Article

Transcriptional interferences ensure one olfactory receptor per ant neuron

https://doi.org/10.1038/s41586-025-09664-x

Received: 8 December 2024

Accepted: 22 September 2025

Published online: 22 October 2025

Check for updates

Bogdan Sieriebriennikov[1,2], Olena Kolumba[1,3], Aurore de Beaurepaire[1], Jennifer Wu[1], Valentina Fambri[1,3], Eva Bardol[1], Yuwei Zhong[1], Ildar Gainetdinov[1], Danny Reinberg[4]✉, Hua Yan[2]✉ & Claude Desplan[1,3]✉

To ensure specificity, sensory neurons must select and express a single receptor from often vast gene families, adhering to the rule of 'one receptor per neuron'. For example, each olfactory sensory neuron in mammals expresses only one odorant receptor (Or) gene[1,2]. In *Drosophila*, which has about 60 Or genes, this selection is deterministic[3]. By contrast, mice face the challenge of choosing one Or gene from over 1,000 options[4]. They solve this through a complex system of stochastic choices[5–9]. Ants also possess many Or genes, most of which are organized into tandem arrays similar to those in mammals, but their regulatory mechanisms have evolved independently. Here we show that, in the ant *Harpegnathos saltator*, each olfactory sensory neuron activates a single promoter within an Or gene array, producing a mature capped and polyadenylated mRNA. While the promoters of downstream genes in the array are inactive, all downstream genes are nonetheless transcribed due to transcriptional readthrough from the active promoter, probably caused by inefficient RNA polymerase II termination. This readthrough appears to suppress downstream promoters through transcriptional interference, resulting in aberrant non-capped transcripts that are not translated, ensuring that only the active gene is expressed. Simultaneously, long antisense transcription originating from the chosen Or promoter covers upstream genes, presumably silencing them. Ants therefore appear to have evolved a unique transcriptional-interference-based mechanism to express a single OR protein from an array of Or genes with functionally similar promoters.

Gene families can contain hundreds of highly similar sequences, many of which perform related yet non-redundant functions and are specifically expressed in different cells. This raises the fundamental question of how cells select the correct gene to be activated at the right place and time. This issue is particularly pronounced in the case of sensory receptor genes, such as Or genes, which represent an extreme case of this regulatory challenge. The mouse genome encodes over 1,400 Or genes, making up 6% of the total gene complement[4] (Fig. 1a). Despite this large number of potential options, each olfactory sensory neuron (OSN) expresses only a single Or gene in the mature state[1,2]. This is achieved through a combination of several mechanisms. First, spatial patterning of the olfactory epithelium restricts a subset of Or genes that can be expressed in each region[5]. Second, during OSN development, multiple Or enhancers assemble into a superenhancer hub, which stochastically activates a single Or allele out of all those available for expression[6]. Once an Or promoter is chosen, negative feedback through both the unfolded protein response pathway and OR protein-independent mechanisms prevents the activation of all other Or genes, including the second allele of the chosen gene, ensuring that only one allele of one Or gene is expressed per neuron[7,8]. Stochastic selection followed by negative feedback is therefore considered the paradigm for gene choice in large Or gene families[9].

By contrast, smaller Or gene families, such as those in *Drosophila*, are subject to deterministic gene choice during development. The *Drosophila melanogaster* genome contains just 60 Or genes (Fig. 1a), which are organized into zones of expression on the antenna and maxillary palps[10]. Expression of a distinct set of transcription factors (TFs) in each OSN type[11,12] appears to be required for the expression of specific subsets of Or genes (for example, *Acj6* and *Pdm3* have been experimentally demonstrated to control the OSN fate)[3,13,14]. Similar to mice, the Or gene choice is first restricted by a subset of spatial patterning genes that specify the fate of sensory organ precursors in the antennal disc[15], which then undergo two Notch-mediated binary fate decisions[16], leading to the specification of Or gene identity. Consistent with deterministic specification, ectopic expression of Or genes does not elicit negative feedback on endogenously expressed genes[17]. This deterministic mechanism has been accepted as the gene choice paradigm applicable to the small Or gene families in insects[3].

[1]Department of Biology, New York University, New York, NY, USA. [2]Department of Biology, University of Florida, Gainesville, FL, USA. [3]Program in Biology and Center for Genomics and Systems Biology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. [4]Howard Hughes Medical Institute, Sylvester Comprehensive Cancer Center, Miller School of Medicine, University of Miami, Miami, FL, USA. ✉e-mail: dxr1274@miami.edu; hua.yan@ufl.edu; cd38@nyu.edu

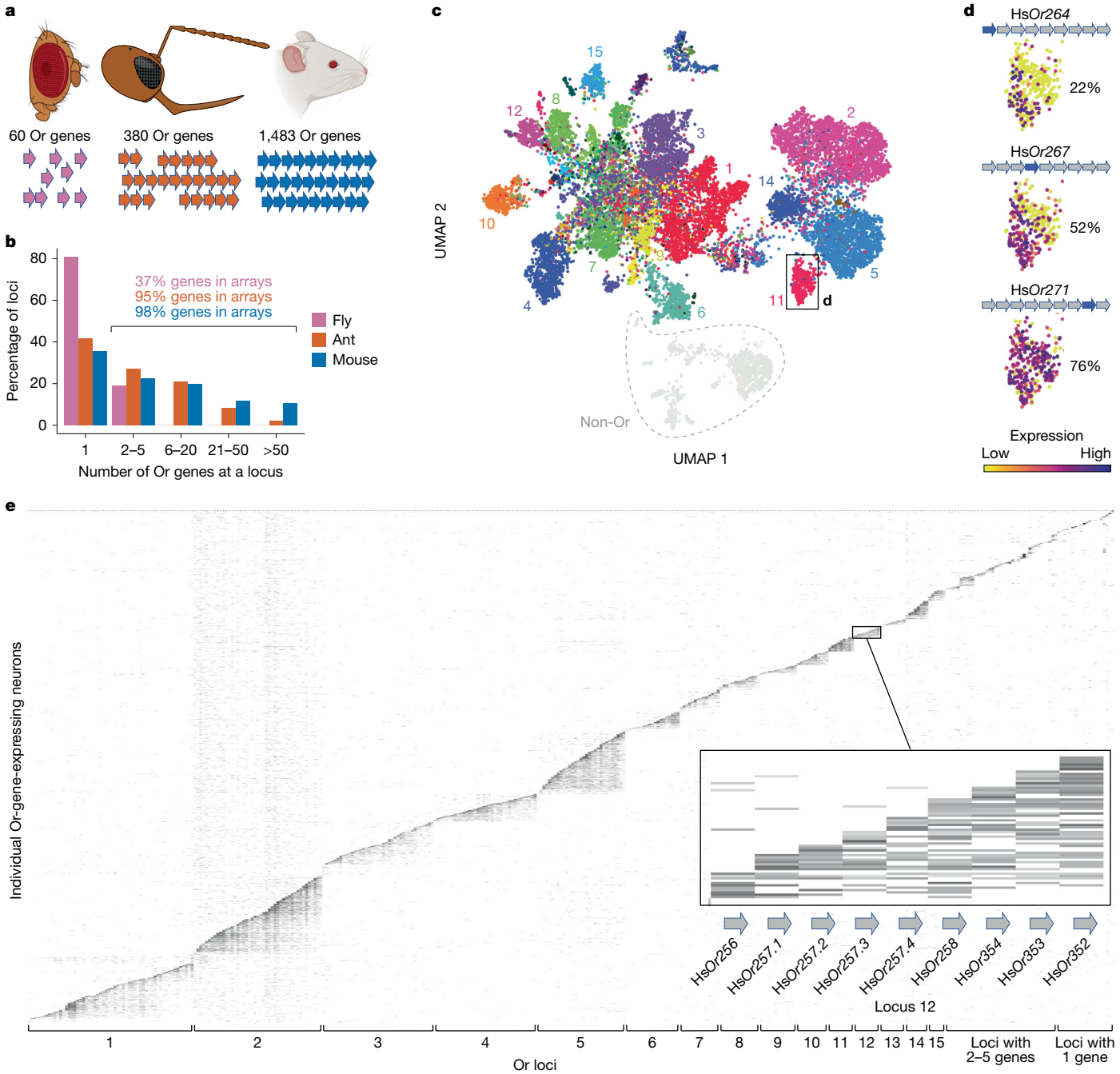


Fig. 1 | Genomic arrangement and expression pattern of Or genes in *H. saltator*. **a**, The number of Or genes in different species and a schematic of their genomic organization (illustrative of, but not an exact depiction of the proportion of clustered versus singleton Or genes); the arrows represent individual genes and groups of arrows represent tandem arrays in the genome. The images of the fly and mouse heads were created in BioRender. **b**, The size distribution of Or loci in different species. **c**, RNA-based UMAP of neurons in *H. saltator*. Colours represent genomic loci to which the expressed Or genes belong. The largest loci are numbered. **d**, Expression pattern of selected Or genes from locus 11. The percentages represent the fractions of cells in the cluster expressing the gene. **e**, The expression level of Or genes in Or-gene-expressing OSNs. Inset: co-expression of individual Or genes at locus 12.

However, social insects such as ants, bees and wasps—which rely heavily on olfactory communication to navigate their social environment, recognize colony members and maintain hierarchical structures within their colonies[18]—have large numbers of Or genes. For example, the jumping ant *H. saltator* possesses 380 Or genes[19] (Fig. 1a) and the clonal raider ant *Ooceraea biroi* has more than 500 (ref. 20). The increased number of Or genes in *Harpegnathos* has arisen primarily through tandem duplications, such that 95% of all Or genes belong to arrays containing between 2 and 58 genes, which are almost exclusively arranged head-to-tail within the arrays (Fig. 1b).

This arrangement of Or genes presents a unique regulatory challenge. Although the genes within arrays are closely related, the proteins that they encode respond to different ligands[21,22], suggesting that precise regulation is necessary to ensure that the correct OR protein is expressed in each OSN. Indeed, an earlier study in *O. biroi* showed that, although clustered Or genes may be co-expressed, only a single Or mRNA is exported into the cytoplasm in each OSN[23]. To investigate the regulatory mechanisms enabling singular OR protein expression in ants, we analysed single-nucleus RNA-sequencing (snRNA-seq) data in *H. saltator*. We identified a striking and consistent pattern of Or

gene co-transcription within arrays reminiscent of that described for *O. biroi*. By integrating short- and long-read snRNA-seq, long-read bulk RNA-seq and single-nucleus assay for transposase-accessible chromatin with high-throughput sequencing (snATAC–seq), we elucidated a transcription-interference-based mechanism that allows for the expression of a single protein in a given OSN. In the absence of shared regulatory elements, such as locus control regions, this mechanism probably evolved to regulate neighbouring promoters, preventing the simultaneous activation of multiple functionally similar genes.

Co-transcription through Pol II readthrough

To investigate Or-gene expression patterns in the ant antennae, we performed multimodal snRNA-seq and snATAC–seq (multiome) experiments and combined the RNA-seq portion of the data with additional snRNA-seq data[24]. We subsetted neurons and classified them by the receptor genes that they expressed. We then visualized the data using a uniform manifold approximation and projection (UMAP) projection, colouring cells based on the genomic array to which the expressed Or genes belonged. Neurons expressing Or genes from the same genomic array clustered together, suggesting that they expressed shared genes (Fig. 1c). Further analysis of such clusters revealed a distinct expression pattern across the Or genes within the array: the more 3′ an Or gene was located in the array, the more broadly it was expressed, and the smallest subset of cells expressed the most 5′ genes along with all of the other 3′ genes (Fig. 1d). To illustrate expression in individual cells, we plotted a heat map in which the Or genes were grouped by genomic array along the *x* axis, and neurons were arranged by their most upstream expressed gene along the *y* axis (Fig. 1e). The heat map displayed a clear 'stair-step' pattern: whenever a gene was expressed, all downstream genes in the array were also transcribed. In summary, arrayed Or genes in *H. saltator* exhibit a distinct pattern of co-expression of downstream genes, similar to what has been reported in *O. biroi*[23].

