, L. Shearwin-Whyatt, J. Li, Z. Song, T. Hayakawa, D. Stevens, J. C. Felton, E. Peel, Y. Cheng, F. Pajpach, N. Bradley, H. Suzuki, M. Nikaïdo, J. Damas, T. Daish, T. Perry, Z. Zhu, Y. Geng, A. Rhie, Y. Sims, J. Wood, B. Haase, J. Mountcastle, O. Fedrigo, Q. Li, H. Yang, J. Wang, S. D. Johnston, A. M. Phillippy, K. Howe, E. D. Jarvis, O. A. Ryder, H. Kaessmann, P. Donnelly, J. Korlach, H. A. Lewin, J. Graves, K. Belov, M. B. Renfree, F. Grutzner, Q. Zhou, G. Zhang, Platypus and echidna genomes reveal mammalian biology and evolution. *Nature* **592**, 756–762 (2021).
114. R. N. Johnson, D. O’Meally, Z. Chen, G. J. Etherington, S. Y. W. Ho, W. J. Nash, C. E. Grueber, Y. Cheng, C. M. Whittington, S. Dennison, E. Peel, W. Haerty, R. J. O’Neill, D. Colgan, T. L. Russell, D. E. Alquezar-Planas, V. Attenbrow, J. G. Bragg, P. A. Brandies, A. Y.-Y. Chong, J. E. Deakin, D. B. Palma, Z. Duda, M. D. B. Eldridge, K. M. Ewart, C. J. Hogg, G. J. Frankham, A. Georges, A. K. Gillett, M. Govendir, A. D. Greenwood, T. Hayakawa, K. M. Helgen, M. Hobbs, C. E. Holleley, T. N. Heider, E. A. Jones, A. King, D. Madden, J. A. M. Graves, K. M. Morris, L. E. Neaves, H. R. Patel, A. Polkinghorne, M. B. Renfree, C. Robin, R. Salinas, K. Tsangaras, P. D. Waters, S. A. Waters, B. Wright, M. R. Wilkins, P. Timms, K. Belov, Adaptation and conservation insights from the koala genome. *Nat. Genet.* **50**, 1102–1111 (2018).
115. D. Jebb, Z. Huang, M. Pippel, G. M. Hughes, K. Lavrichenko, P. Devanna, S. Winkler, L. S. Jermin, E. C. Skirmuntt, A. Katzourakis, L. Burkitt-Gray, D. A. Ray, K. A. M. Sullivan, J. G. Roscito, B. M. Kirilenko, L. M. Dávalos, A. P. Corthals, M. L. Power, G. Jones, R. D. Ransome, D. K. N. Dechmann, A. G. Locatelli, S. J. Puechmaille, O. Fedrigo, E. D. Jarvis, M. Hiller, S. C. Vernes, E. W. Myers, E. C. Teeling, Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).
116. P. Andrade, C. Pinho, G. Pérez I de Lanuza, S. Afonso, J. Brejcha, C.-J. Rubin, O. Wallerman, P. Pereira, S. J. Sabatino, A. Bellati, D. Pellitteri-Rosa, Z. Bosakova, I. Bunikis, M. A. Carretero, N. Feiner, P. Marsik, F. Paupério, D. Salvi, L. Soler, G. M. White, T. Uller, E. Font, L. Andersson, M. Carneiro, Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5633–5642 (2019).
117. M. R. Stammenitz, K. Gori, Y. M. Kwon, E. Harry, F. J. Martin, K. Billis, Y. Cheng, A. Baez-Ortega, W. Chow, S. Comte, H. Eggertsson, S. Fox, R. Hamed, M. Jones, B. Lazenby, S. Peck, R. Pye, M. A. Quail, K. Swift, J. Wang, J. Wood, K. Howe, M. R. Stratton, Z. Ning, E. P. Murchison, The evolution of two transmissible cancers in Tasmanian devils. *Science* **380**, 283–293 (2023).
118. A. D. Foote, Y. Liu, G. W. C. Thomas, T. Vinař, J. Alföldi, J. Deng, S. Dugan, C. E. van Elk, M. E. Hunter, V. Joshi, Z. Khan, C. Kovar, S. L. Lee, K. Lindblad-Toh, A. Mancina, R. Nielsen, X. Qin, J. Qu, B. J. Raney, N. Vijay, J. B. W. Wolf, M. W. Hahn, D. M. Muzny, K. C. Worley, M. T. P. Gilbert, R. A. Gibbs, Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **47**, 272–275 (2015).
119. Y. Fan, Z.-Y. Huang, C.-C. Cao, C.-S. Chen, Y.-X. Chen, D.-D. Fan, J. He, H.-L. Hou, L. Hu, X.-T. Hu, X.-T. Jiang, R. Lai, Y.-S. Lang, B. Liang, S.-G. Liao, D. Mu, Y.-Y. Ma, Y.-Y. Niu, X.-Q. Sun, J.-Q. Xia, J. Xiao, Z.-Q. Xiong, L. Xu, L. Yang, Y. Zhang, W. Zhao, X.-D. Zhao, Y.-T. Zheng, J.-M. Zhou, Y.-B. Zhu, G.-J. Zhang, J. Wang, Y.-G. Yao, Genome of the Chinese tree shrew. *Nat. Commun.* **4**, 1426 (2013).
120. T. Mitros, J. B. Lyons, A. M. Session, J. Jenkins, S. Shu, T. Kwon, M. Lane, C. Ng, T. C. Grammer, M. K. Khokha, J. Grimwood, J. Schmutz, R. M. Harland, D. S. Rokhsar, A chromosome-scale genome assembly and dense genetic map for *Xenopus tropicalis*. *Dev. Biol.* **452**, 8–20 (2019).

