



# A family of olfactory receptors uniquely expanded in marsupial and monotreme genomes are expressed by a T cell subset also unique to marsupials and monotremes

Jordan M. Sampson<sup>a</sup>, Kimberly A. Morrissey<sup>a</sup>, Daniel C. Douek<sup>b</sup>, Robert D. Miller<sup>a,\*</sup>

<sup>a</sup> Center for Evolutionary & Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, NM, USA

<sup>b</sup> Human Immunology Section, Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

## ARTICLE INFO

### Keywords:

OR  
Annotation  
Evolution  
Ectopic expression  
Atypical T cells  
Marsupials

## ABSTRACT

Olfactory receptors (OR), expressed on olfactory neurons, mediate the sense of smell. Recently, OR have also been shown to be expressed in non-olfactory tissues, including cells of the immune system. An analysis of single-cell transcriptomes of splenocytes of the grey short-tailed opossum (*Monodelphis domestica*) found OR are expressed on a subset of T cells, the  $\gamma\mu$  T cells, that are unique to marsupials and monotremes. A majority of opossum  $\gamma\mu$  T cells transcriptomes contain OR family 14 transcripts, specifically, from the OR14C subfamily. Amongst the mammals, the OR14 gene family is expanded in the genomes of marsupials and monotremes, and rarer or absent in placental mammals. In summary, here we demonstrate the intriguing correlation that a family of OR genes, abundant in the genomes of marsupials and monotremes, are ectopically expressed in a particular subset of T cells unique to the marsupials and monotremes.

## 1. Introduction

There are three living lineages of mammals: the so-called placentals (e.g. humans and mice), the marsupials (e.g. opossums and kangaroos), and the egg-laying monotremes (e.g. the duckbill platypus). Comparative analyses between the three mammalian lineages have revealed gains, losses, and expansions of various genetic components of different systems. For example, nearly all jawed vertebrates have two lineages of T cells defined by their T cell receptors (TCR): the  $\alpha\beta$  and  $\gamma\delta$  T cells (Rast et al., 1997; Morrissey et al., 2022). However, we previously showed that marsupials and monotremes have a third lineage, the  $\gamma\mu$  T cell, that is ancient in mammals but lost in the placental mammal lineage (Parra et al., 2007, 2008; Wang et al., 2011; Morrissey et al., 2021). Similarly, comparative genomics revealed a particular olfactory receptor (OR) family, OR14, is expanded in the genomes of the laboratory opossum (*Monodelphis domestica*) and platypus (*Ornithorhynchus anatinus*) (Warren et al., 2008; Olender et al., 2008). In contrast, the OR14 family is comparatively rare in placental mammal genomes (Warren et al., 2008; Olender et al., 2008; this paper).

Recently, ectopic OR expression has been demonstrated for a variety of non-olfactory tissues (Feldmesser et al., 2006; Kang and Koo, 2012; Chen et al., 2018). The role of OR in these non-olfactory cells and tissues

remains largely unexplored and controversial (Feldmesser et al., 2006). Of interest, OR have been found on mature lymphocytes where they appear to influence cellular chemotaxis (Malki et al., 2015; Clark et al., 2016). Using a single-cell transcriptome dataset, we present evidence of OR transcription by opossum immune cells. Noteworthy is the preferential transcription of OR14 gene family members, the OR gene family that is expanded in the genomes of marsupials and monotremes, by  $\gamma\mu$  T cells, which are also unique to marsupials and monotremes.

## 2. Materials and methods

### 2.1. Single cell transcriptome dataset analysis

A single-cell RNA sequencing (scRNASeq) dataset using opossum spleen, thymus, and blood cells was described previously (Morrissey et al., 2021; Schraven et al., 2021; accession numbers PRJNA635959 and PRJNA980639). Differential gene expression (DE) analysis was performed using the R package DESeq2 (v4.1.2; R Core Team 2021; Love et al., 2014). Only genes with >10 reads were used in the analysis. Volcano plots of differentially expressed transcripts were constructed using the R package EnhancedVolcano with an adjusted *P*-value <0.05 (v4.1.2; R Core Team 2021; Blighe et al., 2020, Fig. 1).

\* Corresponding author.

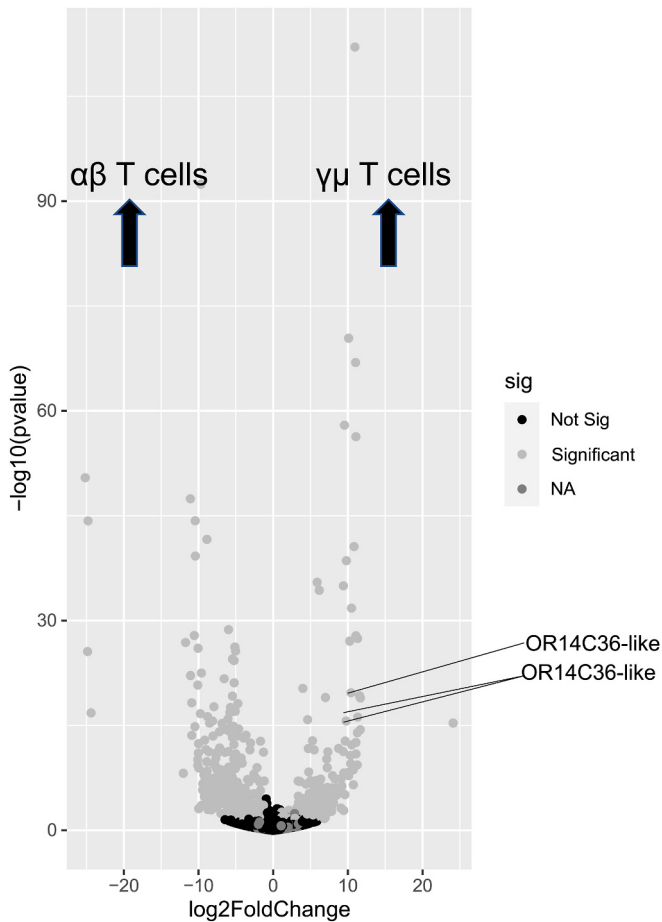
E-mail address: [rdmiller@unm.edu](mailto:rdmiller@unm.edu) (R.D. Miller).