This pattern suggests that, once a promoter is chosen within the array, Pol II may read through all downstream genes, leading to their co-expression. We noticed that the expression level on the heat map dropped as the distance from the first expressed gene increased (Fig. 1e). We quantified this in each array by averaging the expression levels of the first transcribed genes, the second transcribed genes and so on across cells expressing different sets of genes within the same array. The expression level progressively decreased the farther downstream the gene was from the first transcribed gene (Fig. 2a). This gradual drop in expression is consistent with the idea that polymerase readthrough is responsible for the observed gene co-expression as the mechanisms causing transcription termination to fail may not be fully effective, and Pol II may be more likely to disengage after travelling through multiple polyadenylation sites. To further probe the idea of readthrough, we turned to our multiome data. We grouped neurons based on the first expressed gene and assessed chromatin accessibility at each promoter within the array. We found that the promoter of the first transcribed gene was the only accessible promoter in the entire array with a peak of ATAC–seq (or, in some arrays, substantially more accessible than the downstream promoters; Fig. 2b,c). This further supports the notion that transcription initiates at a single promoter, after which Pol II transcribes the downstream genes through readthrough. Thus, runaway transcription appears to cause the observed stair-step co-expression pattern.

Normal 3′ cleavage and polyadenylation

One possible explanation for the Pol II readthrough could be defective polyadenylation, which was suggested to cause a similar stair-step transcription pattern at the *Drosophila* locus containing three chemoreceptor genes, *Ir75c*, *Ir75b* and *Ir75a*: a canonical polyadenylation site was found only after *Ir75a* at the 3′ end of the locus[25]. To test whether a similar mechanism was responsible for the readthrough observed in ant Or gene arrays, we performed bulk long-read RNA-seq (Iso-seq) expecting to capture long chimeric transcripts encompassing multiple neighbouring Or genes. However, we found that the vast majority of Or-gene transcripts was separated into individual polyadenylated mRNAs (Fig. 2d). Although we did sometimes observe chimeric transcripts encompassing two genes, frequent occurrences of such transcripts were restricted to less than 5% of gene pairs, such as *H. saltator Or253* (Hs*Or253*) and Hs*Or254* (Hs*Or253* transcripts either terminate in the first two introns of Hs*Or254* or are spliced with the exons of Hs*Or254*). Moreover, poly(A) tail fragments captured by Iso-seq enabled us to determine cleavage sites and scan regions within 100 bp of them for common motifs. The top enriched motif was AAAATAAA, which contains the canonical polyadenylation signal AATAAA[26]. The putative polyadenylation signal was consistently located 21 nucleotides upstream of the cleavage site (median, −21; median absolute deviation, 24; Fig. 2e). Thus, Or-gene transcripts contain functional polyadenylation signals and are properly processed into individual gene mRNAs, ruling out defective polyadenylation as the cause of transcriptional readthrough.

Defective transcription termination

Once Pol II has transcribed the polyadenylation site, the mRNA is cleaved and polyadenylated while Pol II continues transcribing sequences 3′ of the polyadenylation site. However, the exonuclease Xrn2 (also known as Rat1) is recruited at the 5′ of the runaway transcript and rapidly degrades the nascent RNA, catching up with slower-moving Pol II and releasing it from the DNA template[27,28] (Fig. 2f). If exonuclease activity was reduced in Or-gene-expressing cells, it could allow Pol II to continue transcribing downstream genes, resulting in the observed readthrough. Support for this hypothesis came from single-nucleus long-read RNA-seq (snMAS-seq) data. Although this technique introduces a pronounced 3′ bias[29], examining the 5′ ends of reads mapping to the most 5′ expressed gene or to downstream Or genes still revealed a marked difference between them. The starting positions of reads mapping to the first expressed gene accumulated precisely at the transcription start site (TSS), as expected. By contrast, reads mapping to the downstream-expressed Or genes did not show such accumulation at their TSSs and, instead, started at apparently random locations both upstream and downstream of them (Fig. 2b). Combined with no or strongly reduced chromatin accessibility at the downstream promoters (see above), this strongly suggests that transcription initiates only at the first expressed gene in each given cell. By contrast, the 5′ ends of transcripts mapping to downstream genes appear to not represent mature mRNAs but instead to be incomplete products of digestion of nascent transcripts by the slow-moving exonuclease, which in turn allows faster Pol II to run away and continue transcribing subsequent downstream genes (Fig. 2g). Accordingly, gene coverage profiles in bulk Iso-seq data, which are enriched for mature cytoplasmic transcripts, show a sharp increase at the TSSs for both Or and non-Or genes, whereas the same profiles generated using snMAS-seq data, which contain many immature transcripts, show a sharp increase only for non-Or genes (Extended Data Fig. 1a). This suggests that Or-gene transcripts of most downstream genes are not the products of transcription initiation at the TSS but, rather, the products of aberrant termination of the previous gene.

Or promoters produce antisense RNAs

In addition to sense transcription, we found that each Or promoter drives antisense transcription. When aligning snRNA-seq reads to the genome, we observed that, in cells in which transcription initiates at a particular Or gene, extensive antisense transcription was present upstream of this gene, such that the entire locus was covered by either

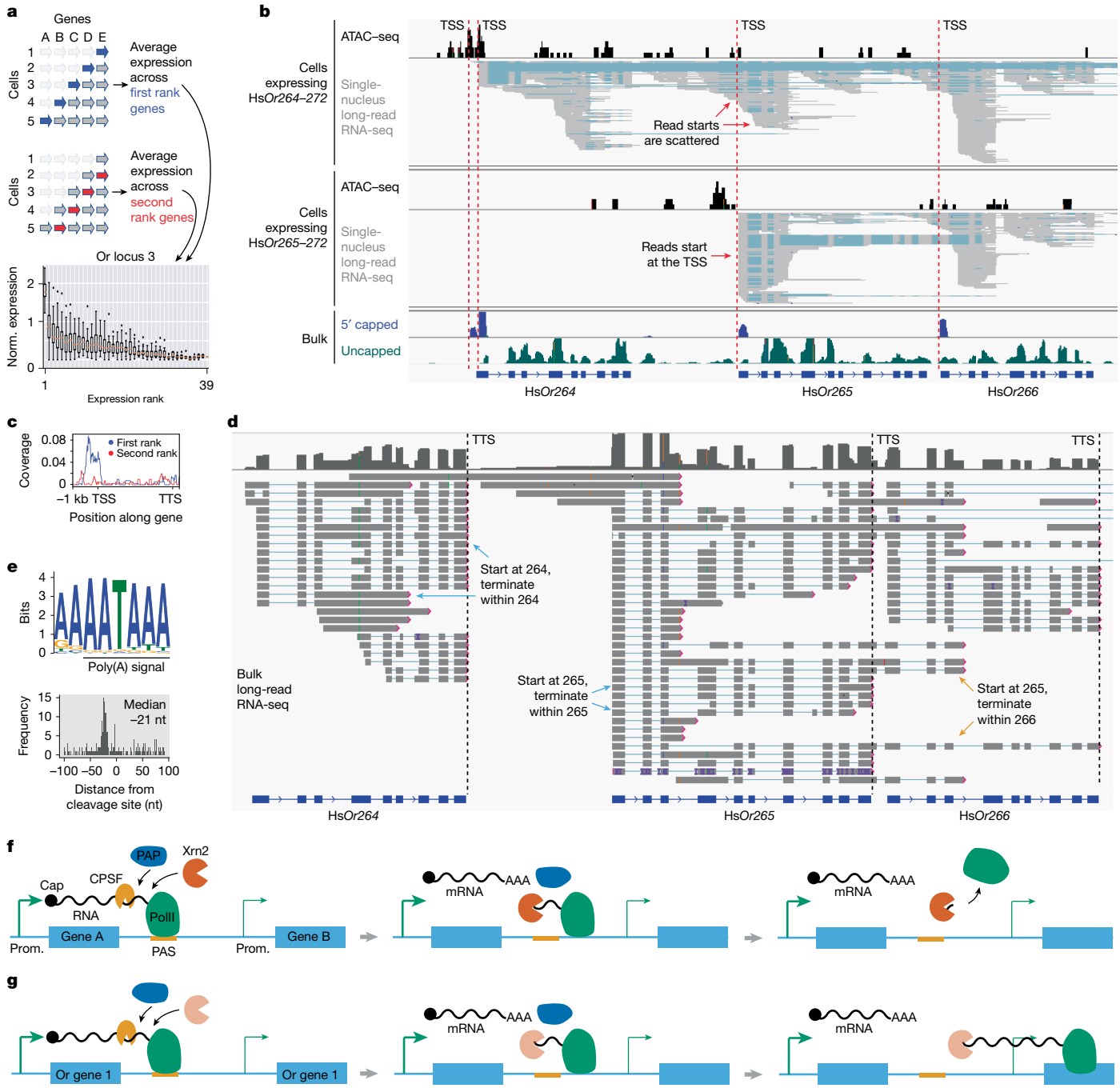


Fig. 2 | Transcriptional readthrough at Or loci. a, The expression level of genes of different rank (for example, first expressed, second expressed and so on) at locus 3. $n = 39$ possible first-rank genes for rank 1, 38 genes for rank 2. For the box plots, the box limits extend from the first to the third quartile; the centre line represents the median; and the whiskers extend to the farthest datapoint within 1.5× the interquartile range from the box. Outlier points are those past the end of the whiskers. **b**, snATAC–seq and MAS-seq data for cells choosing the promoter of Hs*Or264* (top) or Hs*Or265* (middle), showing promoter accessibility and transcript structure of Hs*Or264–266*. Bottom, bulk RNA-seq of 5′ ends of capped and uncapped transcripts. The MAS-seq and the bulk RNA-seq data shown are from one representative biological replicate out of two. **c**, Average ATAC–seq coverage across genes of Or locus 3 when a given gene is the first or the second transcribed gene. **d**, An example of bulk Iso-seq data showing the transcript structure of Hs*Or264–266*. TTS, transcription termination site. One representative biological replicate out of two is shown. **e**, The sequence and location of the putative polyadenylation signal in Or loci. **f**, The torpedo model of transcriptional termination. PAP, poly(A) polymerase; CPSF, cleavage and polyadenylation specificity factor; PAS, polyadenylation signal; prom., promoter; Xrn2, exonuclease Xrn2. **g**, Hypothesis about how low exonuclease activity may enable readthrough at Or loci: Xrn2 (pink) recruited at the 5′ end of the processed transcript is unable to catch up with Pol II (green) and 'torpedo' (disengage) it.

sense or antisense transcripts that were mutually exclusive (Fig. 3a). We used bulk Iso-seq data to study the nature of these transcripts and determined that they were generally long and spliced. In many instances, they terminated at transcription termination sites of neighbouring co-directional non-Or genes. In other cases, they presumably used cryptic termination signals (Fig. 3b). To better characterize the antisense promoters, we identified their TSSs by sequencing the 5′ ends of capped RNAs in bulk and aligning them to the genome. We found that, in all promoters within Or gene arrays (as well as in some singleton Or promoters), reads aligned to two locations on the opposite strands,