**Acknowledgments:** We thank the members of the Mallarino laboratory and the Donia laboratory for discussions and assistance with animal husbandry; N. Crossland (Boston University) for visualization and analysis of histology slides; Princeton LAR (K. Gerhart, G. Barnett, J. McGuire, D. D. L. Aler, R. Franco, P. K. Cunningham, and B. Tereszczyn) for help with animal husbandry and care; J. M. Miller and J. Arley Volmar from the LSI Genomics Core (Princeton) for help with library preparation; C. J. DeCoste from the Flow Cytometry Core (Princeton) for help with the PI assay; O. Perotti for help in the laboratory; and F. Rogers for help with statistics. Histology scanning used a Vectra Polaris slide scanner acquired through NIH S10 funding (S10OD030269). **Funding:** This project was supported by an NIH grant to R.M. (R35GM133758) and a Princeton University Dean for Research Innovation Award to R.M. R.M. is partially supported by the Searle Scholars Program, the Sloan Foundation, and the Vallee Scholars Program. M.S.D. is supported by the Gordon and Betty Moore Foundation (grants

7620 and 9199). C.Y.F. was supported by an NIH fellowship (F32 GM139240-01). Y.P. was supported by NIH/NIAID grant DP2AI171161 and by Ludwig Institute for Cancer Research. **Author contributions:** J.P., M.S.D., and R.M. conceived the project and designed experiments. J.P. and C.Y.F. conducted computational gene annotation, phylogenetic analyses, and tissue harvesting. W.K. conducted RCMC experiments. W.K., J.P., and Y.P. analyzed RCMC data. A.K. performed ELISA experiments and analyzed the resulting data. A.H.G. conducted histology. J.P. conducted all the other experiments and data analyses. J.P., M.S.D., and R.M. wrote the manuscript with the input from all authors. **Competing interests:** M.S.D. is a scientific cofounder and CSO at Pragma Bio. The authors declare that they have no other competing

interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The scRNA-seq, RNA-seq, ATAC-seq, and RCMC data are submitted under NCBI BioProject accession code PRJNA1141454.

Submitted 21 August 2024  
Accepted 12 March 2025  
Published 16 April 2025  
10.1126/sciadv.ads6359