<https://doi.org/10.1016/j.dci.2024.105149>

Received 13 November 2023; Received in revised form 7 February 2024; Accepted 7 February 2024

Available online 8 February 2024

0145-305X/© 2024 Elsevier Ltd. All rights reserved.



**Fig. 1.** Results of the DE analysis. Genes upregulated in  $\alpha\beta$  T cells are on the left, and genes upregulated in  $\gamma\mu$  T cells are on the right. The genes with significant upregulation in either T cell type are shown in grey, with not significant genes in black. All transcripts annotated as OR14C36-like are labeled.

## 2.2. Genome annotation

Current convention for the annotation of OR genes in any species is based on nucleotide identity between OR genes within and between species (Glusman et al., 2000). These criteria was used here for the reannotation of the opossum OR genes. The criteria for an OR being in a family is having  $\geq 40\%$  nucleotide identity amongst family members and  $\geq 60\%$  nucleotide identity, for subfamily members. Family numerical designations maintained based on their nucleotide identity with previously annotated human OR annotated in Ensembl and NCBI. To annotate the opossum OR, we used the nomenclature according to Glusman et al. (2000) and Olender et al. (2020). This nomenclature states that the names of the OR should start with a way to note species, the OR notation, a number for the family name, a letter for the subfamily, and a number for the gene member. For example, the first gene in the *M. domestica* OR family 14, subfamily C would be *ModoOR14C1* in the *M. domestica* genome (version MonDom5, accession number AAFR00000000.3). The OR14 loci within the opossum were re-annotated using the “Modo” designation to signify *M. domestica*, A, C, or I for the subfamily, and a number for the individual gene member (Glusman et al., 2000; Olender et al., 2020; Silva et al., 2023).

The criteria for being in a family or subfamily is having  $\geq 40\%$  or  $\geq 60\%$  nucleotide identity, respectively (Glusman et al., 2000). OR genes were numbered based on their order along the chromosome, starting with the most 5' locus. For subfamily members found on different chromosomes, the numerically lower chromosome was annotated first. For instance, the *ModoOR14A* on chromosome 2 were numbered first,

followed by the *ModoOR14A* on chromosome 4.

## 2.3. OR14 gene count

Representative OR14s from each human subfamily A, C, I, J, K, and L, available from the HORDE database (Olender et al., 2013) were used to search the genomes of 20 mammalian species in NCBI using BLASTn (Altschul et al., 1990, Table 1). The genomes searched were platypus (*Ornithorhynchus anatinus*, accession number PRJNA534073), echidna (*Tachyglossus aculeatus*, accession number JADRJE000000000.1), opossum (*M. domestica*, accession number AAFR000000000.3), Tasmanian devil (*Sarcophilus harrisii*, accession number CACPPN000000000.1), wombat (*Vombatus ursinus*, accession number UNPS000000000.2), koala (*Phascolarctos cinereus*, accession number MSTS000000000.1), human (*Homo sapiens*, accession number GCA\_000001405.28), big brown bat (*Eptesicus fuscus*, accession number ALEH000000000.1), black flying fox (*Pteropus alecto*, accession number ALWS000000000.1), cat (*Felis catus*, accession number AANG000000000.4), cow (*Bos taurus*, accession number NKLS000000000.2), dog (*Canis lupus familiaris*, accession number JAAQRD000000000.1), Egyptian fruit bat (*Rousettus aegyptiacus*, accession number GCA\_001466805.2), giant panda (*Ailuropoda melanoleuca*, accession number LNAT000000000.2), greater horseshoe bat (*Rhinolophus ferrumequinum*, accession number RXPC000000000.1), hyrax (*Procavia capensis*, accession number PVIO000000000.2), large

**Table 1**

A total of all OR14s found within the genomes of several placental, marsupial and monotreme species.

| Species                      | OR14              |                   |    |   |   |    | Total OR14s | Total ORs           |
|------------------------------|-------------------|-------------------|----|---|---|----|-------------|---------------------|
|                              | A                 | C                 | I  | J | K | L  |             |                     |
| Platypus <sup>a</sup>        | 34                | 0                 | 0  | 0 | 1 | 2  | 37          | 299 <sup>b</sup>    |
| Echidna <sup>a</sup>         | 40                | 0                 | 6  | 1 | 7 | 10 | 64          | 693 <sup>b</sup>    |
| Opossum <sup>b</sup>         | 15                | 23                | 17 | 0 | 0 | 0  | 55          | 1327                |
| Tasmanian Devil <sup>b</sup> | 9                 | 13                | 12 | 0 | 0 | 0  | 34          | 1090                |
| Common Wombat <sup>b</sup>   | 18                | 8                 | 8  | 0 | 0 | 0  | 34          | n.d.                |
| Koala <sup>b</sup>           | 10                | 2                 | 1  | 0 | 0 | 0  | 13          | n.d.                |
| Human                        | 2                 | 1                 | 1  | 1 | 1 | 1  | 7           | 416                 |
| Orangutan                    | 2                 | 1                 | 2  | 1 | 1 | 0  | 7           | (438 <sup>ψ</sup> ) |
| Rhesus                       | 2                 | 0                 | 2  | 1 | 1 | 0  | 6           | 931 <sup>d</sup>    |
| Macaque                      |                   |                   |    |   |   |    |             | 733 <sup>d</sup>    |
| Mouse <sup>c</sup>           | 4                 | 1                 | 0  | 2 | 0 | 0  | 9           | 1,264 <sup>d</sup>  |
|                              | (1 <sup>ψ</sup> ) | (1 <sup>ψ</sup> ) |    |   |   |    |             |                     |
| Big Brown Bat                | 0                 | 0                 | 0  | 0 | 0 | 0  | 0           | n.d.                |
| Black Flying Fox             | 0                 | 0                 | 0  | 0 | 0 | 0  | 0           | n.d.                |
| Egyptian Fruit Bat           | 0                 | 0                 | 0  | 0 | 0 | 0  | 0           | n.d.                |
| Greater Horseshoe Bat        | 0                 | 0                 | 0  | 0 | 0 | 0  | 0           | n.d.                |
| Large Flying Fox             | 0                 | 0                 | 0  | 0 | 0 | 0  | 0           | 672 <sup>d</sup>    |
| Little Brown Bat             | 0                 | 0                 | 0  | 0 | 0 | 0  | 0           | 659 <sup>d</sup>    |
| Vesper bat                   | 0                 | 0                 | 0  | 0 | 0 | 0  | 0           | n.d.                |
| Cat                          | 5                 | 1                 | 0  | 1 | 2 | 0  | 9           | 850 <sup>d</sup>    |
| Dog                          | 0                 | 1                 | 1  | 1 | 0 | 0  | 3           | 1,044 <sup>d</sup>  |
| Cow                          | 0                 | 0                 | 0  | 2 | 0 | 0  | 2           | 2,381 <sup>d</sup>  |
| Giant Panda                  | 1                 | 0                 | 1  | 1 | 1 | 0  | 4           | n.d.                |
| Hyrax                        | 0                 | 0                 | 0  | 1 | 0 | 0  | 1           | 678 <sup>d</sup>    |

OR14 counts taken from NCBI.