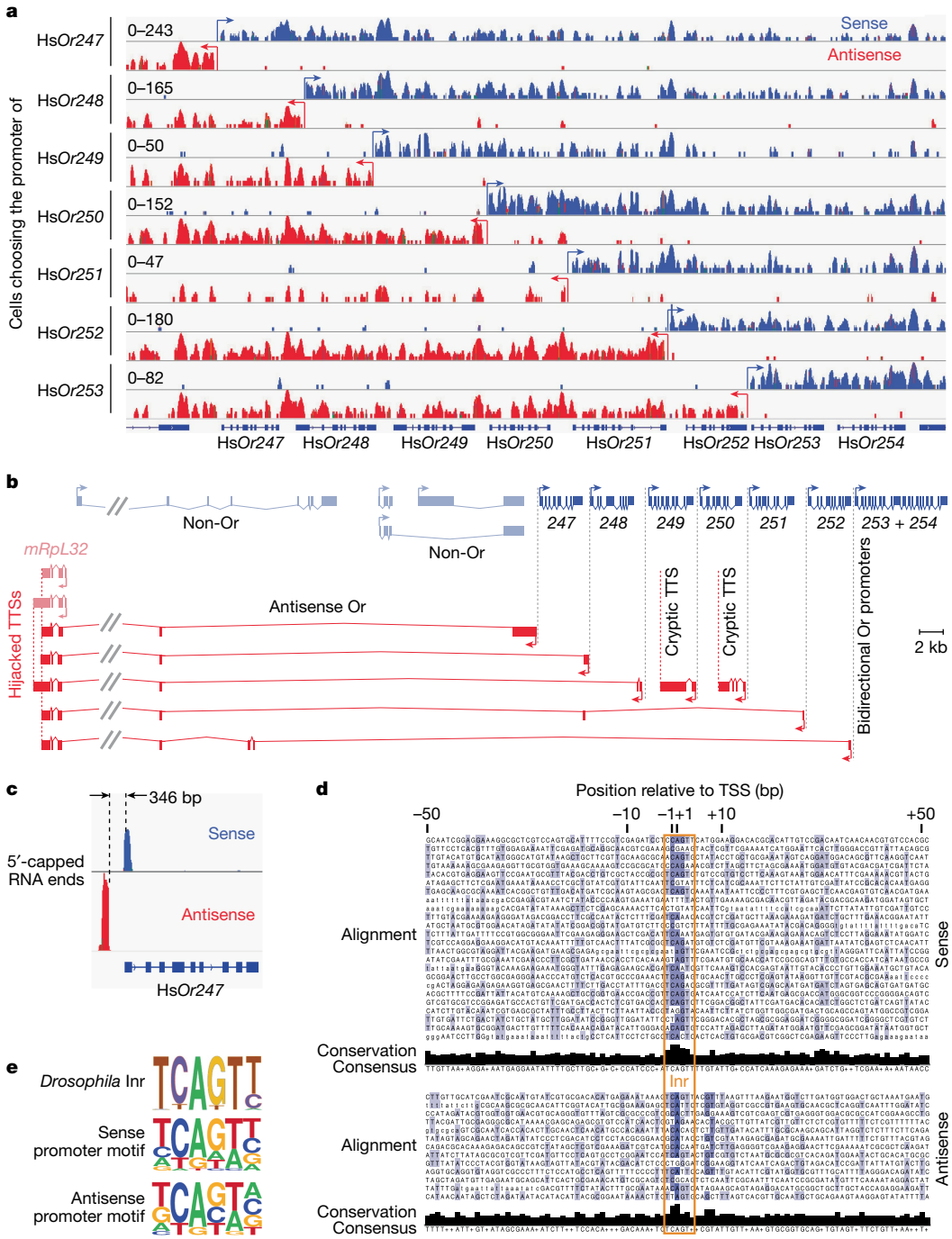


Fig. 3 | Antisense transcription at Or loci. a, log-transformedsnRNA-seq coverage of the sense and antisense strands for cells choosing promoters of different genes in locus 14. In each pair of tracks, sense and antisense tracks have identical scaling, and the scale for each pair is shown on the left. One representative biological replicate of standalone (not multiome) wild-type snRNA-seq out of three is shown. **b**, The structure of representative sense and antisense transcripts at locus 14. **c**, An example of bulk RNA-seq analysis of the 5′ ends of capped transcripts, showing peaks on both strands. One representative biological replicate out of two is shown. **d**, Alignment of peaked sense and antisense promoters. **e**, *Drosophila melanogaster* Initiator (Inr) motif[30] and the putative Inr motif in *H. saltator* Or promoters. Adapted from ref. 30, BioMed Central.

indicating that each such promoter contains a pair of core elements on the sense and on the antisense strands (Fig. 3c). Consistently, some promoters, such as Hs*Or206*, had two ATAC–seq peaks, presumably corresponding to the sense and the antisense core promoters. Thus, the two core promoters in each pair were oriented divergently, allowing simultaneous transcription in both directions (Fig. 3c). Aligning promoters at which transcription in one direction predominantly initiated at the same base yielded the consensus motif TCAGTT for the sense promoters and TCAGT for the antisense promoters. In both cases, the motif was located at −2 bp from the TSS (Fig. 3d,e). This sequence is part of the initiator promoter element, which includes the TSS[30]. Initiator is the second most abundant core promoter motif in *D. melanogaster*[31] and is also present in mammals[32]. In summary, Or promoters contain divergently oriented core promoter motifs on opposite strands, which drive antisense transcription that covers all upstream Or genes and that is mutually exclusive with sense transcription of those genes (Fig. 3a).

No obvious shared enhancers in Or gene arrays

Despite being transcriptionally active in specific neurons, Or loci are globally repressed. Bulk CUT&RUN analysis revealed that Or gene arrays are enriched for the repressive histone modification H3K27me3 and depleted for markers of active transcription, including H3K4me3 and H3K27ac. They are also in a state of inaccessible chromatin, as seen in pseudobulk ATAC–seq coverage plot generated from the snATAC–seq data (Fig. 4a,b). These observations indicate that Or gene arrays are maintained in a repressed state in most OSNs.

Based on how nuclei are clustered in the snRNA-seq-based UMAP plot (Fig. 1c), one would have expected that all OSNs within a given UMAP cluster have the potential to express genes from one gene array and are therefore regulated in a similar manner, which would be reflected in patterns of chromatin accessibility. However, this was not observed. Using snATAC–seq data, we generated a UMAP of chromatin accessibility at the single-cell level and found that neurons expressing Or genes did not segregate based on the array they expressed. Instead, cells expressing Or genes formed two broad clusters: one expressing a subset of nine-exon Or genes associated with pheromone sensing[21] and the other expressing the remaining Or genes (Fig. 4c). This suggests that chromatin accessibility alone does not distinguish between individual Or gene arrays. Furthermore, comparing pseudobulk ATAC–seq tracks of neurons expressing different Or gene arrays, we did not find that surrounding chromatin in cells expressing a given gene array was more accessible compared with in non-expressing cells. In particular, no regions of increased accessibility that could correspond to shared enhancers could be found in or near the expressed array (Fig. 4d). We could observe accessible chromatin only near the TSSs of the individual genes expressed in a given set of OSNs (Fig. 2b). This lack of specific regulatory regions suggests that each Or promoter may itself contain all of the necessary regulatory elements, similar to the situation in *Drosophila*, in which Or promoters are sufficient for gene expression choice within an array[17].

Finally, studies in other systems implicated H3K36me3 in suppression of cryptic TSSs, where H3K36me3 recruits the DNA methyltransferase Dnmt3b to prevent transcription initiation within the gene body[33]. To test whether transcriptional interference in ant Or gene arrays was accompanied by increased accumulation of this mark, we performed a CUT&RUN experiment targeting H3K36me3. However, we did not observe increased accumulation of H3K36me3 in Or gene regions (Extended Data Fig. 1b). As the function of H3K36me3 in mammals is to methylate cryptic TSSs, this is not surprising as DNA methylation does not have a role in gene expression in ants, and H3K36me3 might therefore not be needed[34].

Discussion

Our results reveal a mechanism of gene regulation within the Or gene arrays in ants that appears to be fundamentally distinct from how the classic 'one neuron, one receptor' rule observed in mammals is implemented. In ants, once transcription initiates at a single Or gene, polymerase reads through and transcribes all downstream genes within the array. Our data suggest that this co-transcription is driven by defective transcriptional termination. The Or genes themselves possess functional polyadenylation signals, ruling out defects in polyadenylation as the cause of readthrough. Instead, the evidence is consistent with defective exonuclease-mediated termination, whereby exonuclease does not efficiently digest nascent RNAs left after cleavage at the polyadenylation site and therefore does not catch up to Pol II to disengage it (Fig. 2f,g). The defect in this termination step may allow Pol II to continue transcribing downstream Or genes, leading to the observed transcription of downstream genes. The unusual defective termination process appears to have an important function: runaway transcription through downstream genes should repress transcription initiation at their promoters through transcriptional interference[28,35]. This effect may be direct: for example, a study of two co-directional genes in yeast, in which the transcription unit of the upstream gene overlaps the promoter of the downstream gene[36], showed that transcription of the upstream gene represses the downstream gene because transcribing Pol II disassembles nucleosomes in its path and then reassembles them in its wake, thereby actively assembling a high level of nucleosomes over the downstream promoter and rendering the chromatin inaccessible. Moreover, this effect may also be indirect: another study of two other co-directional genes in yeast[37] showed that transcriptional interference from the upstream gene prevents binding of an activator TF to the promoter of the downstream gene[38]. As for the RNAs generated during the runaway transcription, they probably do not mature into functional mRNAs as their 5′ ends are not 5′ capped. The only capped transcripts are those of the genes that are the first being transcribed. The 5′ ends of downstream genes are therefore probably the result of partial digestion by the 5′ exonuclease and are uncapped and therefore degraded. Our proposed mechanism may explain the pattern earlier observed in the clonal raider ant *O. biroi*, in which only the transcripts of the first expressed Or gene in the array are exported from the nucleus[23].

We propose that, similar to the downstream genes, the upstream genes in the array are repressed through transcriptional interference, but in this case by antisense transcription. We detected divergent transcription from each Or promoter that is chosen to be first transcribed, with antisense transcripts covering all upstream genes. This antisense transcription probably represses the upstream genes through transcriptional interference, as transcription of overlapping convergently oriented genes is mutually exclusive[39,40]. In this regard, the gene regulation that we are describing appears different from another well-characterized example of clustered genes that possess bidirectional promoters: protocadherins[41]. Sense and antisense promoters in protocadherins are oriented convergently, such that the expression is mutually exclusive: antisense is expressed first, and it then recruits a DNA demethylase to activate the sense promoter later in development[42]. Although the exact mechanism in *H. saltator* remains speculative, transcriptional interference alone may be sufficient to silence upstream genes[43].

A key question is whether the receptor choice within arrays is fully stochastic or whether it has at least some elements of deterministic specification (Extended Data Figs. 1c and 2a,b). Furthermore, stochastic specification would prompt a question of how to reconcile receptor choices between the two alleles. Mammals achieve this through negative feedback of a chosen receptor onto all remaining Or gene sequences, including the second allele of the chosen Or gene, but such a mechanism is not part of the *cis* mechanism described here that allows one gene within one cluster on one chromosome to be expressed. Finally, regarding how only a single array is selected, the mechanism of locus choice in ants probably resembles the mechanism of gene choice in flies, which is deterministic and requires a specific set of TFs for activation. However, while a given combination of TFs leads to the expression of one individual gene in flies, ants choose an entire array as subclustering cells that express Or genes from the same locus shows that they are not transcriptionally distinct at a level missed in the full-sample UMAP (Extended Data Fig. 2a,b). Presumably, the same set of TFs might bind to each of the promoters in this array, but the first gene to be activated represses all other Or genes in the array through the mechanism described above. For example, in our data, the gene encoding the TF Buttonless is exclusively expressed in OSNs expressing Or genes of locus 15 (Extended Data Fig. 1c), and it appears to be expressed in all such OSNs regardless of which gene within the locus they choose. Future knockdown studies of locus-specific TFs will test this hypothesis. Regardless of how a promoter is chosen from a set of equivalent promoters, once this happens, transcriptional interference mediated by defective termination and antisense

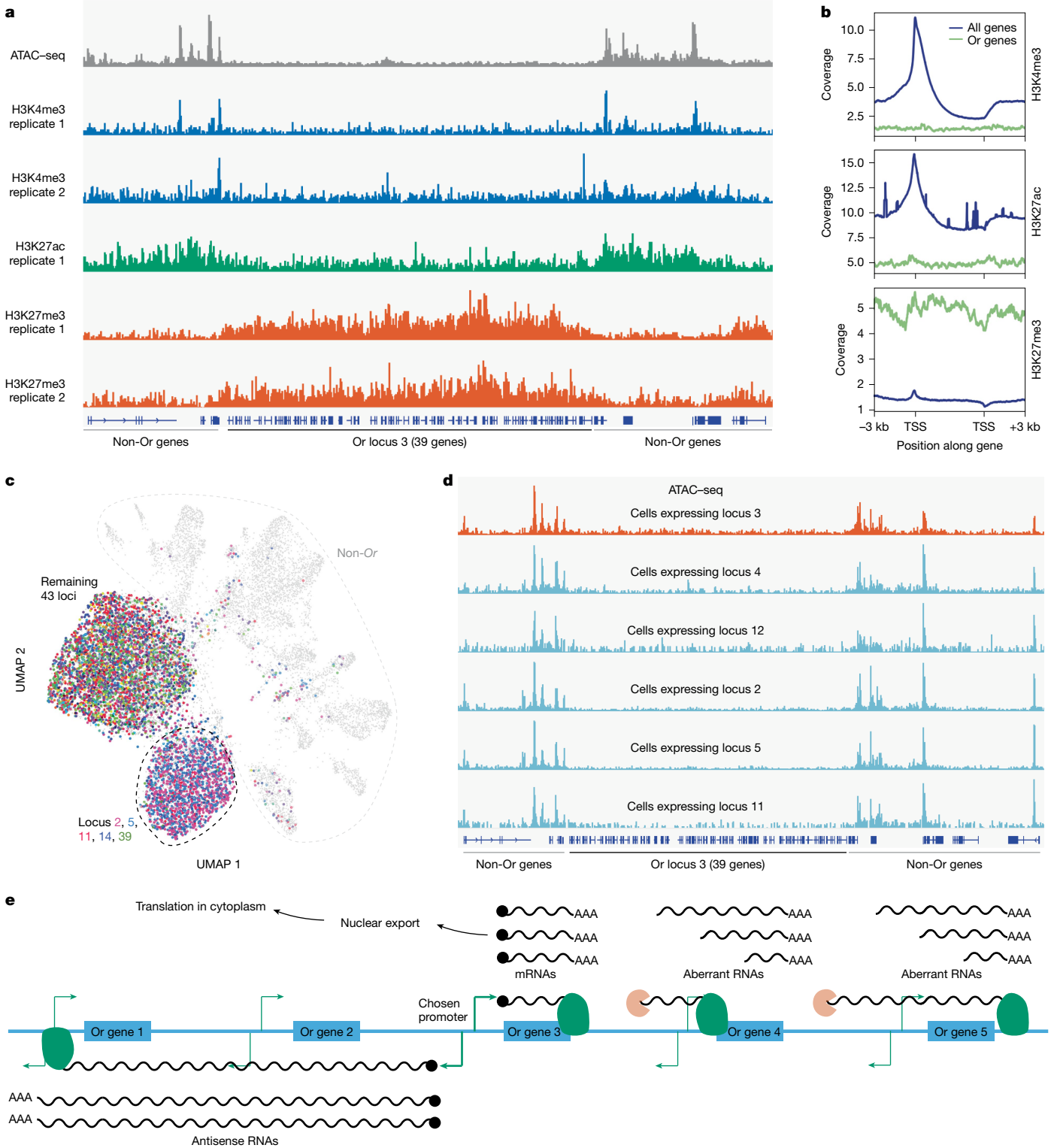


Fig. 4 | The chromatin environment of Or loci. a, Pseudobulk snATAC–seq and bulk CUT&RUN data for major chromatin modifications at locus 3. **b**, The average coverage profile of the signal from the chromatin modifications shown in **a** across all genes and Or genes. One representative biological replicate out of three (H3K4me3) or two (H3K27me3, H3K27ac) is shown. **c**, ATAC-based UMAP of antennal cells. Colours represent genomic loci to which the expressed Or genes belong. **d**, snATAC–seq data showing chromatin accessibility at locus 3 in cells that express genes from locus 3 and cells that express genes from other loci. **e**, The proposed mechanism that leads to the production of a single capped mRNA and therefore a single OR protein in each cell.

transcription might ensure that only one protein is expressed per cell (Fig. 4e).

The stair-step expression pattern is conserved in other hymenopterans, such as *O. biroi* and the honey bee *Apis mellifera*, suggesting that this transcriptional strategy is widespread across the group[23,44]. Importantly, the tandem arrangement of Or genes in hymenopterans mirrors the genomic organization of mammalian Or gene clusters. However, the transcriptional logic differs considerably. In mammals,

mature OSNs express only one Or gene per cell, which is achieved through competition between promoters for shared enhancers and a negative-feedback mechanism between the chosen promoter and the remaining Or alleles[6,7]. By contrast, ants co-express the chosen gene and all downstream genes within an array, but the transcripts of these genes are not productive, achieving both single promoter choice observed in mice and feedback transcriptional repression of non-chosen genes. The upstream Or genes are not transcribed in these cells because of promoter interference by the antisense transcripts. This highlights differences in how insects and mammals solved the problem of gene regulation when Or gene repertoires expanded through tandem duplications.

The co-expression pattern seen in ants resembles the transcription pattern in several chemosensory receptor loci in *Drosophila*, such as *Ir75c*, *Ir75b* and *Ir75a*[11,25]. Nevertheless, there appear to be important mechanistic distinctions. First, the transcriptional readthrough at the *Drosophila Ir75c/b/a* locus was proposed to stem from the lack of canonical polyadenylation signals at the end of *Ir75c* and *Ir75b*, while we show that ant Or genes possess polyadenylation signals. Instead, we propose that transcriptional termination is defective due to inefficient removal of RNA Pol II by Xrn2 exonuclease. Second, antisense RNAs potentially repressing genes upstream of the chosen promoter have not been reported in any *Drosophila* chemoreceptor arrays. Instead, cells that co-express *Ir75b* and *Ir75a* express a TF that represses the more upstream *Ir75c*, and cells that express only *Ir75a* repress both *Ir75c* and *Ir75b* through this mechanism[25]. Considering that the duplications giving rise to the *Ir75c/b/a* complex occurred a relatively long time (45–60 million years) ago[45], it is plausible that repression of individual genes by TFs evolves gradually over time, rendering the transcription interference obsolete. The mechanism we propose here—repression of upstream genes through antisense transcription—may serve as an immediate regulatory solution after extensive gene duplications.

On the mechanistic side, it remains unclear why exonuclease activity in the ant Or gene arrays is reduced to such an extent that transcriptional termination is defective, enabling readthrough. This phenomenon could be specific to Or genes. For example, Or RNAs may possess structural motifs that block or slow down the exonuclease, preventing it from catching up with Pol II. However, this would lead to the production of nascent transcripts truncated at specific locations, which is not what we observed. Alternatively, the chromatin environment in Or loci could inhibit exonuclease activity through the recruitment of a factor that slows down the exonuclease. For example, H3K9me3 in *Drosophila* piRNA loci recruits the Rhino–Deadlock–Cutoff complex, where Cutoff binds to the ends of nascent RNAs, protecting them from the Xrn2 exonuclease[46]. We do observe accumulation of a repressive mark, H3K27me3, at Or loci. Alternatively, exonuclease expression could be specifically downregulated in OSNs, or perhaps *H. saltator* exonuclease acquired mutations lead to its reduced processivity. As a possible indication that reduced exonuclease activity is found outside of OSNs, we have previously described that Or genes are sometimes expressed in non-neuronal tissues if they are adjacent to a non-Or gene expressed there; it is conceivable that this apparent ectopic expression is caused by readthrough from the non-Or genes[24]. Further investigation into expression levels of various termination factors in ant OSNs and the sequence and structure of ant exonuclease may shed light on this question.

In mammals, shared enhancers or locus control regions mediate Or gene choice[6], ensuring that only one gene is expressed per cell. However, ant Or loci seem to lack such shared enhancers. Instead, all of the regulatory elements necessary for gene expression (including the enhancers of individual Or genes) may be contained within relatively small regions around each promoter, similar to the situation in flies[17]. This organization poses a problem: without shared enhancers, any TF combination that would activate a single gene in flies might activate an entire array in ants. The evolution of defective termination with readthrough sense transcription along with antisense transcription provides a solution to this problem: this interferes with the initiation of transcription of both upstream and downstream genes and prevents the translation of genes downstream of the chosen one of which the mRNAs are not capped, allowing each OSN to produce a single functional protein. Thus, ants have probably co-opted transcriptional interference as a regulatory tool, turning defective termination and antisense transcription into a workable mechanism for gene expression control. Unravelling the molecular mechanisms behind defective exonuclease-mediated termination—whether RNA-based, chromatin-related or due to mutations—will be crucial to fully understand this regulatory strategy. As ants have limited amenability to genetic manipulation, such experiments could be performed in a heterologous system. For example, if an ant Or gene array transformed into *Drosophila* showed the same readthrough, it would demonstrate that the readthrough is determined by the locus sequence rather than a defective transcription termination system in ants. This setup would also enable follow-up experiments to determine the exact sequences responsible for the readthrough.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09664-x.

1. Hanchate, N. K. et al. Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis. *Science* **350**, 1251–1255 (2015).
2. Tan, L., Li, Q. & Xie, X. S. Olfactory sensory neurons transiently express multiple olfactory receptors during development. *Mol. Syst. Biol.* **11**, 844 (2015).
3. Fuss, S. H. & Ray, A. Mechanisms of odorant receptor gene choice in *Drosophila* and vertebrates. *Mol. Cell. Neurosci.* **41**, 101–112 (2009).
4. Barnes, I. H. A. et al. Expert curation of the human and mouse olfactory receptor gene repertoires identifies conserved coding regions split across two exons. *BMC Genom.* **21**, 196 (2020).
5. Ressler, K. J., Sullivan, S. L. & Buck, L. B. A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* **73**, 597–609 (1993).
6. Markenscoff-Papadimitriou, E. et al. Enhancer interaction networks as a means for singular olfactory receptor expression. *Cell* **159**, 543–557 (2014).
7. Dalton, R. P., Lyons, D. B. & Lomvardas, S. Co-opting the unfolded protein response to elicit olfactory receptor feedback. *Cell* **155**, 321–332 (2013).
8. Pourmorady, A. D. et al. RNA-mediated symmetry breaking enables singular olfactory receptor choice. *Nature* **625**, 181–188 (2024).
9. Dalton, R. P. & Lomvardas, S. Chemosensory receptor specificity and regulation. *Annu. Rev. Neurosci.* **38**, 331–349 (2015).
10. Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**, 725–736 (1999).
11. McLaughlin, C. N. et al. Single-cell transcriptomes of developing and adult olfactory receptor neurons in *Drosophila*. *eLife* **10**, e63856 (2021).
12. Mermet, J. et al. Multilayer regulation underlies the functional precision and evolvability of the olfactory system. Preprint at *bioRxiv* https://doi.org/10.1101/2025.01.16.632932 (2025).
13. Tichy, A. L., Ray, A. & Carlson, J. R. A new *Drosophila* POU gene, *pdm3*, acts in odor receptor expression and axon targeting of olfactory neurons. *J. Neurosci.* **28**, 7121–7129 (2008).
14. Clyne, P. J. et al. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* **22**, 327–338 (1999).
15. Li, Q. et al. A functionally conserved gene regulatory network module governing olfactory neuron diversity. *PLoS Genet.* **12**, e1005780 (2016).
16. Endo, K., Aoki, T., Yoda, Y., Kimura, K.-I. & Hama, C. Notch signal organizes the *Drosophila* olfactory circuitry by diversifying the sensory neuronal lineages. *Nat. Neurosci.* **10**, 153–160 (2007).
17. Ray, A., van Naters, W., van der, G., Shiraiwa, T. & Carlson, J. R. Mechanisms of odor receptor gene choice in *Drosophila*. *Neuron* **53**, 353–369 (2007).
18. Yan, H. et al. An engineered *orco* mutation produces aberrant social behavior and defective neural development in ants. *Cell* **170**, 736–747 (2017).
19. Zhou, X. et al. Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. *PLoS Genet.* **8**, e1002930 (2012).
20. McKenzie, S. K. & Kronauer, D. J. C. The genomic architecture and molecular evolution of ant odorant receptors. *Genome Res.* **28**, 1757–1765 (2018).
21. Pask, G. M. et al. Specialized odorant receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. *Nat. Commun.* **8**, 297 (2017).
22. Slone, J. D. et al. Functional characterization of odorant receptors in the ponerine ant, *Harpegnathos saltator*. *Proc. Natl Acad. Sci. USA* **114**, 8586–8591 (2017).