The marsupials are highlighted in orange.

<sup>a</sup> The duckbill platypus and the echidna are highlighted in yellow.

<sup>b</sup> Total counts taken from Zhou et al., 2021).

<sup>d</sup> Total counts taken from Hayden et al. (2010).

<sup>e</sup> Mouse equivalent determined by nucleotide percent identity.

lying fox (*Pteropus vampyrus*, accession number ABRP000000000.2), little brown bat (*Myotis lucifugus*, accession number AAPE000000000.2), mouse (*Mus musculus*, accession number GCA\_000001635.9), Vesper bat (*Myotis davidii*, accession number ALWT000000000.1), orangutan (*Pongo abelii*, accession number NDHI000000000.3), and the Rhesus macaque (*Macaca mulatta*, accession number QNVO000000000.2).

Total OR gene counts, in addition to OR14 counts for human, opossum, and Tasmanian devil, were obtained using the genomes available in NCBI (Supplementary Table 1). For all other species, the total count was taken from published data (Table 1).

## 2.4. Phylogenetic analysis

A phylogenetic tree was constructed using nucleotide alignments using Clustal W and performed using MEGAX, iTOL, and the neighbor-joining method (Thompson et al., 1994; Tamura et al., 2004; Kumar et al., 2018; Letunic and Bork, 2021). Bootstrap values were processed using the maximum composite likelihood method (Felsenstein, 1981). The OR14 used include all seven human OR14, and representative OR14 from opossum, Tasmanian devil, and platypus. Representative OR from families 1, 2, 3, 4, 5, 6, 7, 8, 9, 51, and 52 were also used from human, opossum, Tasmanian devil, and platypus. GenBank accession numbers of the OR sequences used in the phylogeny are found in Supplementary Table 2.

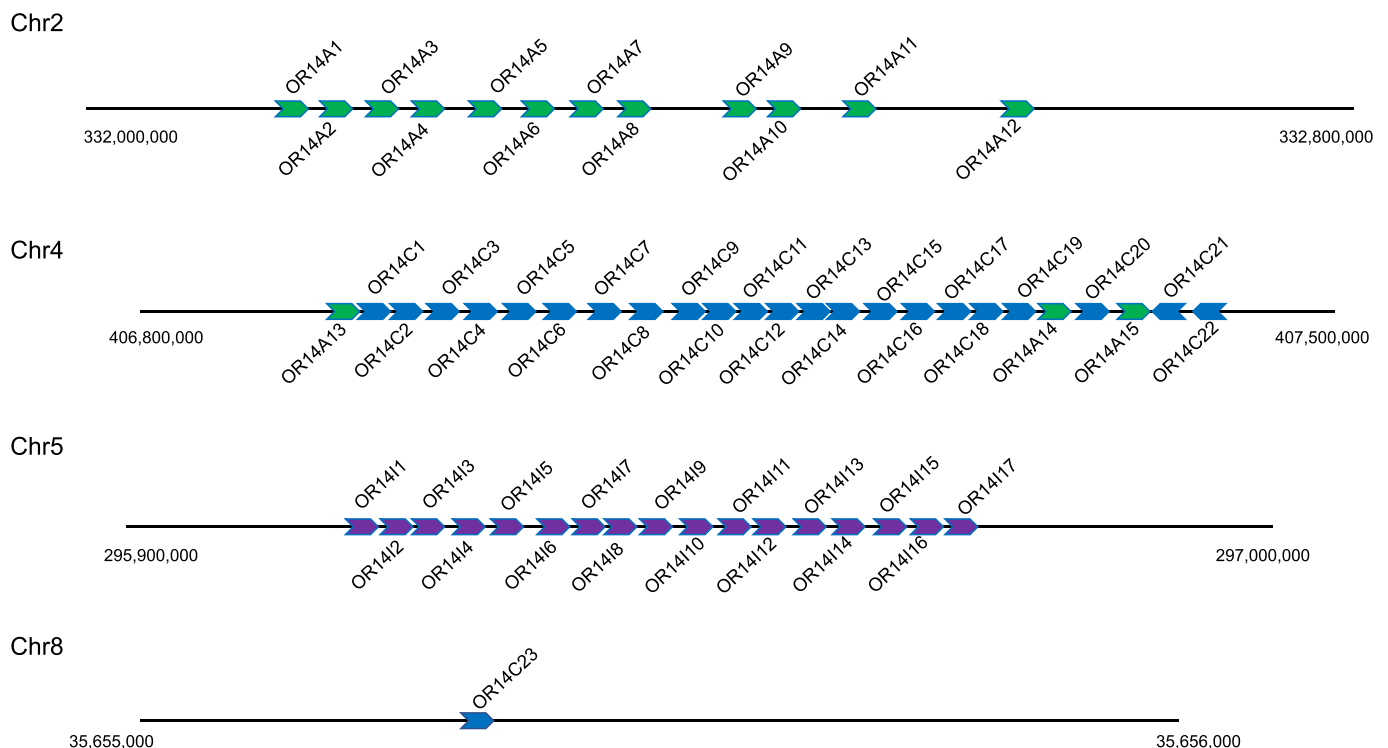
## 3. Results

$\gamma\mu$  T cells are a subset of T cells unique to marsupial and monotreme mammals (Hansen and Miller, 2015). DE analysis of scRNASeq datasets revealed an overabundance of OR transcripts in  $\gamma\mu$  T cells relative to  $\alpha\beta$  T cells (Fig. 1). Specifically, transcripts encoding three unique OR are abundant in the  $\gamma\mu$  T cell transcriptomes (labeled in Fig. 1). These are encoded by genes previously annotated as OR14C36-like in the opossum

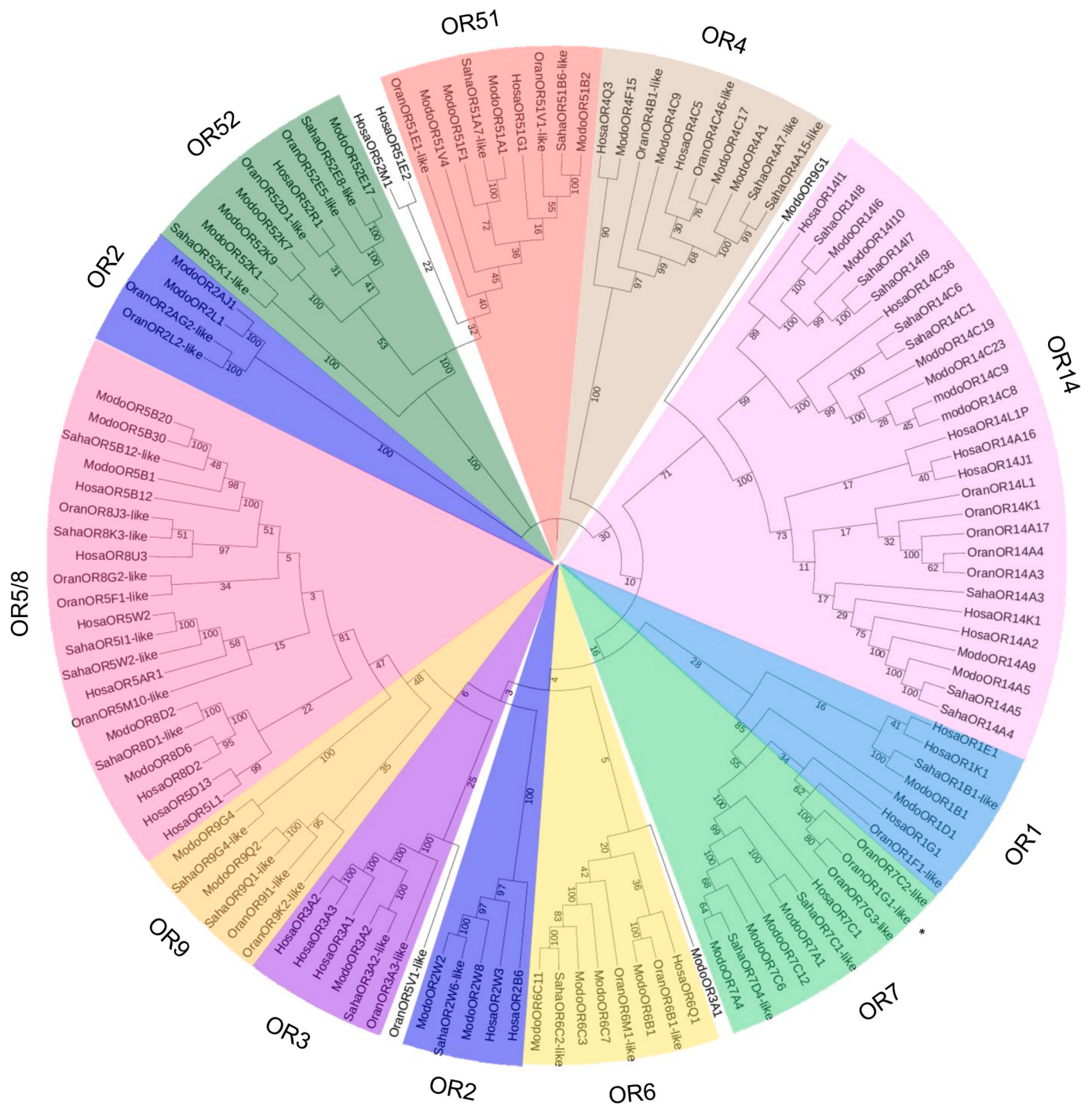
due to their relatedness to the human OR14C subfamily. Multiple members of the same OR families and subfamilies in species outside of humans and mice were often annotated with identical names based on similarity to the single OR14 family member in the human genome. To investigate specific transcription of OR genes in opossum immune cells, it was necessary to first annotate the OR genes in the opossum genome, which had not been done for any marsupial species. Since the OR transcripts that are upregulated in  $\gamma\mu$  T cells appeared to be related to the OR14 family, which had previously been shown to be expanded in marsupials, we focused specifically on this family for further analyses (Warren et al., 2008; Olender et al., 2008).

There are 55 OR14 family gene copies in the opossum genome, and they are found on chromosomes (chr) 2, 4, 5, and 8 (Fig. 2, Supplementary Table 1). *ModoOR14s* distinctly cluster with the other OR14s from placentals and monotremes when compared to other OR families (Fig. 3). Following this reannotation of the *ModoOR14*, the expansion of the OR14 loci in the genomes of marsupials and monotremes, relative to placental mammals, was confirmed (Table 1; Fig. 3; Warren et al., 2008; Olender et al., 2008). Indeed, the most species rich placental mammalian lineage, the bats, lacked the OR14 family altogether (Table 1). All the OR14 subfamilies that have been annotated in placental mammals (A, C, I, J, K, and L) are also represented in marsupials and monotremes collectively (Table 1). However, marsupials only have subfamilies A, C, and I, whereas monotremes have A, I, J, K, and L (Table 1). Amongst the marsupials, the opossum has the greatest number of OR14 genes, and amongst all mammals the echidna has the greatest number, so far.