23. Brahma, A. et al. Transcriptional and post-transcriptional control of odorant receptor choice in ants. *Curr. Biol.* **33**, 5456–5466 (2023).
24. Sieriebriennikov, B. et al. Orco-dependent survival of odorant receptor neurons in ants. *Sci. Adv.* **10**, eadk9000 (2024).
25. Mika, K. et al. Olfactory receptor-dependent receptor repression in *Drosophila*. *Sci. Adv.* **7**, eabe3745 (2021).
26. Gruber, A. J. et al. A comprehensive analysis of 3′ end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res.* **26**, 1145–1159 (2016).
27. Zeng, Y., Zhang, H.-W., Wu, X.-X. & Zhang, Y. Structural basis of exoribonuclease-mediated mRNA transcription termination. *Nature* **628**, 887–893 (2024).
28. Proudfoot, N. J. Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science* **352**, aad9926 (2016).
29. Calvo-Roitberg, E. et al. Challenges in identifying mRNA transcript starts and ends from long-read sequencing data. *Genome Res.* **34**, 1719–1734 (2024).
30. Ohler, U., Liao, G.-C., Niemann, H. & Rubin, G. M. Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol.* **3**, R87 (2002).
31. FitzGerald, P. C., Sturgill, D., Shyakhtenko, A., Oliver, B. & Vinson, C. Comparative genomics of *Drosophila* and human core promoters. *Genome Biol.* **7**, R53 (2006).
32. Vo Ngoc, L., Cassidy, C. J., Huang, C. Y., Duttke, S. H. C. & Kadonaga, J. T. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev.* **31**, 6–11 (2017).
33. Neri, F. et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77 (2017).
34. Sieriebriennikov, B., Reinberg, D. & Desplan, C. A molecular toolkit for superorganisms. *Trends Genet.* **37**, 846–859 (2021).
35. Greger, I. H. & Proudfoot, N. J. Poly(A) signals control both transcriptional termination and initiation between the tandem *GAL10* and *GAL7* genes of *Saccharomyces cerevisiae*. *EMBO J.* **17**, 4771–4779 (1998).
36. Hainer, S. J., Pruneski, J. A., Mitchell, R. D., Monteverde, R. M. & Martens, J. A. Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev.* **25**, 29–40 (2011).
37. Greger, I. H., Aranda, A. & Proudfoot, N. Balancing transcriptional interference and initiation on the GAL7 promoter of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **97**, 8415–8420 (2000).
38. Tsompana, M. & Buck, M. J. Chromatin accessibility: a window into the genome. *Epigenet. Chromatin* **7**, 33 (2014).
39. Makalowska, I., Lin, C.-F. & Makalowski, W. Overlapping genes in vertebrate genomes. *Comput. Biol. Chem.* **29**, 1–12 (2005).
40. Rosa, S., Duncan, S. & Dean, C. Mutually exclusive sense–antisense transcription at *FLC* facilitates environmentally induced gene repression. *Nat. Commun.* **7**, 13031 (2016).
41. Kiefer, L. et al. WAPL functions as a rheostat of protocadherin isoform diversity that controls neural wiring. *Science* **380**, eadf8440 (2023).
42. Canzio, D. et al. Antisense lncRNA transcription mediates DNA demethylation to drive stochastic protocadherin α promoter choice. *Cell* **177**, 639–653 (2019).
43. Hobson, D. J., Wei, W., Steinmetz, L. M. & Svejstrup, J. Q. RNA polymerase II collision interrupts convergent transcription. *Mol. Cell* **48**, 365–374 (2012).
44. Zhang, W. et al. Evolutionary process underlying receptor gene expansion and cellular divergence of olfactory sensory neurons in honeybees. *Mol. Biol. Evol.* **42**, msaf080 (2025).
45. Prieto-Godino, L. L. et al. Evolution of acid-sensing olfactory circuits in drosophilids. *Neuron* **93**, 661–676 (2017).
46. Chen, Y.-C. A. et al. Cutoff suppresses RNA polymerase II termination to ensure expression of piRNA precursors. *Mol. Cell* **63**, 97–109 (2016).

Methods

Ant rearing

Wild-type *H. saltator* were reared at 25 °C inside a USDA designated ant room at New York University. Colonies were maintained in plastic boxes with a plaster floor, containing a depression covered by a glass plate, which served as the nest area[47]. The plaster was watered regularly to maintain humidity, and ants were fed with crickets (length, one-quarter to three-eighth inch (6.35 mm to 9.53 mm)) purchased from Ghann's Cricket Farm. The animals used in all of the experiments were worker females randomly picked from stock colonies.

Multiome sequencing

At least 29 individual antennae per reaction were dissected and ground in a metal cup on dry ice. After grinding, the cup and pestle were moved on wet ice to thaw. Once the sample in the metal cup was fully thawed, 1 ml of NP40 lysis buffer (10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Nonidet P40 Substitute, 1 mM DTT, 1 U μl^{-1} Protector RNase inhibitor) was added, taking care to collect as much sample from the walls as possible. The mixture was transferred into a 1 ml Dounce homogenizer pre-wetted with the NP40 lysis buffer. Nuclei were released by applying 20 strokes of loose pestle, 20 strokes of tight pestle and again 20 strokes of tight pestle. Care was taken to avoid foaming, and the douncer was kept on ice between each set of 20 strokes. After douncing, the sample was split in half, and each half was passed through a 40-µm Flowmi and a pre-wetted 20-µm pluriStainer placed over a 1.5-ml tube. The tubes were centrifuged for 10 min at 500*g* and 4 °C. The supernatant was carefully discarded, and the pellet was resuspended in 250 µl PBS with 1% BSA and 0.4 U μl^{-1} Protector RNase inhibitor and mixed thoroughly by pipetting up and down 20 times. The two subsamples were then combined and filtered through a 40-µm Flowmi into a pre-wetted 10-µm pluriStrainer placed over a 1.5-ml tube.

The sample was centrifuged for 10 min at 500*g* and 4 °C. The supernatant was carefully discarded, and the pellet was resuspended in 100 µl 0.1× lysis buffer (10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.01% Tween-20, 0.01% Nonidet P40 Substitute, 0.001% digitonin, 1% BSA, 1 mM DTT, 1 U μl^{-1} Protector RNase inhibitor) and mixed gently by pipetting up and down five times. The sample was incubated on ice for 2 min. Immediately thereafter, 1 ml of wash buffer (10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 1% BSA, 1 mM DTT, 1 U μl^{-1} Protector RNase inhibitor) was added, and the tube was centrifuged for 10 min at 500*g* and 4 °C. The supernatant was carefully discarded, and 10 µl diluted nucleus buffer (10x Genomics) was added to the remaining sample. Then, 5 µl of the nucleus suspension was processed using the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression kit (10x Genomics). The resulting libraries were sequenced either on the NovaSeq 6000 or NovaSeq X Plus instrument.

Iso-seq

Twenty individual antennae per reaction were dissected and ground in a metal cup on dry ice. The powder was resuspended in TRIzol (Invitrogen) and RNA was isolated using TRIzol-chloroform extraction, followed up by clean-up of the aqueous phase using the RNA Clean & Concentrator-5 kit (Zymo Research). 200 ng of RNA was converted to cDNA using NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module (New England Biolabs), and Iso-seq libraries were prepared using SMRTbell prep kit 3.0 (Pacific Biosciences). The experiment included two biological replicates. Each library was sequenced on a single SMRT Cell of a Sequel IIe instrument.

MAS-seq

Two MAS-seq libraries (representing biological replicates) were prepared from stored cDNA samples from the previously published snRNA-seq experiment[24]. cDNA (75 ng) was used for library preparation using the MAS-seq for 10x Single Cell 3′ kit (Pacific Biosciences). One library was sequenced on a single SMRT Cell of a Sequel IIe instrument and the other library was sequenced on a single SMRT Cell of a Revio instrument.

Sequencing of 5′-capped and 5′-phosphorylated RNA ends

5′-Phosphorylated RNAs were captured by direct adapter ligation to phosphorylated 5′ ends, as in the degradome method[48,49]. To reduce bias caused by sequence-dependent efficiency of ligation[50], the 5′ adaptor was a mix of two sequences containing random bases[51] (Supplementary Table 2). To capture 5′-capped RNAs, total RNA was first dephosphorylated to remove 5′-phosphorylated species, and then decapped, which resulted in phosphorylated 5′ ends of formerly capped molecules only. This approach is based on the TERA-seq method[52]. The resulting RNAs were processed as the originally 5′-phosphorylated RNAs.

To extract RNA, at least 25 individual antennae per reaction were dissected and ground in a metal cup on dry ice. The powder was resuspended in TRIzol (Invitrogen) and RNA was isolated using TRIzol–chloroform extraction, followed up by clean-up of the aqueous phase using the RNA Clean & Concentrator-5 kit (Zymo Research) and elution in 13 µl of water. rRNA was depleted using the NEBNext RNA Depletion Core Reagent Set with RNA sample purification beads (NEB, E7870) using a custom set of probes provided in Supplementary Note 1. At the end, 15 µl of rRNA-depleted RNA was transferred into two 1.5-ml low-binding tubes for further processing.

Next, one of the two samples was dephosphorylation and decapped. Dephosphorylation of transcripts was conducted by adding 145 µl of water, 20 µl of 10x rCutSmart buffer (New England Biolabs), and 20 µl of Quick CIP (New England Biolabs) to the samples. After mixing by pipetting, the samples were incubated in a thermomixer at 37 °C and 300 rpm for 1 h. Clean-up was then again conducted using RNA Clean & Concentrator-5, according to the manufacturer's protocols and eluted in 42 µl of water. To decap the remaining RNAs, 5 µl of 10× Thermopol buffer (New England Biolabs) and 5 µl of RNA 5′ pyrophosphohydrolase (New England Biolabs) were added, and the samples were mixed by pipetting. The samples were again incubated in a thermomixer at 37 °C and 300 rpm, and clean-up was conducted using the RNA Clean & Concentrator-5 kit according to the manufacturer's protocols but eluted in 16 µl of water.