The annotation of the *ModoOR14* family members allowed for the identification of those expressed in each cell type of the single-cell datasets, which included splenocytes, peripheral blood mononuclear cells (PBMCs), and thymocytes (Table 2). Since there were 23 *ModoOR14C* genes identified in the genome, each one was used to search the single-cell transcriptomes to identify the individual *ModoOR14C* genes present and the cell type transcribing them (Fig. 2). *ModoOR14s* were



**Fig. 2.** Gene maps of all 55 *M. domestica* *ModoOR14* genes. *ModoOR14A1-ModoOR14A12* are found on chromosome 2 and *ModoOR14A13-ModoOR14A15* are found on chromosome 4 in green. *ModoOR14C1-ModoOR14C22* are found on chromosome 4 and *ModoOR14C23* is found on chromosome 8 in blue. *ModoOR14I1-ModoOR14I17* are found on chromosome 5 in purple. *ModoOR14s* were organized into family or subfamily based on having  $\geq 40\%$  or  $\geq 60\%$  nucleotide identity, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 3.** Olfactory receptor (OR) phylogeny. Representative OR from platypus (Oran), human (Hosa), Tasmanian devil (Saha), and opossum (ModO) were compared via a phylogenetic tree. Phylogenetic tree was constructed using nucleotide alignment using Clustal W of above species OR. Analysis was constructed using neighbor-joining method. Numbers on branches indicate bootstrap values on 1000 replicates and were processed using the maximum composite likelihood method. OR families were separated by color with OR1 in light blue, OR2 in dark blue, OR3 in dark purple, OR4 in beige, OR5/8 in pink, OR6 in yellow, OR7 in light green, OR9 in orange, OR14 in light purple, OR51 in red, and OR52 in dark green. OR that did not follow familial phylogeny were kept white. Asterisk represents a possible annotation error in OranOR1G1-like which may be an OR7. GenBank accession numbers of the OR sequences used in the tree are found in [Supplementary Table 2](#). Opossum genes included in the analysis represent the breadth of diversity within each OR subfamily. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

found to be transcribed in 61% of  $\gamma\mu$  T cells in the spleen (Table 2, Table 3). In comparison, *ModoOR14* transcripts were found in only ~2.7% of  $\alpha\beta$  T cells and 12.5% splenic  $\gamma\delta$  T cells (Table 2, Table 3). The *ModoOR14* were transcribed in other cell types both in the spleen, blood, and thymus, though to a lesser extent than splenic  $\gamma\mu$  T cells (Table 2, Table 3). There was ectopic transcription of other OR loci but not to the

same extent as the *ModoOR14* by splenic  $\gamma\mu$  T cells (Table 2, Supplementary Table 3).

The predominant *ModoOR14* subfamily transcripts found in  $\gamma\mu$  T cells were of the *ModoOR14C* subfamily (Table 3). While most of the  $\gamma\mu$  T cells tended to transcribe only a single subfamily member, some  $\gamma\mu$  T cells expressed more than one *ModoOR14C*; although, in some instances, the

**Table 2**  
The scRNAseq tissue distribution of *ModoOR14s* compared to other ORs.

| Tissue            | Cell Type    |               |                      |                       |                    |               | Total     |
|-------------------|--------------|---------------|----------------------|-----------------------|--------------------|---------------|-----------|
|                   | B cell       | T Total       | $\alpha\beta$ T cell | $\gamma\delta$ T cell | $\gamma\mu$ T cell | Other         |           |
| Blood             | 0/97         | 2/175         | 2/165 (1.2%)         | 0/10                  | 0/0                | 3/209 (1.4%)  | 5/480     |
| Spleen            | 1/142 (0.7%) | 33/132        | 2/75 (2.7%)          | 1/8 (13%)             | 30/49 (61%)        | 4/192 (2.1%)  | 38/466    |
| Thymus            | 0/0          | 1/162         | 1/162 (0.62%)        | 0/0                   | 0/0                | 1/302 (0.33%) | 2/464     |
| Total OR14        | 1/239        | 36/469        | 5/402                | 1/18                  | 30/49              | 8/703         | 45/1,410  |
| Total NonOR14 ORs | 7/239 (2.9%) | 29/469 (6.2%) | 19/402 (4.7%)        | 3/18 (17%)            | 7/49 (14%)         | 79/703 (11%)  | 144/1,410 |

**Table 3**  
Number of  $\gamma\mu$  T cells,  $\alpha\beta$  T cells  $\gamma\delta$  T cells, B cells, and other cells *ModoOR14s* were found on.

| <i>ModoOR14</i> Subfamily Member | Number of $\gamma\mu$ T cells | Number of $\alpha\beta$ T cells | Number of $\gamma\delta$ T cells | Number of B cells | Number of other cells |
|----------------------------------|-------------------------------|---------------------------------|----------------------------------|-------------------|-----------------------|
| C8                               | 1                             | 0                               | 0                                | 0                 | 0                     |
| C9                               | 19                            | 1                               | 0                                | 0                 | 0                     |
| C8/C9                            | 2                             | 0                               | 0                                | 0                 | 0                     |
| C19                              | 0                             | 0                               | 0                                | 1                 | 0                     |
| <i>ModoOR14</i> Unk              | 8                             | 1                               | 0                                | 0                 | 0                     |
| No <i>ModoOR14</i>               | 19                            | 400                             | 18                               | 238               | 703                   |
| Total                            | 49                            | 402                             | 18                               | 239               | 703                   |

*ModoOR14C* reads were too short to assign them to a subfamily member (Table 3). There are two other subfamilies found in the opossum, *ModoOR14A* and *ModoOR14I*, however, neither was found transcribed in splenocytes, peripheral blood cells, or thymocytes (Table 2, Table 3).

4. Discussion

In addition to the conventional  $\alpha\beta$  and  $\gamma\delta$  T cell lineages found in nearly all jawed vertebrates, marsupials and monotreme mammals have a third lineage which are the  $\gamma\mu$  T cells (Rast et al., 1997; Parra et al., 2007, 2008; Wang et al., 2011; Morrissey et al., 2021). This lineage of T cells was likely found in the last common ancestor of mammals but lost in the placental lineage (Parra et al., 2012a). The TCR $\mu$  chain is distinct by having a third extracellular V domain which is more related to antibody V domains than to the V domains of the conventional TCR chains (Parra et al., 2007, 2012a; Wang et al., 2011; Morrissey et al., 2021). The role  $\gamma\mu$  T cells play in the immune system of non-placental mammals is not known, although single cell RNA sequencing is beginning to reveal additional characteristics of this lineage. One of the more curious findings, described herein, was increased transcription of OR genes relative to the conventional T cells.

OR are a subset of G-protein coupled receptors (GPCR) discovered by their expression by olfactory neurons (Buck and Axel, 1991). Each neuron expresses a single OR in a clonal fashion (Buck and Axel, 1991). When ligands, such as small odorant molecules, bind they signal the olfactory bulb in the brain, triggering the sense of smell (Buck and Axel, 1991). OR are encoded by large, multi-gene families that have evolved by the process of duplication and deletion, also known as the gene birth-and-death model, like many other large multi-gene families (Nii-mura & Nei, 2005, 2007). This has resulted in OR genes demonstrating varying degrees of relatedness that allows them to be grouped into families and subfamilies, often with *trans*-specific relationships (Glus-man et al., 2000; Olender et al., 2020). For example, there are ~1000 total OR genes in the human genome, and these can be grouped into 18 families based on sequence relatedness, and all 18 of these families can be identified in gorillas, and all but one can be identified in cows (Buck and Axel, 1991; Hayden et al., 2010).

Ectopic transcription of OR loci has been found in cells of the

immune system, such as T cells, although their function on immune cells is a matter of debate. One school of thought is that OR ectopic expression is selectively neutral, perhaps even random, and due to leaky trans-cription in areas of open chromatin (Feldmesser et al., 2006). This seems unlikely in the case of OR14 in the opossum, however, given the apparent targeted transcription of this family by  $\gamma\mu$  T cells. Furthermore, OR have been shown to mediate chemotaxis of T cells in other species (Clark et al., 2016). This may not be surprising given that chemokine receptors, that bind endogenous ligands, are also GPCR (Nomi-yama et al., 2011). Attraction to odorants from microbes may play a role in immune cell chemotaxis via exogenous signals. Indeed, in some cases, the stimulation by the OR ligand was able to override chemokine signals (Clark et al., 2016).

Analyses of the OR transcripts found in  $\gamma\mu$  T cells revealed they were related to the OR14 family, which was discovered with the annotation of the platypus genome (Warren et al., 2008). OR14 were originally clas-sified as being part of the human OR5 family, however, they were later recognized as being a distinct family (Warren et al., 2008; Olender et al., 2008, 2013; Zhou et al., 2021). The OR14 family is also present in marsupial, placental mammal, and avian genomes and is, therefore, ancient in amniotes and was present prior to the divergence of the three mammalian lineages starting ~180 million years ago (Olender et al., 2008, 2013; Khan et al., 2015; Bininda-Emonds et al., 2007).

Each of the three extant mammalian lineages appear to have fol-lowed different genomic trajectories with regards to the complexity and diversity of OR14 genes. Monotremes have a large complexity of OR14 genes and an increased number of OR14 gene subfamilies. Marsupials have a comparatively greater complexity of OR14 genes compared to both monotremes and placentals but fewer subfamilies overall. Placental mammals have substantially fewer OR14 genes but have retained more subfamilies, collectively.

The difference in OR14 gene number between placental mammals on one hand, and marsupials and monotremes on the other, begs the question of the genomic content of the last common ancestor. Insights drawn from avian genomes is consistent with a larger complexity of OR14 genes being the ancestral state in amniotes. For example, 428 of 674 (64%) of chicken OR genes are OR14 and 552 of 688 (80%) of zebra finch OR genes are OR14 family (Khan et al., 2015). These observations are consistent with a high complexity of OR14 genes in the genomes of early amniotes being the norm and likely present in the last common mammalian ancestor. This complexity was retained in both monotremes and marsupials but lost in the placental lineage, following the latter's divergence from the marsupials ~160 million years ago (Table 1; Bininda-Emonds et al., 2007).

The expansion and contraction of OR gene complexity and diversity are thought to reflect different ecological adaptations over millions of years (Hayden et al., 2010). It is possible such adaptation has influenced why marsupials and monotremes have retained larger numbers of the OR14 family while being lost in the placental mammals. However, during the evolution of vertebrate immune systems there are many intriguing correlations in concerted gene loss. Some cases appear more easily explained. For example, cod and other Gadiform fishes have lost both their Major Histocompatibility Class II (MHCII) and CD4 genes (Star et al., 2011). The loss of one likely led to the loss of the other given

the direct interaction between CD4 with MHCII proteins in vertebrate immune systems. As stated earlier, the  $\gamma\mu$  T cell lineage is uniquely mammalian, was present in the last common ancestor of mammals, was retained by marsupials and monotremes, but was lost in the placental mammals (Parra et al., 2007; Wang et al., 2011; Morrissey et al., 2021). In other words, the mammalian lineages that retained  $\gamma\mu$  T cells also retained larger numbers of OR14 genes in their genomes. This may be coincidental, however,  $\gamma\mu$  T cells are also the primary immune cells ectopically expressing OR14 genes consistent with a possible relationship.

Birds, which have large numbers of OR14 genes, lack  $\gamma\mu$  T cells, however they do have T cells with an atypical TCR. Both chickens and zebra finches have  $\gamma\delta$  T cells that utilize TCR $\delta$  chains containing V domains more related to antibody heavy chain V than conventional V $\delta$ , called VH $\delta$  (Parra et al., 2012b). VH $\delta$  introgression in the TCR $\delta$  locus likely represents an ancestral step in the evolution of TCR $\mu$  (Parra et al., 2012a). Taken all together, there is an intriguing correlation between the presence of  $\gamma\mu$  T cells, and possibly other atypical T cell lineages, and the presence of the expanded OR14 family, a correlation that deserves further exploration of the role OR play in the biology of T cells.

## Funding

This work was supported by the National Science Foundation Award [IOS-2103367].

## CRediT authorship contribution statement

**Jordan M. Sampson:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Kimberly A. Morrissey:** Formal analysis, Conceptualization. **Daniel C. Douek:** Resources. **Robert D. Miller:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

## Data availability

All data are in public databases specified in the manuscript

## Acknowledgments

We thank Dr. Lijing Bu for his assistance and bioinformatics support. We thank the UNM Center for Advanced Research Computing for providing the high-performance computing and largescale storage resources used in this work.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dci.2024.105149>.