Next, the formerly capped RNA and the 5′-phosphorylated RNA samples were processed together. To ligate adapters, 2 µl of 10× T4 RNA ligase buffer, 1 µl of 25 µM BA5 oligo mix and 2 µl of T4 RNA ligase were added to each sample and mixed by pipetting. The samples were then incubated in a thermomixer at 16 °C and 300 rpm overnight. After incubation, 30 µl of water was added, and clean-up was again conducted using RNA Clean & Concentrator-5 kit according to the manufacturer's protocols but with elution in 11 µl of water. Next, 10 µl of each purified sample were transferred to PCR tubes. To each tube, 5 µl of 5× FS buffer and 1 µl of 50 µM Deg-RT were added and mixed by pipetting. The samples were then placed in a thermal cycler with the heated lid set to 75 °C for 3 min at 65 °C, followed by a hold at 4 °C. After completion, the tubes were immediately removed, centrifuged and placed onto ice. To each sample on ice, 5.25 µl of water, 1.5 µl of dNTPs and 1.25 µl of 100 mM DTT were added, followed by 1 µl of SuperScript III and all of the components were mixed by pipetting. The samples were then placed in a thermal cycler with the heated lid set to 85 °C for 5 min at 25 °C, 60 min at 50 °C, 15 min at 75 °C, and followed by a hold at 4 °C. Next, cDNA was purified using 2 volumes of SPRI beads and eluted in 13.8 µl of water. To each tube, 15 µl of 2× Q5 Master Mix, 0.6 µl of 10 µM Deg-PCR-1L and 0.6 µl of 10 µM Deg-PCR-1R were added and mixed by pipetting. The samples were then placed in a thermal cycler with the heated lid set to 105 °C and subjected to the following programme: 30 s at 98 °C; two cycles of 10 s at 98 °C, 30 s at 59 °C and 12 s at 72 °C; four cycles of 10 s at 98 °C, 10 s at 68 °C, 12 s at 72 °C; and a final incubation of 3 min at 72 °C, followed by a hold at 12 °C. After amplification, 200–400 bp products were selected using 2% BluePippin cassettes

(Sage Science) and cleaned up using 1.5 volumes of SPRI beads and eluted in 13.8 µl water. A second PCR was performed by adding 15 µl of 2× Q5 Master Mix, 0.6 µl of 10 µM BPCRP1, 0.6 µl 10 µM of a corresponding BPCRPIdX indexing primer to each sample. Each sample was then placed into a thermal cycler with the heated lid set to 105 °C, and the following programme was run: 30 s at 98 °C; two cycles of 10 s at 98 °C, 30 s at 59 °C, and 14 s at 72 °C; ten cycles of 10 s at 98 °C and 14 s at 72 °C; and a final incubation of 3 min at 72 °C, followed by a hold at 12 °C. The 250–450 bp products were size-selected using Blue Pippin. The libraries were sequenced on either a NextSeq500 or a NovaSeq X Plus instrument. The assay included two biological replicates.

CUT&RUN

Nuclei were extracted from frozen antennae (a minimum of 15 individual antennae per reaction) as in the snRNA-seq protocol described earlier[24]. The obtained nucleus suspension was used as an input into a CUT&RUN reaction performed using the CUT&RUN Assay Kit (Cell Signaling Technology). All antibodies (Supplementary Table 2) were used at a dilution of 1:50 except for the IgG control, which was used at a dilution of 1:20. Modifications to the manufacturer's protocol included incubating samples for 2 days on an end-to-end rotator after adding Concanavalin A beads and primary antibody and overnight on an end-to-end rotator after adding the pAG-MNase enzyme. DNA spin column purification was conducted after digestion using the Monarch PCR & DNA Cleanup Kit 5 µg (New England Biolabs).

Library preparation was conducted using the NEBNext Ultra II DNA Library Prep Kit for Illumina according to manufacturer's protocols with several modifications aimed at preserving short fragments[53]. Workflow 3B was used in place of 3A to clean up adaptor-ligated DNA without size selection and 202.6 µl (2.1×) of beads was used instead of 87 µl (0.9×) in step 2. In step 3 of PCR enrichment of adaptor-ligated DNA, the annealing/extension time was shortened from 75 s to 25 s and 17 PCR cycles were conducted. During the PCR clean-up step, 60 µl (1.2×) of beads was used instead of 45 µl (0.9×) in step 2 and DNA was eluted in 23 µl of 0.1× TE instead of 32 µl in step 9. The libraries were size-selected using 2% BluePippin cassettes (Sage Science), keeping fragments between 185 and 600 bp. The libraries were sequenced on a Miseq or a NovaSeq X Plus instrument. The assay included three biological replicates for the IgG control and H3K4me3, two biological replicates for H3K27ac and H3K27me3, and one biological replicate for H3K36me3.

Updating gene annotations for *H. saltator*

To refine and standardize Or gene annotations, two genes not previously labelled as Or in the HSAL60 gene annotations[24] but annotated as Or genes in GenBank were added to the Or gene list: *LOC105186223* and *LOC112590023* were merged as labelled as Hs*Or378*, and *LOC105185051* was renamed to Hs*Or379*. Within the region surrounding Hs*Or211*, a previously missed neighbouring Or gene was identified: *LOC109503732* was labelled as Hs*Or211.2* and the original Hs*Or211* was renamed to Hs*Or211.1*. Moreover, *LOC112588771* was removed due to its overlap with Hs*Or211.1*. Similarly, in the vicinity of Hs*Or307*, an additional Or gene was identified. It was designated Hs*Or307.2* and the existing Hs*Or307* was renamed to Hs*Or307.1*. Finally, for each Or gene, only a single isoform was retained, and its TSS and transcription termination site were updated using Iso-seq data. The resulting set of gene predictions was named HSAL70 and it was uploaded to Gene Expression Omnibus under accession code GSE280492. The list of Or loci and genes is also provided in Supplementary Table 1.

Quantifying Or gene clustering in different species

Mouse Or genes[4] were grouped into arrays by applying a clustering criterion whereby any genes separated by less than 500 kb were assigned to the same array[54]. By contrast, the ant genome is considerably more compact[55], making a fixed-distance criterion unnecessary. Instead, Or gene arrays are readily apparent as contiguous groups of genes arranged in tight proximity, and individual arrays are typically confined to different scaffolds. As the assembly is not chromosome level, this naturally eliminates the ambiguity of assigning Or genes spanning large genomic intervals to an array. The full list of Or gene arrays in *H. saltator* is provided in Supplementary Table 1. In *Drosophila*, the genome is even more compact and Or genes are generally not arranged in clear clusters. Consequently, a fixed distance criterion was also not suitable. Instead, we relied on a more ad hoc approach based on published chromosomal maps of Or gene positions[56].

snRNA-seq analysis

The multiome sequencing data were converted to FASTQ using cellranger-arc mkfastq (Cell Ranger ARC v.2.0.1)[57], the genome was indexed using cellranger-arc mkref and reads were counted using cellranger-arc count. The RNA-seq portion of the multiome data generated in this study was combined with the previously published snRNA-seq data[24] using cellranger aggr (Cell Ranger v.7.0.0)[58]. To make the output of cellranger-arc count compatible with cellranger aggr, the name of the reference stored in the genomes attribute of barcode_info in the gex_molecule_info.h5 files was changed to the name of the reference used for processing the standalone snRNA-seq data. Using scanpy (v.1.10.3)[59], counts were depth-normalized (CP10k) and log-transformed, 2,000 highly variable genes were selected using the flavor = "seurat_v3" argument, and only nuclei with 500–1,800 detected genes and no more than 2.5% mitochondrial reads were retained for the analysis. Using scvi-tools (v.1.1.6)[60], an scVI model was set up using technology (multiome versus standalone snRNA-seq) as a categorical covariate and UMI counts and percentage of mitochondrial reads as continuous covariates. The model was trained using the early_stopping = True argument, and obtained latent representation was used to compute the neighbourhood graph and UMAP. Next, Leiden clustering was done using an arbitrary resolution of 0.25 and neuronal clusters were subsetted based on their expression of *LOC105183410*/*Syt1*, *LOC105189534*/*nSyb* and *LOC105183587*/*onecut*[24,61]. The subsetted neuronal data were reanalysed as described above, except nuclei having exactly 0% mitochondrial reads were eliminated.

To identify the Or gene array expressed in each OSN, we first classified each nucleus with respect to which receptor gene (Or gene, Ir gene, Gr gene, the mechanoreceptor *nompC* or the ammonia transporter *Rh50*) it expressed using the approach we previously applied to the standalone snRNA-seq data[24]. In brief, we used the Mann–Whitney *U*-test implemented in scipy.stats.mannwhitneyu[62] to determine whether a nucleus and its nearest neighbours expressed each receptor gene at a higher level than a background subset of non-neuronal cells (the full dataset was used for this instead of the neuronal subset). Next, for each nucleus in the neuronal subset, we determined the receptor gene with the lowest *P* value. If this gene was an Or, we extracted its array identity using the information provided in Supplementary Table 1.

To generate the heat map showing Or gene expression in different neurons, cells were ordered by the expressed array and then by the first transcribed gene within the array. To determine the first transcribed gene in each cell, genes within an array were plotted on the *x* axis and their expression level was plotted on the *y* axis. Then, for each gene, two regression lines were fitted using statsmodels.api.OLS[63]: a line using the points upstream of the chosen gene and a line using the points downstream and including the chosen gene. The difference in slopes was the greatest when the chosen gene was the first transcribed gene. As this strategy would not work for the first or the last gene in the array, five flanking genes from outside the array were added on each side.

To subcluster OSNs expressing Or genes from array 5 (Extended Data Fig. 2), the raw data were subsetted to retain only the nuclei classified as expressing that array, then analysed similarly to as described above, except the only covariate in the scVI model was technology (multiome versus standalone snRNA-seq). Removing Or genes from the list of

variable genes (Extended Data Fig. 2b) was done after highly variable gene selection.

To find TFs differentially expressed between OSN groups expressing Or genes from different arrays, we first subsetted the data to include only genes identified as putative TFs in supplementary table 4 of ref. 64 except *LOC105190671* and to include only nuclei that had an Or gene array assigned to them (that is, Or-gene-expressing OSNs), except for array 36, which only had a single nucleus assigned to it. We next ran sc.tl.rank_genes_groups and sc.tl.filter_rank_genes_groups and further filtered the list of differentially expressed TFs to retain only those that had adjusted $P < 0.05$.

Iso-seq analysis

The first steps of analysis were done in SMRT Link v.7.0.1. Primer sequences were removed using lima and poly(A) tails were trimmed and concatemers were identified and removed using isoseq3 refine. Resulting full-length non-concatemer reads were converted to FASTQ using bedtools bamtofastq (bedtools v.2.29.2)[65], and mapped to the genome using minimap2 (v.2.22)[66] with the following parameters: -ax splice:hq -uf. The aligned reads were sorted using samtools sort (samtools v.1.14)[67], split by strand using samtools view, and indexed using samtools index. Mapped reads from the two libraries were merged using samtools merge. To generate coverage profiles in Extended Data Fig. 1a, alignment BAM files were converted to BigWig files using bamCoverage (deeptools v.3.5.0)[68] with the following parameters: --normalizeUsing RPGC --effectiveGenomeSize 334200000; then, profiles were plotted using plotProfile after running computeMatrix scale-regions with the following parameters: --referencePoint TSS --beforeRegionStartLength 2000 --afterRegionStartLength 2000.

Identification of the putative polyadenylation signal

Iso-seq reads mapped to each strand (see above) were filtered to retain reads ending with at least [A]×10 and having soft-clipped bases at that end. This was accomplished by converting BAM to SAM using samtools view and filtering forward-strand-mapping reads using awk '{if ($10 ~ /AAAAAAAAAA$/ && $6 ~ /S$/) {print}}' and reverse-strand-mapping reads using awk '{if ($10 ~ /^TTTTTTTTTT/ && $6 ~ /^[0–9]+S/) {print}}'. Coordinates of the mapped portions of retained reads were written to BED using bedtools bamtobed. The 3′-most-mapped coordinate of each read, which presumably corresponded to the RNA-cleavage site, was then extracted and also saved in BED format. Overlapping entries in the resulting BED file were collapsed using bedtools sort, followed by bedtools merge using the -c 5 -o sum arguments to count the number of collapsed entries. Obtained cleavage site coordinates were intersected with coordinates of Or genes + 100 bp using bedtools intersect, and the most frequently used cleavage site in each gene (entry with the highest value in the fifth column of the BED) was retained. Sequences of these cleavage sites ±100 bp, as well as negative-control sequences of the same length, but shifted 1 kb upstream, were generated using bedtools getfasta and parsed for common motifs using MEME v.5.5.7 run in the differential enrichment mode (-objfun de)[69]. The top enriched motif was AAAATAAA, which contains the canonical polyadenylation signal AATAAA[26]. Coordinates of the identified motif within each sequence were extracted from the MEME output.

snATAC–seq analysis

The outputs of cellranger-arc count (see above) were combined using cellranger-arc aggr. Next, TSS ± 500 bp of all Or genes was added to the atac_peaks.bed output while making sure to avoid adding overlapping peaks by first running bedtools subtract -A -a Or_promoters.bed -b atac_peaks.bed. Then, cellranger-arc aggr was re-run with the amended peak BED. The ATAC–seq portion of the cellranger-arc aggr output was further analysed using scanpy. Peaks that were detected in fewer than 5% cells were removed, counts were depth-normalized (CP10k) and log-transformed, and only nuclei with 1,000–6,000 detected peaks were retained for the analysis. Using scvi-tools, a peakVI model was set up with total ATAC counts as a continuous covariate. The model was trained using the early_stopping = True argument, and the obtained latent representation was used to compute the neighbourhood graph and UMAP. Cells were classified by an expressed Or gene array as described for snRNA-seq data above. subset-bam (https://github.com/10XGenomics/subset-bam) was used to generate pseudobulk ATAC–seq.