## References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Bininda-Emonds, O.R., Cardillo, M., Jones, K.E., MacPhee, R.D., Beck, R.M., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., Purvis, A., 2007. The delayed rise of present-day mammals. *Nature* 446, 507–512. <https://doi.org/10.1038/nature05634>.
- Blighe, K., Rana, S., Lewis, M., 2020. EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. R package version 1.6.0. <https://github.com/kevinblighe/EnhancedVolcano>.
- Buck, L., Axel, R., 1991. A Novel Multigene Family May Encode Odorant Receptors: A. Chen, Z., Zhao, H., Fu, N., Chen, L., 2018. The diversified function and potential therapy of ectopic olfactory receptors in non-olfactory tissues. *J. Cell. Physiol.* 233, 2104–2115. <https://doi.org/10.1002/jcp.25929>.
- Clark, A.A., Nurmukhambetova, S., Li, X., Munger, S.D., Lees, J.R., 2016. Odorants specifically modulate chemotaxis and tissue retention of CD4+ T cells via cyclic adenosine monophosphate induction. *J. Leukoc. Biol.* 100, 699–709. <https://doi.org/10.1189/jlb.1A0914-425RR>.
- Feldmesser, E., Olender, T., Khen, M., Yanai, I., Ophir, R., Lancet, D., 2006. Widespread ectopic expression of olfactory receptor genes. *BMC Genom.* 7, 1–18. <https://doi.org/10.1186/1471-2164-7-121>.
- Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376. <https://doi.org/10.1007/BF01734359>.
- Glusman, G., Bahar, A., Sharon, D., Pilpel, Y., White, J., Lancet, D., 2000. The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mamm. Genome* 11, 1016–1023. <https://doi.org/10.1007/s003350010196>.
- Hansen, V.L., Miller, R.D., 2015. The evolution and structure of atypical T cell receptors. In: Hsu, E., Du, Pasquier (Eds.), *Pathogen-Host Interactions: Antigenic Variation vs. Somatic Adaptation*, vol. 57. Springer, Cham, pp. 265–278. [https://doi.org/10.1007/978-3-319-20819-0\\_11](https://doi.org/10.1007/978-3-319-20819-0_11).
- Hayden, S., Bekaert, M., Crider, T.A., Mariani, S., Murphy, W.J., Teeling, E.C., 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Res.* 20, 1–9. <https://doi.org/10.1101/gr.099416.109>.
- Kang, N., Koo, J., 2012. Olfactory receptors in non-chemosensory tissues. *BMB Reports* 45, 612–622. <https://doi.org/10.5483/bmbrep.2012.45.11.232>.
- Khan, I., Yang, Z., Maldonado, E., Li, C., Zhang, G., Gilbert, M.T.P., Jarvis, E.D., O'Brien, S.J., Johnson, W.E., Antunes, A., 2015. Olfactory receptor subgenomes linked with broad ecological adaptations in Sauropsida. *Mol. Biol. Evol.* 32, 2832–2843. <https://doi.org/10.1093/molbev/msv155>.
- Kumar, S., Stecher, G., Li, M., Nkayaz, C., Tamura, K., 2018. Mega X: molecular evolutionary genetics analysis cross computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- Letunic, I., Bork, P., 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. <https://doi.org/10.1093/nar/gkab301>.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Malki, A., Fiedler, J., Fricke, K., Ballweg, I., Pfaffl, M.W., Krautwurst, D., 2015. Class I odorant receptors, TAS1R and TAS2R taste receptors, are markers for subpopulations of circulating leukocytes. *J. Leukoc. Biol.* 97, 533–545. <https://doi.org/10.1189/jlb.2a0714-331rr>.
- Morrissey, K.A., Wegrecki, M., Praveena, T., Hansen, V.L., Bu, L., Sivaraman, K.K., Darko, S., Douek, D.C., Rossjohn, J., Miller, R.D., LeNours, J., 2021. The molecular assembly of the marsupial  $\gamma\mu$  T cell receptor defines a third T cell lineage. *Science* 371, 1383–1388. <https://doi.org/10.1126/science.abc7070>.
- Morrissey, K.A., Sampson, J.M., Rivera, M., Bu, L., Hansen, V.L., Gemmell, N.J., Gardner, M.G., Bertozzi, T., Miller, R.D., 2022. Comparison of reptilian genomes reveals deletions associated with the natural loss of  $\gamma\delta$  T cells in squamates. *J. Immunol.* 208, 1960–1967. <https://doi.org/10.4049/jimmunol.2101158>.
- Niimura, Y., Nei, M., 2005. Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene* 346, 13–21. <https://doi.org/10.1016/j.gene.2004.09.025>.
- Niimura, Y., Nei, M., 2007. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS One* 2, e708. <https://doi.org/10.1371/journal.pone.0000708>.
- Nomiyama, H., Osada, N., Yoshie, O., 2011. A family tree of vertebrate chemokine receptors for a unified nomenclature. *Dev. Comp. Immunol.* 35, 705–715. <https://doi.org/10.1016/j.dci.2011.01.019>.
- Olender, T., Lancet, D., Nebert, D.W., 2008. Update on the olfactory receptor (OR) gene superfamily. *Hum. Genom.* 3, 1–11. <https://doi.org/10.1186/1479-7364-3-1-87>.
- Olender, T., Nativ, N., Lancet, D., 2013. HORDE: comprehensive resource for olfactory receptor genomics. *Olfactory Receptors: Methods and Protocols* 23–38. [https://doi.org/10.1007/978-1-62703-377-0\\_2](https://doi.org/10.1007/978-1-62703-377-0_2).
- Olender, T., Jones, T., Bruford, E., Lancet, D., 2020. A unified nomenclature for vertebrate olfactory receptors. *BMC Evol. Biol.* 20, 1–12. <https://doi.org/10.1186/s12862-020-01607-6>.
- Parra, Z.E., Baker, M.L., Schwarz, R.S., Deakin, J.E., Lindblad-Toh, K., Miller, R.D., 2007. A unique T cell receptor discovered in marsupials. *Proc. Natl. Acad. Sci. USA* 104, 9776–9781. <https://doi.org/10.1073/pnas.0609106104>.
- Parra, Z.E., Baker, M.L., Hathaway, J., Lopez, A.M., Trujillo, J., Sharp, A., Miller, R.D., 2008. Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica*. *BMC Genom.* 9, 1–9. <https://doi.org/10.1186/1471-2164-9-111>.
- Parra, Z.E., Lillie, M., Miller, R.D., 2012a. A model for the evolution of the mammalian t-cell receptor  $\alpha/\delta$  and  $\mu$  loci based on evidence from the duckbill Platypus. *Mol. Biol. Evol.* 29, 3205–3214. <https://doi.org/10.1093/molbev/mss128>.
- Parra, Z.E., Mitchell, K., Dalloul, R.A., Miller, R.D., 2012b. A second TCR $\delta$  locus in Galliformes uses antibody-like V domains: insight into the evolution of TCR $\delta$  and TCR $\mu$  genes in tetrapods. *J. Immunol.* 188, 3912–3919. <https://doi.org/10.4049/jimmunol.1103521>.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Rast, J.P., Anderson, M.K., Strong, S.J., Luer, C., Litman, R.T., Litman, G.W., 1997.  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  T cell antigen receptor genes arose early in vertebrate phylogeny. *Immunity* 6, 1–11. [https://doi.org/10.1016/S1074-7613\(00\)80237-X](https://doi.org/10.1016/S1074-7613(00)80237-X).
- Schraven, A.L., Hansen, V.L., Morrissey, K.A., Stannard, H.J., Ong, O.T., Douek, D.C., Miller, R.D., Old, J.M., 2021. Single-cell transcriptome analysis of the B-cell repertoire reveals the usage of immunoglobulins in the gray short-tailed opossum (*Monodelphis domestica*). *Dev. Comp. Immunol.* 123, 104141. <https://doi.org/10.1016/j.dci.2021.104141>.
- Silva, L., Mendes, T., Ramos, L., Zhang, G., Antunes, A., 2023. Parallel evolution of fish bi-modal breathing and expansion of olfactory receptor (OR) genes: toward a