MAS-seq analysis

The first steps of analysis were done in SMRT Link v.13.0.0. Reads were segmented using skera split, primer sequences were removed using lima, UMIs and cell barcodes were clipped from the reads using isoseq tag with the --design argument set to T-12U-16B, and poly(A) tails were trimmed and concatemers were identified and removed using isoseq refine. Cell barcodes were corrected and cells versus empty droplets were identified using isoseq correct using the following list of reference barcodes: https://downloads.pacbcloud.com/public/dataset/MAS-Seq/REF-10x_barcodes/3M-february-2018-REVERSE-COMPLEMENTED.txt.gz. PCR deduplication was performed using isoseq groupdedup and the resulting reads were mapped to the genome using pbmm2 align with the following parameters: --preset ISOSEQ --sort. The aligned reads were split by strand using samtools view, and indexed using samtools index. As the MAS-seq libraries were generated from the same cDNA as the previously published 10x data[24], barcodes corresponding to nuclei expressing a given set of Or genes were extracted from the 10x data, reverse-complemented and used as an input to subset-bam to generate pseudobulk MAS-seq data for the desired subset of nuclei. To generate coverage profiles in Extended Data Fig. 1a, alignment BAM files were converted to BigWig files using bamCoverage (deeptools v.3.5.0)[68] with the following parameters: --normalizeUsing RPGC --effectiveGenomeSize 334200000, then profiles were plotted using plotProfile after running computeMatrix scale-regions with the following parameters: --referencePoint TSS --beforeRegionStartLength 2000 --afterRegionStartLength 2000.

Analysis of 5′-capped and 5′-phosphorylated RNA end sequencing

Reads were mapped to the genome using bwa-mem2 mem (bwa-mem2 v.2.1)[70]. Alignment files were sorted using samtools sort, split by strand using samtools view and indexed using samtools index[67].

Identification of the Initiator motif

Using Iso-seq data, Or genes containing peaked promoters (that is, promoters at which transcription in one direction predominantly initiated at the same base[71]) were identified, and TSS ± 50 bp regions of these genes were extracted using bedtools getfasta[65]. To ensure that any sequence conservation was not due to selection of recent duplicates, no more than one gene per array was selected. The resulting set of sequences was visualized using Jalview (v.2.11.4.1)[72]. Visual examination identified a single conserved region surrounding the TSS. Its sequence matched closely to the Initiator element of *Drosophila*[30].

CUT&RUN analysis

Reads were mapped to the genome using bwa-mem2 mem. Alignment files were sorted using samtools sort and indexed using samtools index[67]. They were then converted to BigWig files using bamCoverage (deeptools v.3.5.0)[68], and coverage profiles were plotted using plotHeatmap after running computeMatrix scale-regions with the following parameters: --beforeRegionStartLength 3000 --regionBodyLength 5000 --afterRegionStartLength 3000.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sequencing data generated in this study have been deposited at the Gene Expression Omnibus (GSE280477 and GSE280492) and SRA (PRJNA1178663, PRJNA1178688 and PRJNA1261453).

47. Sieber, K. et al. Embryo injections for CRISPR-mediated mutagenesis in the ant *Harpegnathos saltator*. *J. Vis. Exp.* https://doi.org/10.3791/61930 (2021).
48. Addo-Quaye, C., Eshoo, T. W., Bartel, D. P. & Axtell, M. J. Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr. Biol.* **18**, 758–762 (2008).
49. Wang, W. et al. The initial uridine of primary piRNAs does not create the tenth adenine that Is the hallmark of secondary piRNAs. *Mol. Cell* **56**, 708–716 (2014).
50. Kim, H. et al. Bias-minimized quantification of microRNA reveals widespread alternative processing and 3′ end modification. *Nucleic Acids Res.* **47**, 2630–2640 (2019).
51. Fu, Y., Wu, P.-H., Beane, T., Zamore, P. D. & Weng, Z. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genom.* **19**, 531 (2018).
52. Ibrahim, F., Oppelt, J., Maragkakis, M. & Mourelatos, Z. TERA-seq: true end-to-end sequencing of native RNA molecules for transcriptome characterization. *Nucleic Acids Res.* **49**, e115 (2021).
53. Liu, N. et al. Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. *Cell* **173**, 430–442 (2018).
54. Niimura, Y. & Nei, M. Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene* **346**, 13–21 (2005).
55. Shields, E. J., Sheng, L., Weiner, A. K., Garcia, B. A. & Bonasio, R. High-quality genome assemblies reveal long non-coding RNAs expressed in ant brains. *Cell Rep.* **23**, 3078–3090 (2018).
56. Gomez-Diaz, C., Martin, F., Garcia-Fernandez, J. M. & Alcorta, E. The two main olfactory receptor families in *Drosophila*, ORs and IRs: acomparative approach. *Front. Cell. Neurosci.* **12**, 253 (2018).
57. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
58. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
59. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
60. Gayoso, A. et al. A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166 (2022).
61. Li, H. et al. Fly Cell Atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022).
62. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
63. Seabold, S. & Perktold, J. Statsmodels: econometric and statistical modeling with Python. In *Proc. Python in Science Conference* (eds van der Walt, S. & Millman, J.) 92–96 (SciPy, 2010).
64. Gospocic, J. et al. Kr-h1 maintains distinct caste-specific neurotranscriptomes in response to socially regulated hormones. *Cell* **184**, 5807–5823 (2021).
65. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
66. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).
67. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
68. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
69. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
70. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *Proc. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (IEEE, 2019).
71. Hoskins, R. A. et al. Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* **21**, 182–192 (2011).
72. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

Acknowledgements We thank the members of Desplan and Yan laboratories for discussions and Y.-C. Lin for maintaining ant colonies. We acknowledge the library preparation and DNA sequencing support provided by the ICBR NGS Shared Resource at the University of Florida and the Genomics Core at New York University. This work was funded by the National Institutes of Health grants K99DC021991 to B.S.; R01AG058762 to C.D. and D.R.; R01EY13010 and Tamkeen under the NYUAD Center for Genomics and Systems Biology (ADHPG-CGSB) to C.D.; R01DC020203 to H.Y.; and the Human Frontier Science Program grant LT000010/2020-L to B.S.

Author contributions Conceptualization: B.S. and C.D. Methodology: B.S., O.K., A.d.B., J.W., V.F. and I.G. Software: B.S., O.K., A.d.B. and V.F. Validation: O.K., J.W. and E.B. Formal analysis: B.S., O.K., A.d.B. and V.F. Investigation: B.S., O.K., A.d.B., J.W., V.F. and E.B. Resources: Y.Z. Data curation: B.S. Writing—original draft: B.S., J.W. and C.D. Writing—review and editing: B.S., O.K., A.d.B., J.W., V.F., Y.Z., I.G., D.R., H.Y. and C.D. Visualization: B.S. and Y.Z. Supervision: D.R., H.Y. and C.D. Project administration: B.S. and C.D. Funding acquisition: B.S., D.R., H.Y. and C.D.; B.S. conceived the study with C.D. and performed most of the experiments and analyses. O.K. developed a protocol for snRNA-seq, prepared libraries, performed analyses that uncovered the stair-step expression pattern indicating co-expression within arrays, and conceptualized the next steps with B.S. and C.D.; A.d.B. performed RNA-seq analyses that identified antisense transcripts, quantified the expression of different genes in sets of co-expressed genes, developed a protocol for multiome sequencing and prepared multiome libraries. J.W. developed a protocol for CUT&RUN, prepared CUT&RUN libraries and performed sequencing of 5′-capped and 5′-phosphorylated RNA ends. V.F. analysed TF expression and identified that different arrays tend to express different sets of TFs. Y.Z. assisted with the majority of experiments and participated in conceptual discussions. I.G. supervised the study and suggested methods of sequencing of 5′-capped and 5′-phosphorylated RNA ends. D.R. supervised the study and provided conceptual feedback. H.Y. initiated snRNA-seq experiments on ant antennae and supervised the study. C.D. conceived the study with B.S. and supervised every stage of the research.

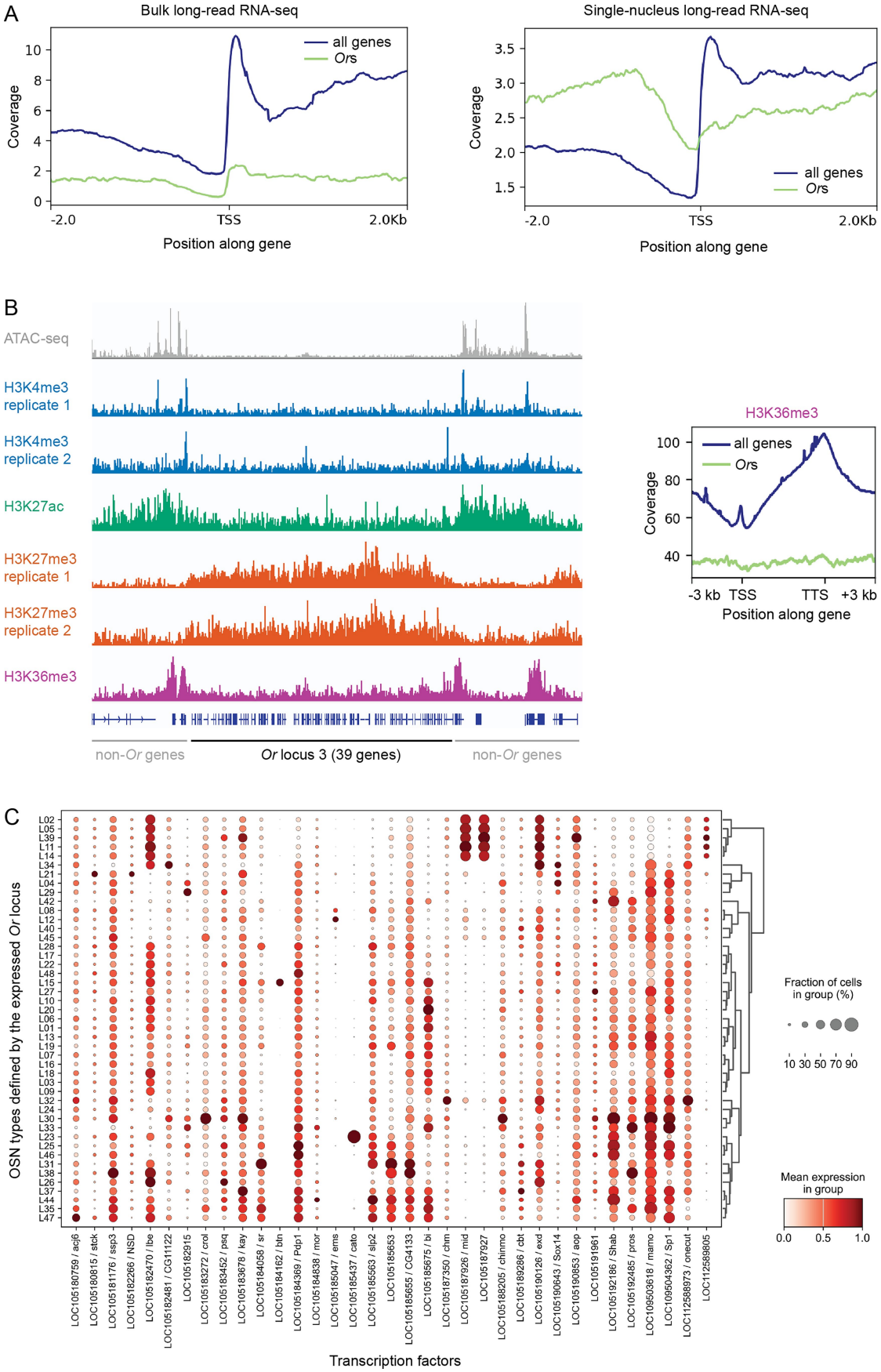
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09664-x.

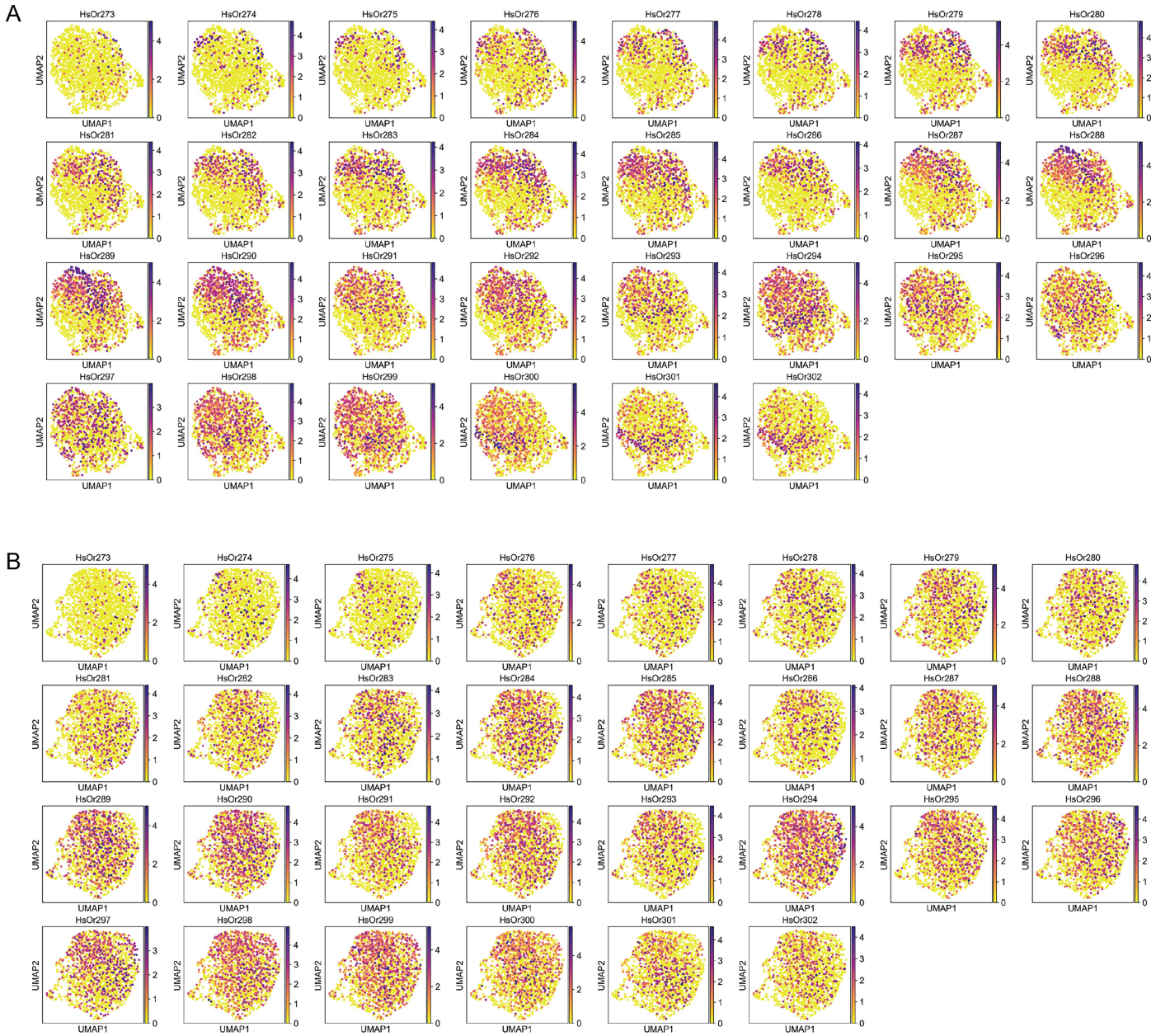
Correspondence and requests for materials should be addressed to Danny Reinberg, Hua Yan or Claude Desplan.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Additional plots. (**A**) Long-read RNA-seq coverage profile in bulk and combined single-nucleus data. One representative biological replicate out of two is shown. (**B**) CUT&RUN for H3K36me3 in Or locus 3, alongside the ATAC-seq and CUT&RUN tracks shown in Fig. 4a. (**C**) Expression of putative transcription factors in OSN types defined by the expressed Or locus. These genes were identified by searching for protein domains often present in transcription factors, and they may include genes that are not *bona fide* transcription factors.



Extended Data Fig. 2 | Subclustering of OSNs expressing locus 5 Or genes. The plots show expression level of individual Or genes in the locus (*HsOr273* is the most 5′ gene). (**A**) Subclustering where Or genes were permitted to remain in the list of highly variable genes. Nuclei are separated by the Or genes they express. (**B**) Subclustering where Or genes were removed from the list of highly variable genes. Nuclei expressing different Or genes are intermixed, showing that non-Or genes (including any TFs) do not separate OSNs expressing different Or genes from the same array.

Corresponding author(s): Claude Desplan
Hua Yan
Danny Reinberg

Last updated by author(s): Jan 3, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Cell Ranger ARC v2.0.1 Cell Ranger v7.0.0 scanpy v1.10.3 scvi-tools v1.1.6 scipy v1.14.1 statsmodels v0.14.3 SMRT Link v7.0.1 SMRT Link v13.0.0 bedtools v2.29.2 minimap2 v2.22 samtools v1.14 MEME v5.5.7 subset-bam v1.1.0 bwa-mem2 v2.1 Jalview v2.11.4.1 deeptools v3.5.0
Data analysis	Cell Ranger ARC v2.0.1 Cell Ranger v7.0.0 scanpy v1.10.3 scvi-tools v1.1.6

scipy v1.14.1
statsmodels v0.14.3
SMRT Link v7.0.1
SMRT Link v13.0.0
bedtools v2.29.2
minimap2 v2.22
samtools v1.14
MEME v5.5.7
subset-bam v1.1.0
bwa-mem2 v2.1
Jalview v2.11.4.1
deeptools v3.5.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Sequencing data generated in this study have been deposited at GEO (accessions GSE280477 and GSE280492) and SRA (accessions PRJNA1178663, PRJNA1178688, and PRJNA1261453).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No human data were used or collected.
Reporting on race, ethnicity, or other socially relevant groupings	No human data were used or collected.
Population characteristics	No human data were used or collected.
Recruitment	No human data were used or collected.
Ethics oversight	No human data were used or collected.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the multiome experiment, we targeted the number of recovered nuclei similar to what was recovered in our previously published snRNA-seq experiment (Sieriebriennikov, Sieber et al. 2024).
Data exclusions	No data were excluded from the analysis.
Replication	Iso-Seq was performed in duplicates - we visually assessed transcript structure in Or loci using IGV and concluded that the two replicates are highly consistent with each other and therefore sufficient. MAS-Seq was performed in duplicates - we visually assessed transcript structure in Or loci using IGV and concluded that the two replicates are highly consistent with each other and therefore sufficient. CUT&RUN was performed either in triplicates (IgG control and H3K4me3) or duplicates (H3K27ac and H3K27me3) - we visually assessed coverage in Or loci using IGV and concluded that all replicates are highly consistent with each other and therefore duplicates are sufficient. CUT&RUN against H3K36me3 had a single replicate because this experiment was done as a quick test of an unlikely model. The result was negative, so it is mentioned in the manuscript text, but none of the major conclusions are based on this experiment. Sequencing of 5'-capped and 5'-

monophosphorylated RNA was performed in duplicates - we visually assessed coverage in Or loci using IGV and concluded that the two replicates are highly consistent with each other and therefore sufficient.

Randomization: We did not perform any group comparisons.

Blinding: We did not perform any group comparisons.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Tri-Methyl-Histone H3 (Lys4) (C42D8) Rabbit mAb - Cell Signaling Technology Cat#9751
Tri-Methyl-Histone H3 (Lys27) (C36B11) Rabbit mAb - Cell Signaling Technology Cat#9733
Acetyl-Histone H3 (Lys27) (D5E4) XP Rabbit mAb - Cell Signaling Technology Cat#8173
Anti-Histone H3 (tri methyl K36) antibody - ChIP Grade - abcam Cat#ab9050
Rabbit (DA1E) mAb IgG XP Isotype Control (CUT&RUN) - Cell Signaling Technology Cat#66362

Validation

Validation statements copied from the manufacturer's websites:

#9751
"This antibody has been validated using SimpleChIP® Enzymatic Chromatin IP Kits."
"Tri-Methyl-Histone H3 (Lys4) Antibody detects endogenous levels of histone H3 when tri-methylated on Lys4. This antibody shows some cross-reactivity with histone H3 that is di-methylated on Lys4, but does not cross-react with non-methylated or mono-methylated histone H3 Lys4. In addition, the antibody does not cross-react with methylated histone H3 Lys9, Lys27, Lys36 or methylated histone H4 Lys20.

Species Reactivity:

Human, Mouse, Rat, Monkey, D. melanogaster, S. cerevisiae"

#9733
"This antibody has been validated using SimpleChIP® Enzymatic Chromatin IP Kits."
"Tri-Methyl-Histone H3 (Lys27) (C36B11) Rabbit mAb detects endogenous levels of histone H3 only when tri-methylated on Lys27. The antibody does not cross-react with non-methylated, mono-methylated or di-methylated Lys27. In addition, the antibody does not cross-react with mono-methylated, di-methylated or tri-methylated histone H3 at Lys4, Lys9, Lys36 or Histone H4 at Lys20.

Species Reactivity:

Human, Mouse, Rat, Monkey"

#8173
"This antibody has been validated using SimpleChIP® Enzymatic Chromatin IP Kits."
"Acetyl-Histone H3 (Lys27) (D5E4) XP® Rabbit mAb recognizes endogenous levels of histone H3 protein only when acetylated at Lys27. This antibody does not cross react with histone H3 acetylated at Lys9, 14, 18, 23, or 56. This antibody shows some cross-reactivity with acetyl-histone H2B lysine 5.

Species Reactivity:

Human, Mouse, Rat, Monkey"

#ab9050
"Abcam's high quality validation processes ensure Anti-Histone H3 (tri methyl K36) antibody - ChIP Grade (ab9050) has high sensitivity and specificity."
"Anti-Histone H3 (tri methyl K36) antibody - ChIP Grade (ab9050) specifically detects Histone H3 Tri Methyl-K36 (UniProt ID: P68431; Molecular weight: 15kDa)."

Species Reactivity (ChIP):
Tested - human
Expected - cow
Predicted - Mouse, Rat, Saccharomyces cerevisiae, Xenopus laevis, Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster, Plants, Schizosaccharomyces pombe, Zebrafish, Silk worm, Rice, Xenopus tropicalis, Trypanosoma brucei

#66362
"Rabbit (DA1E) mAb IgG XP® Isotype Control (CUT&RUN) is not directed against any known antigen. It functions as an isotype control for rabbit IgG monoclonal antibodies in the CUT&RUN assay."

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Harpegnathos saltator worker females of random age
Wild animals	No wild animals were used.
Reporting on sex	The methods section contains a statement that only females were used in the study (ant males are rare and short-lived). The manuscript makes no mention of any sex differences or the effect of sex.
Field-collected samples	No samples were collected from the field.
Ethics oversight	No ethics approval was necessary.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	No plants were used in the study.
Novel plant genotypes	No plants were used in the study.
Authentication	No plants were used in the study.