- universal ORs nomenclature. *Journal of Genetics and Genomics* 8, 600–610. <https://doi.org/10.1016/j.jgg.2023.03.006>.
- Star, B., Nederbragt, A.J., Jentoft, S., Grimholt, U., Malmstrom, M., Gregers, T.F., Rounge, T.B., Paulsen, J., Solbakken, M.H., Sharma, A., Wetten, O.F., Lanzen, A., Winer, R., Knight, J., Vogel, J.H., Aken, B., Andersen, O., Lagesen, K., Tooming-Klunderud, A., Edvardsen, R.B., Tina, K.G., Espelund, M., Nepal, C., Previti, C., Karlsen, B.O., Moum, T., Skage, M., Berg, P.R., Gjoen, T., Kuhl, H., Thorsen, J., Malde, K., Reinhardt, R., Du, L., Johansen, S.D., Searle, S., Lien, S., Nilsen, F., Jonassen, I., Omholt, S.W., Stenseth, N.C., Jakobsen, K.S., 2011. The genome sequence of Atlantic cod reveals a unique immune system. *Nature* 477, 207–210. <https://doi.org/10.1038/nature10342>.
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA* 101, 11030–11035. <https://doi.org/10.1073/pnas.0404206101>.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>.
- Wang, X., Parra, Z.E., Miller, R.D., 2011. Platypus TCR $\mu$  provides insight into the origins and evolution of a uniquely mammalian TCR locus. *J. Immunol.* 187, 5246–5254. <https://doi.org/10.4049/jimmunol.1101113>.
- Warren, W.C., Hillier, L.W., Graves, J.A.M., Birney, E., Ponting, C.P., Grutzner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T., Yang, S.P., Heger, A., Locke, D.P., Miethke, P., Waters, P.D., Veyrunes, F., Fulton, L., Fulton, B., Graves, T., Wallis, J., Puente, X.S., Lopez-Otin, C., Ordóñez, G.R., Eichler, E.E., Chen, L., Cheng, Z., Deakin, J.E., Alsop, A., Thompson, K., Kirby, P., Papenfuss, A.T., Wakefield, M.J., Olender, T., Lancet, D., Huttley, G.A., Smit, A.F.A., Pask, A., Temple-Smith, P., Batzer, M.A., Walker, J.A., Konkel, M.K., Harris, R.S., Whittington, C.M., Wong, E.S.W., Gemmell, N.J., Buschiazio, E., Jentzsch, I.M.V., Merkel, A., Schmitz, J., Zemmann, A., Churakov, G., Kriegs, J.O., Brosius, J., Murchison, E.P., Sachidanandam, R., Smith, C., Hannon, G.J., Tsend-Ayush, E., McMillan, D., Attenborough, R., Rens, W., Ferguson-Smith, M., Lefevre, C.M., Sharp, J.A., Nicholas, K.R., Ray, D.A., Kube, M., Reinhardt, R., Pringle, T.H., Taylor, J., Jones, R. C., Nixon, B., Dacheux, J.L., Niwa, H., Sekita, Y., Huang, X.Q., Stark, A., Kheradpour, P., Kellis, M., Flicek, P., Chen, Y., Webber, C., Hardison, R., Nelson, J., Hallsworth-Pepin, K., Delehaunty, K., Markovic, C., Minx, P., Feng, Y.C., Kremitzki, C., Mitreva, M., Glasscock, J., Wylie, T., Wohldmann, P., Thiru, P., Nhan, M.N., Pohl, C.S., Smith, S.M., Hou, S.F., Renfree, M.B., Mardis, E.R., Wilson, R. K., 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453, 175–183. <https://doi.org/10.1038/nature06936>.
- Zhou, Y., Shearwin-Whyatt, L., Li, J., Song, Z., Hayakawa, T., Stevens, D., Fenelon, J.C., Peel, E., Cheng, Y., Pajpach, F., Bradley, N., Suzuki, H., Nikaido, M., Damas, J., Daish, T., Perry, T., Zhu, Z., Geng, Y., Rhie, A., Sims, Y., Wood, J., Fedrigo, O., Li, Q., Yang, H., Wang, J., Johnston, S.D., Phillippy, A.M., Howe, K., Jarvis, E.D., Ryder, O. A., Kaessmann, H., Donnelly, P., Korlach, J., Lewin, H.A., Graves, J., Belov, K., Renfree, M.B., Grutzner, F., Zhou, Q., Zhang, G., 2021. Platypus and echidna genomes reveal mammalian biology and evolution. *Nature* 592, 756–762. <https://doi.org/10.1038/s41586-020-03039-0>.
- Molecular Basis for Odor Recognition. *Cell* 65, 175–187. [https://dx.doi.org/10.1016/0092-8674\(91\)90418-x](https://dx.doi.org/10.1016/0092-8674(91)90418-x).